Making Do with Less: An Introduction to Compressed Sensing 2

Kurt Bryan

July 5, 2023

The Marble Problem Again

Recall the marble problem. In the 10-marble version we need to find the sparsest possible solution to the system of 5 linear equations

$$\begin{aligned} x_1 + x_2 + x_3 + x_5 + x_7 + x_9 &= 0.14 \\ x_5 + x_6 + x_7 + x_9 &= 0.44 \\ x_1 + x_3 + x_4 + x_5 + x_7 + x_8 &= -0.30 \\ x_3 + x_6 + x_8 + x_9 &= 0.14 \\ x_3 + x_5 + x_6 + x_7 + x_8 + x_{10} &= -0.30 \end{aligned}$$

in 10 unknowns. There are many infinitely many solutions—but the sparsest is $x_3 = -0.3, x_9 = 0.44$, and all other $x_i = 0$ (but we don't know this).

The Marble Problem Again

Recall the marble problem. In the 10-marble version we need to find the sparsest possible solution to the system of 5 linear equations

$$\begin{aligned} x_1 + x_2 + x_3 + x_5 + x_7 + x_9 &= 0.14 \\ x_5 + x_6 + x_7 + x_9 &= 0.44 \\ x_1 + x_3 + x_4 + x_5 + x_7 + x_8 &= -0.30 \\ x_3 + x_6 + x_8 + x_9 &= 0.14 \\ x_3 + x_5 + x_6 + x_7 + x_8 + x_{10} &= -0.30 \end{aligned}$$

in 10 unknowns. There are many infinitely many solutions—but the sparsest is $x_3 = -0.3, x_9 = 0.44$, and all other $x_i = 0$ (but we don't know this). We want to find this sparse solution.

Matrix Formulation

This linear system can be expressed as $\mathbf{Ax} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.14 \\ 0.44 \\ -0.30 \\ 0.14 \\ -0.30 \end{bmatrix}.$$

Matrix Formulation

This linear system can be expressed as $\mathbf{Ax} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.14 \\ 0.44 \\ -0.30 \\ 0.14 \\ -0.30 \end{bmatrix}.$$

The matrix $\mathbf{A}$ is called the *sensing matrix* or the *measurement matrix*. The vector $\mathbf{b}$ stems from collected data (weight measurements).

Matrix Formulation

Our central goal today is to understand some conditions under which an underdetermined linear system $\mathbf{Ax} = \mathbf{b}$ possesses a unique sparse solution (if it has one at all).

Matrix Formulation

Our central goal today is to understand some conditions under which an underdetermined linear system $\mathbf{Ax} = \mathbf{b}$ possesses a unique sparse solution (if it has one at all).

But the matrix $\mathbf{A}$ does not have to be a $0-1$ matrix. It could (in principle) be anything. However, it is usually a random matrix—entries may be integers, real numbers, or complex numbers.

Matrix Formulation

Our central goal today is to understand some conditions under which an underdetermined linear system $\mathbf{Ax} = \mathbf{b}$ possesses a unique sparse solution (if it has one at all).

But the matrix $\mathbf{A}$ does not have to be a $0 - 1$ matrix. It could (in principle) be anything. However, it is usually a random matrix—entries may be integers, real numbers, or complex numbers.

Let's look at a few concrete examples for intuition.

Uniqueness

Consider the underdetermined linear system

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Uniqueness

Consider the underdetermined linear system

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

If $\mathbf{b} = \langle 0, 0 \rangle$ then $\mathbf{x} = \langle 0, 0, 0, 0 \rangle$ is obviously the unique sparsest solution (clearly true in all cases if $\mathbf{b}$ is the zero vector).

Uniqueness

Consider the underdetermined linear system

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

If $\mathbf{b} = \langle 0, 0 \rangle$ then $\mathbf{x} = \langle 0, 0, 0, 0 \rangle$ is obviously the unique sparsest solution (clearly true in all cases if $\mathbf{b}$ is the zero vector).

For any nonzero $\mathbf{b}$ there are always multiple (but finitely many) 2-sparse solutions to this system (e.g., set $x_3 = x_4 = 0$, solve for x_1 and x_2).

Uniqueness

Consider the underdetermined linear system

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

If $\mathbf{b} = \langle 0, 0 \rangle$ then $\mathbf{x} = \langle 0, 0, 0, 0 \rangle$ is obviously the unique sparsest solution (clearly true in all cases if $\mathbf{b}$ is the zero vector).

