Linear Inverse Problem

Qin Zhang

SAMSI/ CRSC Undergraduate Workshop, 2008

May 20, 2008

Outline

1. Forward Problem and Inverse Problem
2. Review of linear algebra
3. Ill-poseness of a linear inverse problem
4. Least Square solution for overdetermined system

Spring-mass system

Without external forces the displacement $y(t)$ is described by

$$m\ddot{y} + c\dot{y} + ky = 0$$

the solution is of the form

$$y(t) = Ae^{-\frac{c}{2m}t}\sin(\omega t + \phi)$$

where the constants A and ϕ are determined by the initial conditions, and $\omega = \sqrt{k/m}$.

Question: What if we don't know m, c or k in advance?

Spring-mass system

Without external forces the displacement $y(t)$ is described by

$$m\ddot{y} + c\dot{y} + ky = 0$$

the solution is of the form

$$y(t) = Ae^{-\frac{c}{2m}t}\sin(\omega t + \phi)$$

where the constants A and ϕ are determined by the initial conditions, and $\omega = \sqrt{k/m}$.

Question: What if we don't know m, c or k in advance?

Forward Problem vs Inverse problem

Consider the following model

$$y = F(x, b)$$

where b is the model parameter, x is the input variable, and y is the output variable.

- **Forward problem**:
  Given the parameter b, what is the value of y for x?
- **Inverse problem**:
  Having data (x, y), how to calculate or estimate the parameter b?

In some sense, we need to find the inverse F^{-1}

Forward Problem vs Inverse problem

Consider the following model

$$y = F(x, b)$$

where b is the model parameter, x is the input variable, and y is the output variable.

- **Forward problem**:
  Given the parameter b, what is the value of y for x?
- **Inverse problem**:
  Having data (x, y), how to calculate or estimate the parameter b?

In some sense, we need to find the inverse F^{-1}

Forward Problem vs Inverse problem

Consider the following model

$$y = F(x, b)$$

where b is the model parameter, x is the input variable, and y is the output variable.

- **Forward problem**:
  Given the parameter b, what is the value of y for x?
- **Inverse problem**:
  Having data (x, y), how to calculate or estimate the parameter b?

In some sense, we need to find the inverse F^{-1}

Forward Problem vs Inverse problem

Consider the following model

$$y = F(x, b)$$

where b is the model parameter, x is the input variable, and y is the output variable.

- **Forward problem**:
  Given the parameter b, what is the value of y for x?
- **Inverse problem**:
  Having data (x, y), how to calculate or estimate the parameter b?

In some sense, we need to find the inverse F^{-1}

Linear Inverse Problem

If the function F is a *linear* function in b, then given data (x, y) to solve for the model parameter b, is a *linear* inverse problem.

Example:

$$y = bx$$

$$y = b_0 + b_1 x + b_2 x^2$$

$$y = b_0 + b_1 \cos(x) + b_2 sin(x^2) + b_3 e^x$$

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots b_{n-1} x_{n-1} + b_n x_n$$

$$= \begin{pmatrix} 1 & x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Linear Inverse Problem

If the function F is a *linear* function in b, then given data (x, y) to solve for the model parameter b, is a *linear* inverse problem.

Example:

$$y = bx$$

$$y = b_0 + b_1 x + b_2 x^2$$

$$y = b_0 + b_1 \cos(x) + b_2 sin(x^2) + b_3 e^x$$

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots b_{n-1} x_{n-1} + b_n x_n$$

$$= \begin{pmatrix} 1 & x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Linear Inverse Problem

If the function F is a *linear* function in b, then given data (x, y) to solve for the model parameter b, is a *linear* inverse problem.

Example:

$$y = bx$$

$$y = b_0 + b_1 x + b_2 x^2$$

$$y = b_0 + b_1 \cos(x) + b_2 sin(x^2) + b_3 e^x$$

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots b_{n-1} x_{n-1} + b_n x_n$$

$$= \begin{pmatrix} 1 & x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Linear Inverse Problem

If the function F is a *linear* function in b, then given data (x, y) to solve for the model parameter b, is a *linear* inverse problem.

