Linear Algebra Background and Uniqueness

Bhaskar D Rao
University of California, San Diego
Email: brao@ucsd.edu
Least Squares (LS) Problem

\[ y = Ax + v, \]

where \( A \in R^{N \times M} \), \( x \in R^M \), and \( y \in R^N \).
Least Squares (LS) Problem

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M} \), \( x \in \mathbb{R}^{M} \), and \( y \in \mathbb{R}^{N} \). \( y \) are the measurements. \( y \) and \( A \) are given and one is interested in estimating \( x \).
Least Squares (LS) Problem

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M} \), \( x \in \mathbb{R}^M \), and \( y \in \mathbb{R}^N \).

\( y \) are the measurements. \( y \) and \( A \) are given and one is interested in estimating \( x \).

Assumptions

- \( N \geq M \)

- \( A \) is full rank. Rank of \( A \) is \( M \), i.e. columns of \( A \) are linearly independent.
Least Squares (LS) Problem

\[ y = Ax + v, \]

where \( A \in R^{N \times M} \), \( x \in R^M \), and \( y \in R^N \).

\( y \) are the measurements. \( y \) and \( A \) are given and one is interested in estimating \( x \)

Assumptions

\begin{itemize}
  \item \( N \geq M \)
  \item \( A \) is full rank. Rank of \( A \) is \( M \), i.e. columns of \( A \) are linearly independent.
\end{itemize}

Least Squares solution of \( x \) is given by

\[ x_{LS} = \arg \min_x [\|y - Ax\|^2] = A^+ y, \]

\( A^+ \) is the Moore-Penrose inverse given by \( A^+ = (A^T A)^{-1} A^T \).
Four fundamental subspaces

▶ $\mathbb{R}(A)$ is the range space of $A$, i.e. $\mathbb{R}(A) = \{z: z = A\alpha, \forall \alpha\}$.
Dimension of $\mathbb{R}(A) = M$.

▶ $\mathbb{N}(A)$ is the null space of $A$, i.e. $\mathbb{N}(A) = \{w: A\alpha = 0\}$.
Dimension of $\mathbb{N}(A) = 0$ by virtue of the full rank assumption.

▶ $\mathbb{R}(A^T)$ is the range space of $A^T$, i.e. $\mathbb{R}(A^T) = \{z: z = A^T\alpha, \forall \alpha\}$.
Dimension of $\mathbb{R}(A^T) = M$.

▶ $\mathbb{N}(A^T)$ is the null space of $A^T$, i.e. $\mathbb{N}(A^T) = \{w: A^T\alpha = 0\}$.
Dimension of $\mathbb{N}(A^T) = N - M$.

Some properties:

▶ $\mathbb{R}(A) \perp \mathbb{N}(A^T)$

▶ $\mathbb{R}(A^T) \perp \mathbb{N}(A)$

▶ $\mathbb{R}(A) \oplus \mathbb{N}(A^T) = \mathbb{R}^N$

▶ $\mathbb{R}(A^T) \oplus \mathbb{N}(A) = \mathbb{R}^M$
LS Geometry

Four fundamental subspaces
LS Geometry

Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{z : z = A\alpha, \forall \alpha.\}$
  Dimension of $\mathcal{R}(A) = M$.
LS Geometry

Four fundamental subspaces

- \( \mathcal{R}(A) \) is the range space of \( A \), i.e. \( \mathcal{R}(A) = \{z : z = A\alpha, \ \forall \alpha.\} \)
  Dimension of \( \mathcal{R}(A) = M. \)

- \( \mathcal{N}(A) \) is the null space of \( A \), i.e. \( \mathcal{N}(A) = \{w : Aw = 0\}. \)
  Dimension of \( \mathcal{N}(A) = 0 \) by virtue of the full rank assumption.
LS Geometry

Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{ z : z = A\alpha, \forall \alpha. \}$ Dimension of $\mathcal{R}(A) = M$.

- $\mathcal{N}(A)$ is the null space of $A$, i.e. $\mathcal{N}(A) = \{ w : Aw = 0 \}$. Dimension of $\mathcal{N}(A) = 0$ by virtue of the full rank assumption.