For any nonzero $\mathbf{b}$ there are always multiple (but finitely many) 2-sparse solutions to this system (e.g., set $x_3 = x_4 = 0$, solve for x_1 and x_2).

For any $\mathbf{b}$ there are infinitely many 3 and 4-sparse solutions.

Uniqueness

What about 1-sparse solutions to

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} ?$$

Uniqueness

What about 1-sparse solutions to

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}?$$

For a typical **b** there may not be a 1-sparse solution, but it's easy to see that if a 1-sparse solution does exist, the solution is unique.

Uniqueness

To see this write

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Uniqueness

To see this write

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Then the linear system $\mathbf{Ax} = \mathbf{b}$ can be expressed

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Uniqueness

To see this write

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Then the linear system $\mathbf{Ax} = \mathbf{b}$ can be expressed

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

If a 1-sparse solution exists, $\mathbf{b}$ is a multiple of one of the columns of $\mathbf{A}$, and none of the columns are multiples of each other.

Uniqueness

More generally, if the sensing matrix $\mathbf{A}$ has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, that is,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & | & \mathbf{a}_2 & | & \cdots & | & \mathbf{a}_n \end{bmatrix}$$

then $\mathbf{Ax} = \mathbf{b}$ can be written

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

Uniqueness

More generally, if the sensing matrix $\mathbf{A}$ has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, that is,

$$\mathbf{A} = \left[\mathbf{a}_1 \quad \middle| \quad \mathbf{a}_2 \quad \middle| \quad \cdots \quad \middle| \quad \mathbf{a}_n\right]$$

then $\mathbf{Ax} = \mathbf{b}$ can be written

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

If none of the columns $\mathbf{a}_i$ are multiples of each other (that is, no two are parallel vectors) then any 1-sparse solution will be unique.

Uniqueness

On the other hand, if two columns are multiples of each other, e.g.,

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

then 1-sparse solutions may not be unique!

Uniqueness

On the other hand, if two columns are multiples of each other, e.g.,

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

then 1-sparse solutions may not be unique!

Uniqueness

On the other hand, if two columns are multiples of each other, e.g.,

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

then 1-sparse solutions may not be unique! Consider

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}.$$

Uniqueness

On the other hand, if two columns are multiples of each other, e.g.,

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

then 1-sparse solutions may not be unique! Consider

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}.$$

Both x_3 and x_4 are candidates for a 1-sparse solution.

Coherence

For two vectors $\mathbf{v}$ and $\mathbf{w}$ we define the *coherence* $\mu(\mathbf{v}, \mathbf{w})$ of these two vectors as

$$\mu(\mathbf{v}, \mathbf{w}) = \frac{|\mathbf{v} \cdot \mathbf{w}|}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

where $\mathbf{v} \cdot \mathbf{w}$ is the usual dot product and $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ is the usual Euclidean length of $\mathbf{v}$.

Coherence

For two vectors $\mathbf{v}$ and $\mathbf{w}$ we define the *coherence* $\mu(\mathbf{v}, \mathbf{w})$ of these two vectors as

$$\mu(\mathbf{v}, \mathbf{w}) = \frac{|\mathbf{v} \cdot \mathbf{w}|}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

where $\mathbf{v} \cdot \mathbf{w}$ is the usual dot product and $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ is the usual Euclidean length of $\mathbf{v}$.

Recall from basic vector calculus that $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\theta)$ where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$.

Coherence

For two vectors $\mathbf{v}$ and $\mathbf{w}$ we define the *coherence* $\mu(\mathbf{v}, \mathbf{w})$ of these two vectors as

$$\mu(\mathbf{v}, \mathbf{w}) = \frac{|\mathbf{v} \cdot \mathbf{w}|}{\|\mathbf{v}\| \|\mathbf{w}\|}$$

where $\mathbf{v} \cdot \mathbf{w}$ is the usual dot product and $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ is the usual Euclidean length of $\mathbf{v}$.

Recall from basic vector calculus that $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\theta)$ where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$.

Then $0 \leq \mu(\mathbf{v}, \mathbf{w}) \leq 1$ always. We have $\mu = 1$ exactly when $\mathbf{v}$ and $\mathbf{w}$ are parallel, $\mu = 0$ when they are orthogonal.