Example:

$$y = bx$$

$$y = b_0 + b_1 x + b_2 x^2$$

$$y = b_0 + b_1 \cos(x) + b_2 sin(x^2) + b_3 e^x$$

$$\begin{aligned} y &= b_0 + b_1 x_1 + b_2 x_2 + \ldots b_{n-1} x_{n-1} + b_n x_n \\ &= \begin{pmatrix} 1 & x_1 & \ldots & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} \end{aligned}$$

Multidimension

In general, b and y can be vectors.

Suppose there are n number of parameters $b_1, b_2, \ldots, b_n$, and m number of observation $y_1, y_2, \ldots, y_m$,

$$\begin{aligned}
y_1 &= x_{11}b_1 + x_{12}b_2 + \ldots x_{1n}b_n \\
y_2 &= x_{21}b_1 + x_{22}b_2 + \ldots x_{2n}b_n \\
\vdots & \qquad\qquad\qquad \vdots \\
y_m &= x_{m1}b_1 + x_{22}b_2 + \ldots x_{mn}b_n
\end{aligned}$$

Linear Model

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & & x_{2n} \\ \vdots & & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

In general, the linear model can be written as

$$y = Xb$$

where X is a $m \times n$ matrix, b is a $n \times 1$ vector, and y is a $m \times 1$ vector.

Example

- $$b_1 + 2b_2 = 3$$

- $$\begin{aligned} b_1 + 2b_2 &= 3 \\ 2b_1 - 3b_2 &= -1 \end{aligned}$$

- $$\begin{aligned} b_1 + 2b_2 &= 3 \\ 2b_1 - 3b_2 &= -1 \\ 3b_1 - 2b_2 &= 4 \end{aligned}$$

Linear System of Equations

$$y = Xb$$

where X is a $m \times n$ matrix, b is a $n \times 1$ vector, and y is a $m \times 1$ vector.

- If $m < n$, **underdetermined system**
  usually have infinite many solution.
- If $m = n$, X is a square matrix. If X is nonsingular, then

$$b = X^{-1}y$$

- If $m > n$, **overdetermined system**
  usually have no exact solution.
  Question: Can we find some solution in a proper sense?

Linear System of Equations

$$y = Xb$$

where X is a $m \times n$ matrix, b is a $n \times 1$ vector, and y is a $m \times 1$ vector.

- If $m < n$, **underdetermined system**
  usually have infinite many solution.
- If $m = n$, X is a square matrix. If X is nonsingular, then

$$b = X^{-1}y$$

- If $m > n$, **overdetermined system**
  usually have no exact solution.
  Question: Can we find some solution in a proper sense?

Linear System of Equations

$$y = Xb$$

where X is a $m \times n$ matrix, b is a $n \times 1$ vector, and y is a $m \times 1$ vector.

- If $m < n$, **underdetermined system**
  usually have infinite many solution.
- If $m = n$, X is a square matrix. If X is nonsingular, then

$$b = X^{-1}y$$

- If $m > n$, **overdetermined system**
  usually have no exact solution.
  Question: Can we find some solution in a proper sense?

Transpose of a matrix

The *transpose* of a matrix X is denoted by X^T, whose rows are the columns of X.

For example,

$$X = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 0 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 0 \end{pmatrix}$$

It has the following properties:

$$(A^T)^T = A, \quad (A+B)^T = A^T + B^T, \quad (AB)^T = B^T A^T$$

Transpose of a matrix

The *transpose* of a matrix X is denoted by X^T, whose rows are the columns of X.

For example,

$$X = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 0 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 0 \end{pmatrix}$$

It has the following properties:

$$(A^T)^T = A, \quad (A+B)^T = A^T + B^T, \quad (AB)^T = B^T A^T$$

Transpose of a matrix

The *transpose* of a matrix X is denoted by X^T, whose rows are the columns of X.