- $\mathcal{R}(A^T)$ is the range space of $A^T$, i.e. $\mathcal{R}(A^T) = \{ z : z = A^T\alpha, \forall \alpha. \}$ Dimension of $\mathcal{R}(A^H) = M$. 

- $\mathcal{N}(A^T)$ is the null space of $A^T$, i.e. $\mathcal{N}(A^T) = \{ w : A^Tw = 0 \}$. Dimension of $\mathcal{N}(A^T) = N - M$. 

Some properties

- $\mathcal{R}(A) \perp \mathcal{N}(A^T)$
- $\mathcal{R}(A^T) \perp \mathcal{N}(A)$
- $\mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^N$
- $\mathcal{R}(A^T) \oplus \mathcal{N}(A) = \mathbb{R}^M$
Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{z : z = A\alpha, \forall \alpha.\}$
  Dimension of $\mathcal{R}(A) = M$.

- $\mathcal{N}(A)$ is the null space of $A$, i.e. $\mathcal{N}(A) = \{w : Aw = 0\}$. Dimension of $\mathcal{N}(A) = 0$ by virtue of the full rank assumption.

- $\mathcal{R}(A^T)$ is the range space of $A^T$, i.e.
  $\mathcal{R}(A^T) = \{z : z = A^T\alpha, \forall \alpha.\}$ Dimension of $\mathcal{R}(A^H) = M$.

- $\mathcal{N}(A^T)$ is the null space of $A^T$, i.e. $\mathcal{N}(A^T) = \{w : A^Tw = 0\}$. Dimension of $\mathcal{N}(A^T) = N - M$. 
LS Geometry

Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{ z : z = A\alpha, \ \forall \alpha. \}$
  Dimension of $\mathcal{R}(A) = M$.

- $\mathcal{N}(A)$ is the null space of $A$, i.e. $\mathcal{N}(A) = \{ w : Aw = 0 \}$.
  Dimension of $\mathcal{N}(A) = 0$ by virtue of the full rank assumption.

- $\mathcal{R}(A^T)$ is the range space of $A^T$, i.e.
  $\mathcal{R}(A^T) = \{ z : z = A^T\alpha, \ \forall \alpha. \}$
  Dimension of $\mathcal{R}(A^H) = M$.

- $\mathcal{N}(A^T)$ is the null space of $A^T$, i.e. $\mathcal{N}(A^T) = \{ w : A^Tw = 0 \}$.
  Dimension of $\mathcal{N}(A^T) = N - M$.

Some properties pause

- $\mathcal{R}(A) \perp \mathcal{N}(A^T)$
LS Geometry

Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{z : z = A\alpha, \ \forall \alpha.\}$
  Dimension of $\mathcal{R}(A) = M$.

- $\mathcal{N}(A)$ is the null space of $A$, i.e. $\mathcal{N}(A) = \{w : Aw = 0\}$.
  Dimension of $\mathcal{N}(A) = 0$ by virtue of the full rank assumption.

- $\mathcal{R}(A^T)$ is the range space of $A^T$, i.e.
  $\mathcal{R}(A^T) = \{z : z = A^T\alpha, \ \forall \alpha.\}$
  Dimension of $\mathcal{R}(A^H) = M$.

- $\mathcal{N}(A^T)$ is the null space of $A^T$, i.e. $\mathcal{N}(A^T) = \{w : A^Tw = 0\}$.
  Dimension of $\mathcal{N}(A^T) = N - M$.

Some properties pause

- $\mathcal{R}(A) \perp \mathcal{N}(A^T)$
- $\mathcal{R}(A^T) \perp \mathcal{N}(A)$
LS Geometry

Four fundamental subspaces

- \( \mathcal{R}(A) \) is the range space of \( A \), i.e. \( \mathcal{R}(A) = \{ z : z = A\alpha, \ \forall \alpha \} \)
  Dimension of \( \mathcal{R}(A) = M \).

- \( \mathcal{N}(A) \) is the null space of \( A \), i.e. \( \mathcal{N}(A) = \{ w : Aw = 0 \} \).
  Dimension of \( \mathcal{N}(A) = 0 \) by virtue of the full rank assumption.