Coherence

If a matrix $\mathbf{A}$ has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ then the coherence $\mu(\mathbf{A})$ of $\mathbf{A}$ is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} \mu(\mathbf{a}_i, \mathbf{a}_j).$$

It's clear that $0 \leq \mu(\mathbf{A}) \leq 1$ always.

Coherence

If a matrix $\mathbf{A}$ has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ then the coherence $\mu(\mathbf{A})$ of $\mathbf{A}$ is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} \mu(\mathbf{a}_i, \mathbf{a}_j).$$

It's clear that $0 \leq \mu(\mathbf{A}) \leq 1$ always.

We have $\mu(\mathbf{A}) = 0$ exactly when all columns of $\mathbf{A}$ are orthogonal to each other.

Coherence

If a matrix $\mathbf{A}$ has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ then the coherence $\mu(\mathbf{A})$ of $\mathbf{A}$ is defined as

$$\mu(\mathbf{A}) = \max_{i \neq j} \mu(\mathbf{a}_i, \mathbf{a}_j).$$

It's clear that $0 \leq \mu(\mathbf{A}) \leq 1$ always.

We have $\mu(\mathbf{A}) = 0$ exactly when all columns of $\mathbf{A}$ are orthogonal to each other.

We have $\mu(\mathbf{A}) = 1$ if any two columns of $\mathbf{A}$ are multiples of each other.

Coherence Example 1

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix}.$$

Then $\mu(\mathbf{a}_1, \mathbf{a}_2) = 0$ (these columns are orthogonal). Also

$$\mu(\mathbf{a}_1, \mathbf{a}_3) = \frac{(1)(1) + (0)(1)}{(1)(\sqrt{2})} = 1/\sqrt{2}.$$

Coherence Example 1

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 2 \end{bmatrix}.$$

Then $\mu(\mathbf{a}_1, \mathbf{a}_2) = 0$ (these columns are orthogonal). Also

$$\mu(\mathbf{a}_1, \mathbf{a}_3) = \frac{(1)(1) + (0)(1)}{(1)(\sqrt{2})} = 1/\sqrt{2}.$$

Similar computations give pairwise coherence values $1/\sqrt{5}$, $1/\sqrt{2}$, $2/\sqrt{5}$ and $3/\sqrt{10}$. Then $\mu(\mathbf{A}) = 3/\sqrt{10} \approx 0.948$.

Coherence Example 2

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & -1 & 1 & 2 \end{bmatrix}.$$

Then $\mu(\mathbf{a}_3, \mathbf{a}_4) = 1$ (these columns are parallel) and so $\mu(\mathbf{A}) = 1$.

It's easy to see that, for example, if $\mathbf{b} = \langle b, b \rangle$ then there are many 1-sparse solutions to $\mathbf{Ax} = \mathbf{b}$.

An Easy Theorem

A bit of thought shows we've proven that if $\mu(\mathbf{A}) < 1$ (no columns of $\mathbf{A}$ are parallel or multiples of each other) then any 1-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique (if such a solution exists).

An Easy Theorem

A bit of thought shows we've proven that if $\mu(\mathbf{A}) < 1$ (no columns of $\mathbf{A}$ are parallel or multiples of each other) then any 1-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique (if such a solution exists).

But if $\mu(\mathbf{A}) = 1$, a 1-sparse solution may not be unique (depends on $\mathbf{b}$)).

An Easy Theorem

A bit of thought shows we've proven that if $\mu(\mathbf{A}) < 1$ (no columns of $\mathbf{A}$ are parallel or multiples of each other) then any 1-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique (if such a solution exists).

But if $\mu(\mathbf{A}) = 1$, a 1-sparse solution may not be unique (depends on $\mathbf{b}$)).

Can we relate $\mu(\mathbf{A})$ to uniqueness for solutions of higher sparsity? Lower values of $\mu(\mathbf{A})$ let us assert that solutions of greater sparsity are unique.

Coherence Example 3: Orthogonal Matrices

Consider a matrix $\mathbf{A}$ with $\mu(\mathbf{A}) = 0$. That is, each column of $\mathbf{A}$ is orthogonal to every other column of $\mathbf{A}$, so $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ if $i \neq j$.

Coherence Example 3: Orthogonal Matrices

Consider a matrix $\mathbf{A}$ with $\mu(\mathbf{A}) = 0$. That is, each column of $\mathbf{A}$ is orthogonal to every other column of $\mathbf{A}$, so $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ if $i \neq j$.