For example,

$$X = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 0 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 0 \end{pmatrix}$$

It has the following properties:

$$(A^T)^T = A, \quad (A+B)^T = A^T + B^T, \quad (AB)^T = B^T A^T$$

Inner product

$$(1,2,3)\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = b_1 + 2b_2 + 3b_3$$

In general,

$$(a_1,\ldots,a_n)\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i$$

$$< a, b >= a^T b$$

Inverse of a matrix

An $n \times n$ square matrix X is *nonsingular* (invertible) if there exists a matrix X^{-1} such that $XX^{-1} = X^{-1}X = I$, where I is the identity matrix.

Theorem: X is nonsingular if and only if $det(X) \neq 0$.

If $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $det(X) = ad - bc$,

$$X^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

In Matlab, we can use the command $inv(X)$ to compute X^{-1}.

Solve linear system of equation

$$Xb = y \Rightarrow X^{-1}Xb = X^{-1}y \Rightarrow b = X^{-1}y$$

for example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

In Matlab, we have two ways to solve this problem:

- $inv(X) * y$
- $X \backslash y$

Solve linear system of equation

$$Xb = y \Rightarrow X^{-1}Xb = X^{-1}y \Rightarrow b = X^{-1}y$$

for example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

In Matlab, we have two ways to solve this problem:

- $inv(X) * y$
- $X \backslash y$

Solve linear system of equation

$$Xb = y \Rightarrow X^{-1}Xb = X^{-1}y \Rightarrow b = X^{-1}y$$

for example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 3 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

In Matlab, we have two ways to solve this problem:

- $inv(X) * y$
- $X \backslash y$

Vector norm

The vector norm is a positive valued function to measure the *length* of a vector.

In Euclidean space $\mathbb{R}^2$, the Euclidean norm of a vector $v = (x, y)$ is defined by

$$\|v\| = \sqrt{x^2 + y^2}$$

In general, the 2- norm of a $n \times 1$ vector $y \in \mathbb{R}^n$ is defined by

$$\|y\|_2 = (\sum_{i=1}^{n} y_i^2)^{\frac{1}{2}}$$

so

$$\|y\|_2^2 = \sum_{i=1}^{n} y_i^2 = y^T y =< y, y >$$

Vector norm

The vector norm is a positive valued function to measure the *length* of a vector.

In Euclidean space $\mathbb{R}^2$, the Euclidean norm of a vector $v = (x, y)$ is defined by

$$\|v\| = \sqrt{x^2 + y^2}$$

In general, the 2- norm of a $n \times 1$ vector $y \in \mathbb{R}^n$ is defined by

$$\|y\|_2 = (\sum_{i=1}^{n} y_i^2)^{\frac{1}{2}}$$

so

$$\|y\|_2^2 = \sum_{i=1}^{n} y_i^2 = y^T y = < y, y >$$

Matrix norm

The 1-norm of a $m \times n$ matrix $X \in \mathbb{R}^{m \times n}$ is defined by

$$\|X\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |X_{ij}|$$
$$= \text{the largest absolute column sum}$$

Similarly, we can define the 2-norm or ∞-norm of a matrix X.

In Matlab, we use the command

$$\text{norm}(X, 1)$$

to compute the matrix 1-norm.

ill-conditioned system

Example:

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.067 \end{pmatrix} \to \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.06{\color{red}6} \end{pmatrix} \to \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

You will notice major difference here! Why?

$$det(X) = 10^{-6}$$

A system is *ill-conditioned* if some small perturbation in the system causes a relatively large change in the exact solution.

ill-conditioned system

Example:

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.067 \end{pmatrix} \rightarrow \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.066 \end{pmatrix} \rightarrow \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

You will notice major difference here! Why?

$$det(X) = 10^{-6}$$

A system is *ill-conditioned* if some small perturbation in the system causes a relatively large change in the exact solution.

ill-conditioned system

Example:

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.067 \end{pmatrix} \rightarrow \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$\begin{pmatrix} 0.835 & 0.667 \\ 0.333 & 0.266 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0.168 \\ 0.066 \end{pmatrix} \rightarrow \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

You will notice major difference here! Why?

$$det(X) = 10^{-6}$$

A system is *ill-conditioned* if some small perturbation in the system causes a relatively large change in the exact solution.