- \( \mathcal{R}(A^T) \) is the range space of \( A^T \), i.e.
  \( \mathcal{R}(A^T) = \{ z : z = A^T\alpha, \ \forall \alpha \} \)
  Dimension of \( \mathcal{R}(A^H) = M \).

- \( \mathcal{N}(A^T) \) is the null space of \( A^T \), i.e. \( \mathcal{N}(A^T) = \{ w : A^Tw = 0 \} \).
  Dimension of \( \mathcal{N}(A^T) = N - M \).

Some properties pause

- \( \mathcal{R}(A) \perp \mathcal{N}(A^T) \)
- \( \mathcal{R}(A^T) \perp \mathcal{N}(A) \)
- \( \mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^N \)
LS Geometry

Four fundamental subspaces

- $\mathcal{R}(A)$ is the range space of $A$, i.e. $\mathcal{R}(A) = \{ z : z = A\alpha, \ \forall \alpha \}$. Dimension of $\mathcal{R}(A) = M$.

- $\mathcal{N}(A)$ is the null space of $A$, i.e. $\mathcal{N}(A) = \{ w : Aw = 0 \}$. Dimension of $\mathcal{N}(A) = 0$ by virtue of the full rank assumption.

- $\mathcal{R}(A^T)$ is the range space of $A^T$, i.e.
  $\mathcal{R}(A^T) = \{ z : z = A^T\alpha, \ \forall \alpha \}$. Dimension of $\mathcal{R}(A^H) = M$.

- $\mathcal{N}(A^T)$ is the null space of $A^T$, i.e. $\mathcal{N}(A^T) = \{ w : A^Tw = 0 \}$. Dimension of $\mathcal{N}(A^T) = N - M$.

Some properties pause

- $\mathcal{R}(A) \perp \mathcal{N}(A^T)$
- $\mathcal{R}(A^T) \perp \mathcal{N}(A)$
- $\mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^N$
- $\mathcal{R}(A^T) \oplus \mathcal{N}(A) = \mathbb{R}^M$
Orthogonal projection Operators

$P_A = AA^+ \text{ is the orthogonal projector onto } R(A), \text{ the range space of } A, \forall z \in R(A)$. $P_A$ is independent of the basis used for $R(A)$.

$P_{\perp A} = I - P_A$ is the orthogonal projector onto $N(A^T)$, the null space of $A$.

$y_{LS} = Ax_{LS} = P_A y$

$y_{LS} = y - y_{LS} = P_{\perp A} y \perp R(A)$
Orthogonal projection Operators

\[ P_A = A A^T + \text{is the orthogonal projector onto } R(A), \text{the range space of } A, \forall z \in R(A). \]
\[ P_A \text{ is independent of the basis used for } R(A). \text{ Useful to choose an orthonormal basis.} \]

\[ P_{\perp A} = I - P_A \text{ is the orthogonal projector onto } N(A^T), \text{the null space of } A^T. \]

\[ y_{\text{LS}} = A x_{\text{LS}} = P_A y \]
\[ v_{\text{LS}} = y - y_{\text{LS}} = P_{\perp A} y \perp R(A) \]
Orthogonal projection Operators

- $P_A = AA^+$ is the orthogonal projector onto $\mathcal{R}(A)$, the range space of $A$, i.e. $P_Az \in \mathcal{R}(A), \forall z$. $P_A$ is independent of the basis used for $\mathcal{R}(A)$. Useful to choose an orthonormal basis.

- $P_{\perp} = I - P_A$ is the orthogonal projector onto $\mathcal{N}(A^T)$, the null space of $A$. 

- $y_{LS} = Ax_{LS} = P_Ay$

- $v_{LS} = y - y_{LS} = P_{\perp}y \perp \mathcal{R}(A)$
Orthogonal projection Operators

- $P_A = AA^+$ is the orthogonal projector onto $\mathcal{R}(A)$, the range space of $A$, i.e. $P_A z \in \mathcal{R}(A)$, $\forall z$. $P_A$ is independent of the basis used for $\mathcal{R}(A)$. Useful to choose an orthonormal basis.