In this case, solutions of ANY sparsity are unique.

Coherence Example 3: Orthogonal Matrices

Consider a matrix $\mathbf{A}$ with $\mu(\mathbf{A}) = 0$. That is, each column of $\mathbf{A}$ is orthogonal to every other column of $\mathbf{A}$, so $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ if $i \neq j$.

In this case, solutions of ANY sparsity are unique. To see this, write $\mathbf{Ax} = \mathbf{b}$ as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

Coherence Example 3: Orthogonal Matrices

Consider a matrix $\mathbf{A}$ with $\mu(\mathbf{A}) = 0$. That is, each column of $\mathbf{A}$ is orthogonal to every other column of $\mathbf{A}$, so $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ if $i \neq j$.

In this case, solutions of ANY sparsity are unique. To see this, write $\mathbf{Ax} = \mathbf{b}$ as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

Take the dot product of both sides with $\mathbf{a}_i$ and use column-orthogonality to find $x_i\|\mathbf{a}_i\|^2 = \mathbf{a} \cdot \mathbf{b}$, so $x_i = (\mathbf{a}_i \cdot \mathbf{b})/\|\mathbf{a}_i\|^2$ is uniquely determined.

Low Coherence Matrices

So $\mu(\mathbf{A}) = 0$ would be ideal. Except that if $\mathbf{A}$ is an $m \times n$ matrix (m equations in n unknowns) then the columns can be orthogonal ONLY when $m \geq n$.

Low Coherence Matrices

So $\mu(\mathbf{A}) = 0$ would be ideal. Except that if $\mathbf{A}$ is an $m \times n$ matrix (m equations in n unknowns) then the columns can be orthogonal ONLY when $m \geq n$.

That is the over (or critically) determined case, but we're interested in the underdetermined case $m < n$, probably $m << n$.

Low Coherence Matrices

So $\mu(\mathbf{A}) = 0$ would be ideal. Except that if $\mathbf{A}$ is an $m \times n$ matrix (m equations in n unknowns) then the columns can be orthogonal ONLY when $m \geq n$.

That is the over (or critically) determined case, but we're interested in the underdetermined case $m < n$, probably $m << n$.

So $\mu(\mathbf{A}) = 0$ is off the table. But for a given $m \times n$ matrix $\mathbf{A}$, the lower the value of $\mu(\mathbf{A})$, the better things go for compressed sensing.

Low Coherence Matrices

So $\mu(\mathbf{A}) = 0$ would be ideal. Except that if $\mathbf{A}$ is an $m \times n$ matrix (m equations in n unknowns) then the columns can be orthogonal ONLY when $m \geq n$.

That is the over (or critically) determined case, but we're interested in the underdetermined case $m < n$, probably $m << n$.

So $\mu(\mathbf{A}) = 0$ is off the table. But for a given $m \times n$ matrix $\mathbf{A}$, the lower the value of $\mu(\mathbf{A})$, the better things go for compressed sensing.

One lower bound for the coherence of an $m \times n$ matrix is

$$\mu(\mathbf{A}) \geq \sqrt{\frac{n-m}{m(n-1)}}.$$

A Theorem

The statement that if $\mu(\mathbf{A}) < 1$ then any 1-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique can be generalized to

Theorem: If

$$\mu(\mathbf{A}) < 1/(2k-1)$$

then any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique.

A Theorem

The statement that if $\mu(\mathbf{A}) < 1$ then any 1-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique can be generalized to

Theorem: If

$$\mu(\mathbf{A}) < 1/(2k-1)$$

then any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique.

Alternatively, this inequality can be turned around to read that if

$$k < \frac{\mu(\mathbf{A}) + 1}{2\mu(\mathbf{A})}$$

then any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique.

Coherence in Reality

Coherence in Reality

- Coherence works well for determining when 1-sparse solutions are unique, for it makes it easy to determine when no two columns of the sensing matrix are parallel.

Coherence in Reality

- Coherence works well for determining when 1-sparse solutions are unique, for it makes it easy to determine when no two columns of the sensing matrix are parallel.
- Coherence isn't as sharp a tool for k-sparse solutions if $k \geq 2$; the condition $\mu(\mathbf{A}) < 1/(2k-1)$ is too stringent.