Error estimate and condition number

Theorem:
Let b_e be the solution of $Xb = y$, and let b_c be the solution of $(X + \delta X)b = y + \delta y$, then we have

$$\frac{\|b_e - b_c\|}{\|b_e\|} \leq \|X\|\|X^{-1}\|(\frac{\|\delta X\|}{\|X\|} + \frac{\|\delta y\|}{\|y\|})$$

We see the relative error depends on $\|X\|\|X^{-1}\|$, which is defined as the **condition number** of the nonsingular matrix X.

$$\text{cond}(X) = \|X\|\|X^{-1}\|$$

A matrix X is

- ill-conditioned: If $\text{cond}(X)$ is large.
- well-conditioned: If $\text{cond}(X)$ is small.

Error estimate and condition number

Theorem:
Let b_e be the solution of $Xb = y$, and let b_c be the solution of $(X + \delta X)b = y + \delta y$, then we have

$$\frac{\|b_e - b_c\|}{\|b_e\|} \leq \|X\|\|X^{-1}\|(\frac{\|\delta X\|}{\|X\|} + \frac{\|\delta y\|}{\|y\|})$$

We see the relative error depends on $\|X\|\|X^{-1}\|$, which is defined as the **condition number** of the nonsingular matrix X.

$$\text{cond}(X) = \|X\|\|X^{-1}\|$$

A matrix X is

- ill-conditioned: If $\text{cond}(X)$ is large.
- well-conditioned: If $\text{cond}(X)$ is small.

Example for overdetermined linear system

Consider an overdetermined linear system $y = Xb$, which has more columns than rows ($m > n$), for example

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

NO exact solution!!!

Least Square Solution

For an overdetermined linear system $y = Xb$ with $X \in \mathbb{R}^{m \times n}$ and $m > n$, we want to minimize the sum of squared errors (SSE),

$$\min_b \|y - Xb\|_2^2$$

The solution $\hat{b}$ is called the least square solution of the system.

In the previous example, we want to minimize

$$\text{SSE} = (2 - b_1 - 2b_2)^2 + (3 - b_2)^2 + (4 - 3b_1)^2$$

Vector- partial derivative

Result: Let $\vec{a}$ and $\vec{b}$ be $n \times 1$ vectors, then

$$\frac{\partial a^T b}{\partial b} = a, \quad \frac{\partial b^T a}{\partial b} = a$$

Proof: note that $a^T b = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n = \sum_{i=1}^{n} a_i b_i$. So

$$(\frac{\partial a^T b}{\partial b})_j = \frac{\partial a^T b}{\partial b_j} = a_j$$

Normal equation

$$
\begin{aligned}
0 = \frac{\partial \|y - Xb\|_2^2}{\partial b} &= \frac{\partial (y - Xb)^T (y - Xb)}{\partial b} \\
&= -X^T(y - Xb) + [(y - Xb)^T(-X)]^T \\
&= -X^T(y - Xb) - X^T(y - Xb) \\
&= -2X^T(y - Xb) \\
&= -2(X^T y - X^T X b)
\end{aligned}
$$

Normal Equation:

$$X^T X b = X^T y$$

If $(X^T X)^{-1}$ exists, then

$$\hat{b} = (X^T X)^{-1} X^T y$$

Simple Linear Regression

Consider the model

$$y_i = bx_i, \quad i = 1, \ldots n$$

To write it in matrix form

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} b$$

then

$$X^T X = \sum_{i=1}^{n} x_i^2, \quad X^T Y = \sum_{i=1}^{n} x_i y_i$$

So the least square solution $\hat{b}$ is

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

In summary, given the observations $y = (y_1, y_2, \ldots, y_m)^T$, and assume the model is

$$y = Xb + \epsilon$$

where ϵ is the observation error. The parameter that gives the smallest error between the model and observation is

$$\hat{b} := \min_b \|y - Xb\|_2^2 = (X^T X)^{-1} X^T y$$

Remark:

- the estimation $\hat{b}$ depends on whether $cond(X^T X)$ is large or not
- How do we know if the parameters estimate are good or bad? Uncertainty in the estimation is related to the *measurement error*. (to be addressed in an upcoming tutorial titled "Statistical View of Linear Least Squares" by Dr. Michael D. Porter)