- $P_A^\perp = I - P_A$ is the orthogonal projector onto $\mathcal{N}(A^T)$, the null space of $A^T$. 
Orthogonal projection Operators

▶ $P_A = AA^+$ is the orthogonal projector onto $\mathcal{R}(A)$, the range space of $A$, i.e. $P_A z \in \mathcal{R}(A)$, $\forall z$. $P_A$ is independent of the basis used for $\mathcal{R}(A)$. Useful to choose an orthonormal basis.

▶ $P_A^\perp = I - P_A$ is the orthogonal projector onto $\mathcal{N}(A^T)$, the null space of $A^T$.

▶ $y_{LS} = Ax_{LS} = P_A y$
Orthogonal projection Operators

- $P_A = AA^+$ is the orthogonal projector onto $\mathcal{R}(A)$, the range space of $A$, i.e. $P_A z \in \mathcal{R}(A), \forall z$. $P_A$ is independent of the basis used for $\mathcal{R}(A)$. Useful to choose an orthonormal basis.

- $P_A^\perp = I - P_A$ is the orthogonal projector onto $\mathcal{N}(A^T)$, the null space of $A^T$.

- $y_{LS} = Ax_{LS} = P_A y$

- $v_{LS} = y - y_{LS} = P_A^\perp y \perp \mathcal{R}(A)$
Orthogonal projection Operators

- $P_A = AA^+$ is the orthogonal projector onto $\mathcal{R}(A)$, the range space of $A$, i.e. $P_A z \in \mathcal{R}(A), \forall z$. $P_A$ is independent of the basis used for $\mathcal{R}(A)$. Useful to choose an orthonormal basis.

- $P_A^\perp = I - P_A$ is the orthogonal projector onto $\mathcal{N}(A^T)$, the null space of $A^T$.

- $y_{LS} = Ax_{LS} = P_A y$

- $v_{LS} = y - y_{LS} = P_A^\perp y \perp \mathcal{R}(A)$
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M}, x \in \mathbb{R}^M, \) and \( y \in \mathbb{R}^N. \)
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M} \), \( x \in \mathbb{R}^M \), and \( y \in \mathbb{R}^N \).

Assumptions

- \( N \leq M \)
- Rank of \( A \) is \( N \), i.e. rows of \( A \) are linearly independent.
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M} \), \( x \in \mathbb{R}^M \), and \( y \in \mathbb{R}^N \).

Assumptions

- \( N \leq M \)
- Rank of \( A \) is \( N \), i.e. rows of \( A \) are linearly independent.

Fundamental subspaces
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M}, x \in \mathbb{R}^M, \) and \( y \in \mathbb{R}^N. \)

Assumptions

- \( N \leq M \)
- Rank of \( A \) is \( N \), i.e. rows of \( A \) are linearly independent.

Fundamental subspaces

- \( \mathcal{R}(A) \) is the range space of \( A \), i.e. \( \mathcal{R}(A) = \{z : z = A\alpha, \ \forall \alpha \} \)
  - Dimension of \( \mathcal{R}(A) = N \).
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M}, x \in \mathbb{R}^M, \) and \( y \in \mathbb{R}^N. \)

Assumptions

- \( N \leq M \)
- Rank of \( A \) is \( N \), i.e. rows of \( A \) are linearly independent.

Fundamental subspaces

- \( \mathcal{R}(A) \) is the range space of \( A \), i.e. \( \mathcal{R}(A) = \{ z : z = A\alpha, \ \forall \alpha. \} \)
  Dimension of \( \mathcal{R}(A) = N. \)

- \( \mathcal{N}(A) \) is the null space of \( A \), i.e. \( \mathcal{N}(A) = \{ w : Aw = 0 \}. \)
  Dimension of \( \mathcal{N}(A) = M - N \)

Since we have an under-determined system of equations, minimizing the error \( \| y - Ax \|_2 \), as in LS, will result in zero error.
Under-determined System of Equations

\[ y = Ax + v, \]

where \( A \in \mathbb{R}^{N \times M}, x \in \mathbb{R}^M, \) and \( y \in \mathbb{R}^N. \)

Assumptions

- \( N \leq M \)
- Rank of \( A \) is \( N \), i.e. rows of \( A \) are linearly independent.