Coherence in Reality

- Coherence works well for determining when 1-sparse solutions are unique, for it makes it easy to determine when no two columns of the sensing matrix are parallel.
- Coherence isn't as sharp a tool for k-sparse solutions if $k \geq 2$; the condition $\mu(\mathbf{A}) < 1/(2k-1)$ is too stringent.
- There is a generalization of coherence called the "restricted isometry property (RIP) of order k" we can compute for $\mathbf{A}$ that gives slightly better insight into the uniqueness of sparse solutions.

Coherence in Reality

- Coherence works well for determining when 1-sparse solutions are unique, for it makes it easy to determine when no two columns of the sensing matrix are parallel.
- Coherence isn't as sharp a tool for k-sparse solutions if $k \geq 2$; the condition $\mu(\mathbf{A}) < 1/(2k-1)$ is too stringent.
- There is a generalization of coherence called the "restricted isometry property (RIP) of order k" we can compute for $\mathbf{A}$ that gives slightly better insight into the uniqueness of sparse solutions.
- Unfortunately computing the RIP for any specific matrix $\mathbf{A}$ is itself a combinatorially bad computation, and impractical.

The Nullspace Property

For a given $\mathbf{A}$, how can we prove that any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique?

The Nullspace Property

For a given $\mathbf{A}$, how can we prove that any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique?

Suppose that $\mathbf{x} = \mathbf{v}$ and $\mathbf{x} = \mathbf{w}$ are distinct k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$; let $\mathbf{u} = \mathbf{v} - \mathbf{w}$.

The Nullspace Property

For a given $\mathbf{A}$, how can we prove that any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique?

Suppose that $\mathbf{x} = \mathbf{v}$ and $\mathbf{x} = \mathbf{w}$ are distinct k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$; let $\mathbf{u} = \mathbf{v} - \mathbf{w}$. Note that $\mathbf{u}$ is $2k$-sparse and is not the zero vector, and

$$\mathbf{Au} = \mathbf{0}.$$

The Nullspace Property

For a given $\mathbf{A}$, how can we prove that any k-sparse solution to $\mathbf{Ax} = \mathbf{b}$ is unique?

Suppose that $\mathbf{x} = \mathbf{v}$ and $\mathbf{x} = \mathbf{w}$ are distinct k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$; let $\mathbf{u} = \mathbf{v} - \mathbf{w}$. Note that $\mathbf{u}$ is $2k$-sparse and is not the zero vector, and

$$\mathbf{Au} = \mathbf{0}.$$

Punchline: if there are two distinct k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ then the nullspace of $\mathbf{A}$ must contain a nontrivial $2k$-sparse vector.

The Nullspace Property

So we can thus show k-sparse solutions are unique by showing that the nullspace of $\mathbf{A}$ (the set of all solutions to $\mathbf{Ax} = 0$) contains no $2k$-sparse vectors.

The Nullspace Property

So we can thus show k-sparse solutions are unique by showing that the nullspace of $\mathbf{A}$ (the set of all solutions to $\mathbf{Ax} = 0$) contains no $2k$-sparse vectors.

One approach is brute force. For example, let

$$\mathbf{A} = \begin{bmatrix} -3 & -3 & -1 & 3 & -2 & 4 \\ 5 & -3 & 3 & 5 & 4 & -4 \\ 1 & 2 & 2 & -2 & -2 & 0 \\ -2 & 3 & 5 & -2 & -4 & -4 \end{bmatrix}$$

The Nullspace Property

So we can thus show k-sparse solutions are unique by showing that the nullspace of $\mathbf{A}$ (the set of all solutions to $\mathbf{Ax} = 0$) contains no $2k$-sparse vectors.

One approach is brute force. For example, let

$$\mathbf{A} = \begin{bmatrix} -3 & -3 & -1 & 3 & -2 & 4 \\ 5 & -3 & 3 & 5 & 4 & -4 \\ 1 & 2 & 2 & -2 & -2 & 0 \\ -2 & 3 & 5 & -2 & -4 & -4 \end{bmatrix}$$

Must a 2-sparse solution to $\mathbf{Ax} = \mathbf{b}$ be unique? Let's check the nullspace of $\mathbf{A}$ for 4-sparse vectors!

The Nullspace Property

Suppose $\mathbf{x} = \langle x_1, x_2, x_3, x_4, 0, 0 \rangle$ is a 4-sparse solution (support $\{1, 2, 3, 4\}$) to $\mathbf{Ax} = \mathbf{0}$.

The Nullspace Property

Suppose $\mathbf{x} = \langle x_1, x_2, x_3, x_4, 0, 0\rangle$ is a 4-sparse solution (support $\{1,2,3,4\}$) to $\mathbf{Ax} = \mathbf{0}$. Then

$$\begin{bmatrix} -3 & -3 & -1 & 3 \\ 5 & -3 & 3 & 5 \\ 1 & 2 & 2 & -2 \\ -2 & 3 & 5 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \mathbf{0}.$$

You can check that the only solution is $x_1 = x_2 = x_3 = x_4 = 0$.

The Nullspace Property

Suppose $\mathbf{x} = \langle x_1, x_2, x_3, x_4, 0, 0\rangle$ is a 4-sparse solution (support $\{1, 2, 3, 4\}$) to $\mathbf{Ax} = \mathbf{0}$. Then

$$\begin{bmatrix} -3 & -3 & -1 & 3 \\ 5 & -3 & 3 & 5 \\ 1 & 2 & 2 & -2 \\ -2 & 3 & 5 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \mathbf{0}.$$

You can check that the only solution is $x_1 = x_2 = x_3 = x_4 = 0$.

Now repeat assuming $\mathbf{x}$ has support $\{1, 2, 3, 5\}$, etc., all $\left(\begin{array}{c} 6 \\ 4 \end{array} \right)$ index support subset possibilities.

The Nullspace Property

This approach for proving k-sparse uniqueness on an $m \times n$ system requires solving $\left(\begin{array}{c} n \\ 2k \end{array} \right)$ linear systems of size $m \times 2k$—way too much computation!

The Nullspace Property

This approach for proving k-sparse uniqueness on an $m \times n$ system requires solving $\binom{n}{2k}$ linear systems of size $m \times 2k$—way too much computation!

But here's an alternate take on the matter that will involve randomness. Let $\mathbf{u}$ be a vector in $\mathbb{R}^n$ (might as well be a unit vector).

The Nullspace Property

This approach for proving k-sparse uniqueness on an $m \times n$ system requires solving $\binom{n}{2k}$ linear systems of size $m \times 2k$—way too much computation!

But here's an alternate take on the matter that will involve randomness. Let $\mathbf{u}$ be a vector in $\mathbb{R}^n$ (might as well be a unit vector).

If $\mathbf{Au} = \mathbf{0}$ this is equivalent to saying that

$$\|\mathbf{Au}\|^2 = 0.$$

The Restricted Isometry Property

The Restricted Isometry Property

We could prove that there are no $2k$-sparse vectors in the nullspace of $\mathbf{A}$ by showing that

$$0 < c_1 \leq \|\mathbf{Au}\|^2$$

for all $2k$-sparse vectors $\mathbf{u}$.

The Restricted Isometry Property

We could prove that there are no $2k$-sparse vectors in the nullspace of $\mathbf{A}$ by showing that

$$0 < c_1 \leq \|\mathbf{Au}\|^2$$

for all $2k$-sparse vectors $\mathbf{u}$.

A matrix $\mathbf{A}$ is said to satisfy the restricted isometry property (RIP) of order q if

$$0 < c_1 \leq \|\mathbf{Au}\|^2 \leq c_2$$

for some constants c_1, c_2, and all q-sparse unit vectors $\mathbf{u}$.

The Restricted Isometry Property

Suppose **A** is $m \times n$ and has entries that are random variables (e.g., standard normal/Gaussian).

The Restricted Isometry Property

Suppose $\mathbf{A}$ is $m \times n$ and has entries that are random variables (e.g., standard normal/Gaussian).

In this case, for any fixed unit vector $\mathbf{u}$, the quantity $\|\mathbf{Au}\|^2$ is itself a random variable.

The Restricted Isometry Property

Suppose $\mathbf{A}$ is $m \times n$ and has entries that are random variables (e.g., standard normal/Gaussian).

In this case, for any fixed unit vector $\mathbf{u}$, the quantity $\|\mathbf{Au}\|^2$ is itself a random variable. Its distribution may be impossible to work out, but we can use tools from probability to make estimates.

The Restricted Isometry Property

Suppose $\mathbf{A}$ is $m \times n$ and has entries that are random variables (e.g., standard normal/Gaussian).