Fundamental subspaces

- \( \mathcal{R}(A) \) is the range space of \( A \), i.e. \( \mathcal{R}(A) = \{ z : z = A\alpha, \ \forall \alpha \}. \)
  Dimension of \( \mathcal{R}(A) = N. \)
- \( \mathcal{N}(A) \) is the null space of \( A \), i.e. \( \mathcal{N}(A) = \{ w : Aw = 0 \}. \)
  Dimension of \( \mathcal{N}(A) = M - N. \)

Since we have an under-determined system of equations, minimizing the error \( \| y - Ax \|_2 \), as in LS, will result in zero error.

If \( x_0 \) is a solution, then

\[ x_s = x_0 + z, \]

where \( z \in \mathcal{N}(A) \) is also a solution since \( Az = 0. \).
Regularization

There are infinite number of solutions, how to choose one?

Minimum 2-norm solution of $x$ is given by $x_{mn} = \text{arg min}_{y = Ax} \| x \|_2 = A^+y$, where $A^+$ is the Moore-Penrose inverse given by $A^+ = A^T(AA^T)^{-1}$.

To accommodate for noise, of 2-norm regularization is used $x_{REG} = \text{arg min}_{x} \left[ \| y - Ax \|_2^2 + \lambda \| x \|_2^2 \right] = A^T(AA^T + \lambda I)^{-1}y$. The parameter $\lambda$ provides a mechanism to trade off modeling error versus norm of the solution.

Problem: The minimum 2-norm solution is not sparse.
There are infinite number of solutions, how to choose one?

Minimum 2-norm solution of $\mathbf{x}$ is given by

$$
\mathbf{x}_{mn} = \arg\min_{\mathbf{y} = A\mathbf{x}} \|\mathbf{x}\|_2 = A^+ \mathbf{y},
$$

$A^+$ is the Moore-Penrose inverse given by $A^+ = A^T(AA^T)^{-1}$.

To accommodate for noise, of 2-norm regularization is used

$$
\mathbf{x}_{\text{REG}} = \arg\min_{\mathbf{x}} [\|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2] = A^T(A^2 + \lambda I)^{-1} \mathbf{y}.
$$

$\lambda$ provides a mechanism to trade off modeling error versus norm of the solution.

Problem:
The minimum 2-norm solution is not sparse.
There are infinite number of solutions, how to choose one? Minimum 2-norm solution of $x$ is given by

$$x_{mn} = \arg \min_{\|x\|_2} \|x\|_2 = A^+ y,$$

$A^+$ is the Moore-Penrose inverse given by $A^+ = A^T (A A^T)^{-1}$. To accommodate for noise, of 2-norm regularization is used

$$x_{REG} = \arg \min_x [\|y - Ax\|_2^2 + \lambda \|x\|_2^2] = A^T (A A^T + \lambda I)^{-1} y$$

$\lambda$ provides a mechanism to trade off modeling error versus norm of the solution.
Regularization

There are infinite number of solutions, how to choose one?

Minimum 2-norm solution of $x$ is given by

$$x_{mn} = \arg \min_{y=Ax} \|x\|_2 = A^+ y,$$

$A^+$ is the Moore-Penrose inverse given by $A^+ = A^T (A A^T)^{-1}$.

To accommodate for noise, of 2-norm regularization is used

$$x_{REG} = \arg \min_x [\|y - A x\|_2^2 + \lambda \|x\|_2^2] = A^T (A A^T + \lambda I)^{-1} y$$

$\lambda$ provides a mechanism to trade off modeling error versus norm of the solution.

**Problem:** The minimum 2-norm solution is not sparse.
Uniqueness of Sparse Solutions

Two questions now arise:

To deal with uniqueness in the general case, an important useful property of matrices is the "spark":

Definition
The \( \text{spark} (A) \) is the smallest number of columns in \( A \) that are linearly dependent.

From the above definition, it is easy to see that for a \( N \times M \) full row rank matrix \( A \) where \( N < M \), we have \( 2 \leq \text{spark} (A) \leq N + 1 \).

Example
If \( A = \begin{bmatrix} I \end{bmatrix} \), then \( \text{spark} (A) = 2 \), and \( \text{rank} (A) = N \).