In this case, for any fixed unit vector $\mathbf{u}$, the quantity $\|\mathbf{Au}\|^2$ is itself a random variable. Its distribution may be impossible to work out, but we can use tools from probability to make estimates.

In particular, a bit of detailed analysis shows that for any $\epsilon \in (0, 1)$ and $\delta > 0$, if m, n, and k stand in the proper relation then

$$P(\|\mathbf{Au}\|^2 > \delta) > 1 - \epsilon$$

for all $2k$-sparse unit vectors $\mathbf{u}$.

The Restricted Isometry Property

Crude Summary: For any given n (number of variables) and k (solution sparsity), if m (number of measurements) is large enough and $\mathbf{A}$ is an $m \times n$ matrix with suitable random entries, then the RIP of order $2k$ (so k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique) almost certainly holds.

The Restricted Isometry Property

Crude Summary: For any given n (number of variables) and k (solution sparsity), if m (number of measurements) is large enough and $\mathbf{A}$ is an $m \times n$ matrix with suitable random entries, then the RIP of order $2k$ (so k-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique) almost certainly holds.

This, despite the fact that verifying the RIP of order $2k$ for any sizable matrix is almost impossible.

An Easy Case: Bernoulli Matrices

Let $\mathbf{A}$ be an $m \times n$ matrix with entries that are independent signed-Bernoulli (± 1), equal probability.

An Easy Case: Bernoulli Matrices

Let $\mathbf{A}$ be an $m \times n$ matrix with entries that are independent signed-Bernoulli (± 1), equal probability.

Under what conditions on m and n can we be confident that 1-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique? This requires no two columns are parallel.

An Easy Case: Bernoulli Matrices

Let $\mathbf{A}$ be an $m \times n$ matrix with entries that are independent signed-Bernoulli (± 1), equal probability.

Under what conditions on m and n can we be confident that 1-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique? This requires no two columns are parallel.

A typical column $\mathbf{a}_i$ of $\mathbf{A}$ has one of 2^m realizations (± 1 patterns).

An Easy Case: Bernoulli Matrices

Let $\mathbf{A}$ be an $m \times n$ matrix with entries that are independent signed-Bernoulli (± 1), equal probability.

Under what conditions on m and n can we be confident that 1-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique? This requires no two columns are parallel.

A typical column $\mathbf{a}_i$ of $\mathbf{A}$ has one of 2^m realizations (± 1 patterns).

The probability of another column $\mathbf{a}_j$ being parallel to $\mathbf{a}_i$ is $2/2^m$ or $1/2^{m-1}$ (because this requires $\mathbf{a}_j = \pm \mathbf{a}_i$, two possibilities).

An Easy Case: Bernoulli Matrices

Let $\mathbf{A}$ be an $m \times n$ matrix with entries that are independent signed-Bernoulli (± 1), equal probability.

Under what conditions on m and n can we be confident that 1-sparse solutions to $\mathbf{Ax} = \mathbf{b}$ are unique? This requires no two columns are parallel.

A typical column $\mathbf{a}_i$ of $\mathbf{A}$ has one of 2^m realizations (± 1 patterns).

The probability of another column $\mathbf{a}_j$ being parallel to $\mathbf{a}_i$ is $2/2^m$ or $1/2^{m-1}$ (because this requires $\mathbf{a}_j = \pm \mathbf{a}_i$, two possibilities).

There are n columns, so $n(n-1)/2$ column pairs to check.

Boole's Inequality

Boole's inequality states that if $E_1, \ldots, E_N$ are events in some probability space then

$$P(E_1 \cup E_2 \cup \cdots \cup E_N) \leq P(E_1) + \cdots + P(E_N)$$

where $P(E)$ is the probability of the event E.

Boole's Inequality

Boole's inequality states that if $E_1, \ldots, E_N$ are events in some probability space then

$$P(E_1 \cup E_2 \cup \cdots \cup E_N) \leq P(E_1) + \cdots + P(E_N)$$

where $P(E)$ is the probability of the event E. It's not hard to convince yourself with a Venn diagram.

Boole's Inequality

Boole's inequality states that if $E_1, \ldots, E_N$ are events in some probability space then

$$P(E_1 \cup E_2 \cup \cdots \cup E_N) \leq P(E_1) + \cdots + P(E_N)$$

where $P(E)$ is the probability of the event E. It's not hard to convince yourself with a Venn diagram.

In plain English, the probability of at least one of the E_k occurring is no larger than the sum on the right above.

Coherence for Bernoulli Matrices

For our Bernoulli matrices, the probability that at least one pair of columns is parallel is no larger than

$$\sum_{j=1}^{n(n-1)/2} \frac{1}{2^{m-1}} = \left(\frac{n(n-1)}{2}\right)\left(\frac{1}{2^{m-1}}\right) = \frac{n(n-1)}{2^m}.$$

Coherence for Bernoulli Matrices

For our Bernoulli matrices, the probability that at least one pair of columns is parallel is no larger than

$$\sum_{j=1}^{n(n-1)/2} \frac{1}{2^{m-1}} = \left(\frac{n(n-1)}{2}\right)\left(\frac{1}{2^{m-1}}\right) = \frac{n(n-1)}{2^m}.$$

Or for simplicity, use $n^2 - n < n^2$ to see this occurs with probability less than $n^2/2^m$.

Coherence for Bernoulli Matrices

For our Bernoulli matrices, the probability that at least one pair of columns is parallel is no larger than

$$\sum_{j=1}^{n(n-1)/2} \frac{1}{2^{m-1}} = \left(\frac{n(n-1)}{2}\right)\left(\frac{1}{2^{m-1}}\right) = \frac{n(n-1)}{2^m}.$$

Or for simplicity, use $n^2 - n < n^2$ to see this occurs with probability less than $n^2/2^m$.

Equivalently, the probability that no pair of columns is parallel is larger than $1 - n^2/2^m$. That is

$$P(\mu(\mathbf{A}) < 1) > 1 - n^2/2^m.$$

Coherence for Bernoulli Matrices

For example, based on

$$P(\mu(\mathbf{A}) < 1) > 1 - n^2/2^m$$

we can see that if $n = 100$ and $m = 20$, 1-sparse solutions are unique with probability at least $1 - 10^4/2^{20} \approx 0.99$.

Coherence for Bernoulli Matrices

For example, based on

$$P(\mu(\mathbf{A}) < 1) > 1 - n^2/2^m$$

we can see that if $n = 100$ and $m = 20$, 1-sparse solutions are unique with probability at least $1 - 10^4/2^{20} \approx 0.99$.

If $n = 10000$ and $m = 50$, 1-sparse solutions are unique with probability at least $1 - 10^8/2^{50} \approx 0.0.99999991$.

Conclusion

Let $\mathbf{A}$ be an $m \times n$ matrix.

Conclusion

Let $\mathbf{A}$ be an $m \times n$ matrix. If the entries of $\mathbf{A}$ are certain types of random variables then it can be shown that if $m \geq O(k \ln(n/k))$ then the RIP condition of order $2k$ (uniqueness of k-sparse solutions) is met with high probability.

Conclusion

Let **A** be an $m \times n$ matrix. If the entries of **A** are certain types of random variables then it can be shown that if $m \geq O(k\ln(n/k))$ then the RIP condition of order $2k$ (uniqueness of k-sparse solutions) is met with high probability.

In many cases this becomes (asymptotically) roughly $m \geq 2k\ln(n/k)$.

Conclusion

Let $\mathbf{A}$ be an $m \times n$ matrix. If the entries of $\mathbf{A}$ are certain types of random variables then it can be shown that if $m \geq O(k \ln(n/k))$ then the RIP condition of order $2k$ (uniqueness of k-sparse solutions) is met with high probability.

In many cases this becomes (asymptotically) roughly $m \geq 2k \ln(n/k)$.

So verifying that a specific matrix $\mathbf{A}$ works is hard, but we can prove that "most" random matrices will work, under suitable conditions.

Conclusion

Let $\mathbf{A}$ be an $m \times n$ matrix. If the entries of $\mathbf{A}$ are certain types of random variables then it can be shown that if $m \geq O(k \ln(n/k))$ then the RIP condition of order $2k$ (uniqueness of k-sparse solutions) is met with high probability.

In many cases this becomes (asymptotically) roughly $m \geq 2k \ln(n/k)$.

So verifying that a specific matrix $\mathbf{A}$ works is hard, but we can prove that "most" random matrices will work, under suitable conditions.

One rule of thumb people have formulated is that $m > 4k$ is often sufficient.