Example
If \( A = \begin{bmatrix} a_{i,j} \end{bmatrix} \), \( a_{i,j} \sim \mathcal{N}(0,1) \), then with high probability \( \text{spark} (A) = N + 1 \).
Two questions now arise:

- When is the sparsest solution unique?
Two questions now arise:

- When is the sparsest solution unique?
- How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the "spark":

**Definition**

The spark \((A)\) is the smallest number of columns in \(A\) that are linearly dependent.

From the above definition, it is easy to see that for a \(N \times M\) full row rank matrix \(A\) where \(N < M\), we have \(2 \leq \text{spark}(A) \leq N + 1\).

**Example**

If \(A = \begin{bmatrix} I & I \end{bmatrix}\), then \(\text{spark}(A) = 2\), and \(\text{rank}(A) = N\).

**Example**

If \(A = \begin{bmatrix} a_{i,j} \end{bmatrix}\), \(a_{i,j} \sim N(0, 1)\), then with high probability \(\text{spark}(A) = N + 1\).
Uniqueness of Sparse Solutions

Two questions now arise:

- When is the sparsest solution unique?
- How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the ”spark”:

**Definition**

The **spark** \( (A) \) is the smallest number of columns in \( A \) that are linearly dependent.

From the above definition, it is easy to see that for a \( N \times M \) full row rank matrix \( A \) where \( N < M \), we have \( 2 \leq \text{spark}(A) \leq N + 1 \).

**Example**

If \( A = [I I] \), then \( \text{spark}(A) = 2 \), and \( \text{rank}(A) = N \).

**Example**

If \( A = [a_{ij}] \), \( a_{ij} \sim N(0, 1) \), then with high probability \( \text{spark}(A) = N + 1 \).
Uniqueness of Sparse Solutions

Two questions now arise:

▶ When is the sparsest solution unique?
▶ How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the ”spark”:

Definition
The $\text{spark}(\mathbf{A})$ is the smallest number of columns in $\mathbf{A}$ that are linearly dependent.

**Definition**
The $\text{spark}(\mathbf{A})$ is the smallest number of columns in $\mathbf{A}$ that are linearly dependent.
Uniqueness of Sparse Solutions

Two questions now arise:

- When is the sparsest solution unique?
- How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the ”spark”:

**Definition**
The $\text{spark}(\mathbf{A})$ is the smallest number of columns in $\mathbf{A}$ that are linearly dependent.

From the above definition, it is easy to see that for a $N \times M$ full row rank matrix $\mathbf{A}$ where $N < M$, we have $2 \leq \text{spark}(\mathbf{A}) \leq N + 1.$
Two questions now arise:

▶ When is the sparsest solution unique?
▶ How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the "spark":

**Definition**
The \( \text{spark}(A) \) is the smallest number of columns in \( A \) that are linearly dependent.

From the above definition, it is easy to see that for a \( N \times M \) full row rank matrix \( A \) where \( N < M \), we have \( 2 \leq \text{spark}(A) \leq N + 1 \).

**Example**
If \( A = [I | I] \),
Uniqueness of Sparse Solutions

Two questions now arise:

▶ When is the sparsest solution unique?
▶ How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the "spark":

Definition
The $spark(A)$ is the smallest number of columns in $A$ that are linearly dependent.

From the above definition, it is easy to see that for a $N \times M$ full row rank matrix $A$ where $N < M$, we have $2 \leq spark(A) \leq N + 1$.

Example
If $A = [l \quad I]$, then $spark(A) = 2$, and $rank(A) = N$. 
Uniqueness of Sparse Solutions

Two questions now arise:

▶ When is the sparsest solution unique?
▶ How do we verify a solution as the sparsest one?

To deal with uniqueness in the general case, an important useful property of matrices is the "spark":

**Definition**
The spark$(A)$ is the smallest number of columns in $A$ that are linearly dependent.

From the above definition, it is easy to see that for a $N \times M$ full row rank matrix $A$ where $N < M$, we have $2 \leq $spark$(A) \leq N + 1$.

**Example**
If $A = [I \quad I]$, then $spark(A) = 2$, and $rank(A) = N$.

**Example**
If $A = [a_{i,j}]$, $a_{i,j} \sim N(0, 1)$, then with high probability $spark(A) = N + 1$. 
The following theorems give us sufficient conditions to determine uniqueness of this solution.

\begin{itemize}
  \item Theorem 1
  \begin{enumerate}
    \item For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark} (A)$, then $x$ is the sparsest possible solution and it is unique.
  \end{enumerate}
\end{itemize}

Proof. Let $\|x\|_0$ be a solution to $y = Ax$ which satisfies $\|x\|_0 < \frac{1}{2} \text{spark} (A)$. Let $z$ be any other solution, such that $Ax = Az = y$ and $A(x - z) = 0$. By properties of the spark of a matrix and the triangular inequality, we have $\|x\|_0 + \|z\|_0 \geq \|x - z\|_0 \geq \text{spark} (A)$. The above inequality implies that if $\|x\|_0 < \frac{1}{2} \text{spark} (A)$, then $\|y\|_0 > \frac{1}{2} \text{spark} (A)$. This means that any solution other than $x$ has more non-zero entries than $x$ and $x$ is the sparsest possible solution and it is unique.

\footnote{Theorem 2.4. from Elad, Michael. Sparse and redundant representations: from theory to applications in signal and image processing. Springer, 2010.}
The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem**

1. For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark}(A)$, then $x$ is the sparsest possible solution and it is unique.

---

\footnote{Theorem 2.4. from Elad, Michael. Sparse and redundant representations: from theory to applications in signal and image processing. Springer, 2010.}
The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem**

1. For an arbitrary matrix $\mathbf{A}$, if there exists a solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$ such that $\|\mathbf{x}\|_0 < \frac{1}{2} \text{spark} (\mathbf{A})$, then $\mathbf{x}$ is the sparsest possible solution and it is unique.

**Proof.**

Let $\|\mathbf{x}\|_0$ be a solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$ which satisfies $\|\mathbf{x}\|_0 < \frac{1}{2} \text{spark} (\mathbf{A})$.

\[\text{Proof.}\]

\[\text{Let } \|\mathbf{x}\|_0 \text{ be a solution to } \mathbf{y} = \mathbf{A}\mathbf{x} \text{ which satisfies } \|\mathbf{x}\|_0 < \frac{1}{2} \text{spark} (\mathbf{A}).\]

\[\text{The above inequality implies that if } \|\mathbf{x}\|_0 < \frac{1}{2} \text{spark} (\mathbf{A}), \text{ then } \|\mathbf{y}\|_0 > \frac{1}{2} \text{spark} (\mathbf{A}).\]

This means that any solution other than $\mathbf{x}$ has more non-zero entries than $\mathbf{x}$ and $\mathbf{x}$ is the sparsest possible solution and it is unique.

\[\text{Proof.}\]

\[\text{Let } \|\mathbf{x}\|_0 \text{ be a solution to } \mathbf{y} = \mathbf{A}\mathbf{x} \text{ which satisfies } \|\mathbf{x}\|_0 < \frac{1}{2} \text{spark} (\mathbf{A}).\]

---

\[\text{1Theorem 2.4. from Elad, Michael. Sparse and redundant representations: from theory to applications in signal and image processing. Springer, 2010.}\]
The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem**

1. For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark}(A)$, then $x$ is the sparsest possible solution and it is unique.

**Proof.**

Let $\|x\|_0$ be a solution to $y = Ax$ which satisfies $\|x\|_0 < \frac{1}{2} \text{spark}(A)$. Let $z$ be any other solution, such that $Ax = Az = y$ and $A(x - z) = 0$.

---

The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem 1.** For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark}(A)$, then $x$ is the sparsest possible solution and it is unique.

**Proof.**
Let $\|x\|_0$ be a solution to $y = Ax$ which satisfies $\|x\|_0 < \frac{1}{2} \text{spark}(A)$. Let $z$ be any other solution, such that $Ax = Az = y$ and $A(x - z) = 0$. By properties of the spark of a matrix and the triangular inequality, we have

$$\|x\|_0 + \|z\|_0 \geq \|x - z\|_0 \geq \text{spark}(A)$$

---

The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem**

1. For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark (} A \text{)}$, then $x$ is the sparsest possible solution and it is unique.

**Proof.**

Let $\|x\|_0$ be a solution to $y = Ax$ which satisfies $\|x\|_0 < \frac{1}{2} \text{spark (} A \text{)}$. Let $z$ be any other solution, such that $Ax = Az = y$ and $A(x - z) = 0$. By properties of the spark of a matrix and the triangular inequality, we have

$$\|x\|_0 + \|z\|_0 \geq \|x - z\|_0 \geq \text{spark (} A \text{)}$$

The above inequality implies that if $\|x\|_0 < \frac{1}{2} \text{spark (} A \text{)}$, then $\|y\|_0 > \frac{1}{2} \text{spark (} A \text{)}$.

---

The following theorems give us sufficient conditions to determine uniqueness of this solution.

**Theorem**

1. For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \text{spark}(A)$, then $x$ is the sparsest possible solution and it is unique.

**Proof.**

Let $\|x\|_0$ be a solution to $y = Ax$ which satisfies $\|x\|_0 < \frac{1}{2} \text{spark}(A)$. Let $z$ be any other solution, such that $Ax = Az = y$ and $A(x - z) = 0$. By properties of the spark of a matrix and the triangular inequality, we have

$$\|x\|_0 + \|z\|_0 \geq \|x - z\|_0 \geq \text{spark}(A)$$

The above inequality implies that if $\|x\|_0 < \frac{1}{2} \text{spark}(A)$, then $\|y\|_0 > \frac{1}{2} \text{spark}(A)$. This means that any solution other than $x$ has more non-zero entries than $x$ and $x$ is the sparsest possible solution and it is unique. \(\square\)

---

Mutual Coherence

Definition

Define mutual coherence \( \mu \) as:

\[
\mu(A) = \max_{i, j, i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}
\]  

(1)

Evaluating the mutual coherence of a given matrix \( A \) (using \( A^T A \)) is easy, while calculating the spark of \( A \) can be complicated. The following lemma and theorem relate uniqueness of a given solution to the mutual coherence of a matrix \( A \).

Lemma 2

For an arbitrary matrix \( A \) we have

\[
\text{spark}(A) \geq 1 + \frac{1}{\mu(A)}^2
\]

Define mutual coherence $\mu$ as:

$$
\mu(A) = \max_{i,j, i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}
$$

Evaluating the mutual coherence of a given matrix $A$ (using $A^T A$) is easy, while calculating the spark of $A$ can be complicated. The following lemma and theorem relate uniqueness of a given solution to the mutual coherence of a matrix $A$.

Mutual Coherence

Definition
Define mutual coherence $\mu$ as:

$$
\mu(A) = \max_{i,j, i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}
$$

(1)

Evaluating the mutual coherence of a given matrix $A$ (using $A^T A$) is easy, while calculating the spark of $A$ can be complicated.

Mutual Coherence

Definition
Define mutual coherence $\mu$ as:

$$
\mu(A) = \max_{i,j,i \neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}
$$

(1)

Evaluating the mutual coherence of a given matrix $A$ (using $A^T A$) is easy, while calculating the spark of $A$ can be complicated. The following lemma and theorem relate uniqueness of a given solution to the mutual coherence of a matrix $A$.

---

Mutual Coherence

Definition
Define mutual coherence $\mu$ as:

$$\mu(A) = \max_{i,j,i\neq j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}$$  \hspace{1cm} (1)

Evaluating the mutual coherence of a given matrix $A$ (using $A^T A$) is easy, while calculating the spark of $A$ can be complicated. The following lemma and theorem relate uniqueness of a given solution to the mutual coherence of a matrix $A$.

Lemma
2 For an arbitrary matrix $A$ we have

$$\text{spark}(A) \geq 1 + \frac{1}{\mu(A)}$$

---

Theorem

3 For an arbitrary matrix $A$, if there exists a solution to $y = Ax$ such that $\|x\|_0 < \frac{1}{2} \left( 1 + \frac{1}{\mu(A)} \right)$, then $x$ is the sparsest possible solution and it is unique.

Proof.

The proof follows from above theorems.

---