Least Squares Problems

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x\in\mathbb{R}^m} \|b - Ax\|_2.$$

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x \in \mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x \in \mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x\in\mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x \in \mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.
- Machine Learning.

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x\in\mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.
- Machine Learning.

Direct solution methods:

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x\in\mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.
- Machine Learning.

Direct solution methods:

- Normal Equations Method.

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x \in \mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.
- Machine Learning.

Direct solution methods:

- Normal Equations Method.
- QR Decomposition Method.

Definition

Consider the possibly overdetermined system of equations

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n \times m}$ and a vector $b \in \mathbb{R}^n$, $n \geq m$. It may not have an exact solution if $n > m$ or A is square but singular. The **Least Squares Problem** associated with this system is about finding $x_0 \in \mathbb{R}^m$ such that

$$\|b - Ax_0\|_2 = \min_{x \in \mathbb{R}^m} \|b - Ax\|_2.$$

Applications:

- Polynomial curve fitting.
- Making predictions from existing data.
- Machine Learning.

Direct solution methods:

- Normal Equations Method.
- QR Decomposition Method.
- Singular Value Decomposition Method.

Geometry of the Least Squares Problem

As $A \in \mathbb{R}^{n \times m}$, the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ given by $x \mapsto Ax$, for all $x \in \mathbb{R}^m$ has range and null spaces

$$R(A) = \underbrace{\{Ax : x \in \mathbb{R}^m\}}_{\text{also called Col (A)}} \subset \mathbb{R}^n \text{ and } N(A) = \{x \in \mathbb{R}^m : Ax = 0\} \subset \mathbb{R}^m.$$

Geometry of the Least Squares Problem

As $A \in \mathbb{R}^{n\times m}$, the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ given by $x \mapsto Ax$, for all $x \in \mathbb{R}^m$ has range and null spaces

$$R(A) = \underbrace{\{Ax : x \in \mathbb{R}^m\}}_{\text{also called Col (A)}} \subset \mathbb{R}^n \text{ and } N(A) = \{x \in \mathbb{R}^m : Ax = 0\} \subset \mathbb{R}^m.$$

The LSP is a two stage process involving

1. Find the best approximation to b from $R(A)$ i.e., find $y_0 \in R(A)$ such that $\|b - y_0\|_2 = \min_{y\in R(A)} \|b - y\|_2$.
2. Find $x_0 \in \mathbb{R}^m$ such that $Ax_0 = y_0$.

Geometry of the Least Squares Problem

As $A \in \mathbb{R}^{n \times m}$, the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ given by $x \mapsto Ax$, for all $x \in \mathbb{R}^m$ has range and null spaces

$$R(A) = \underbrace{\{Ax : x \in \mathbb{R}^m\}}_{\text{also called Col (A)}} \subset \mathbb{R}^n \text{ and } N(A) = \{x \in \mathbb{R}^m : Ax = 0\} \subset \mathbb{R}^m.$$

The LSP is a two stage process involving

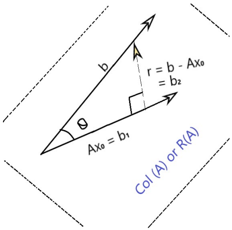
1. Find the best approximation to b from $R(A)$ i.e., find $y_0 \in R(A)$ such that $\|b - y_0\|_2 = \min_{y \in R(A)} \|b - y\|_2$.
2. Find $x_0 \in \mathbb{R}^m$ such that $Ax_0 = y_0$.

Since

$$\mathbb{R}^n = R(A) \oplus R(A)^{\perp},$$

for $b \in R^n$, there exists unique $b_1 \in R(A)$, $b_2 \in R(A)^{\perp}$, such that $b = b_1 + b_2$.

Geometry of the Least Squares Problem



Therefore for all $x \in \mathbb{R}^m$, $b - Ax = \underbrace{b_1 - Ax}_{\in R(A)} + \underbrace{b_2}_{\in R(A)^\perp}$ and

$$\begin{aligned}
\|b - Ax\|_2^2 &= \langle b_1 - Ax + b_2, b_1 - Ax + b_2 \rangle \\
&= \|b_1 - Ax\|_2^2 + \|b_2\|_2^2 \qquad (\text{as } \langle b_1 - Ax, b_2 \rangle = 0 \text{ for all } x \in \mathbb{R}^m) \\
&\geq \|b_2\|_2^2 \\
&= \| \underbrace{b - Ax_0}_{:= r (= b_2)} \|_2^2
\end{aligned}$$

where $x_0 \in \mathbb{R}^m$ such that $Ax_0 = b_1$. Hence,

$$\min_{x \in \mathbb{R}^m} \|b - Ax\|_2^2 \geq \|b - Ax_0\|_2^2 \geq \min_{x \in \mathbb{R}^m} \|b - Ax\|_2^2 \Rightarrow \min_{x \in \mathbb{R}^m} \|b - Ax\|_2 = \|b - Ax_0\|_2.$$

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^\perp$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^{\perp}$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Hence the solution of an LSP problem always exists!

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^\perp$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Hence the solution of an LSP problem always exists!

How many solutions are there?

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^\perp$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Hence the solution of an LSP problem always exists!

How many solutions are there?

The solution is unique if and only if $\operatorname{rank} A = m$.

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^{\perp}$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Hence the solution of an LSP problem always exists!

How many solutions are there?

The solution is unique if and only if $\operatorname{rank} A = m$.
If $\operatorname{rank} A < m$, there are infinitely many solutions!

Geometry of the Least Squares Problem

Hence $x_0 \in \mathbb{R}^m$ is a solution of the least squares problem and

$$r := b - Ax_0 = b_2 \in R(A)^{\perp}$$

is the residual vector.

Note that if P be the orthogonal projection of $\mathbb{R}^n$ onto $R(A)$, then $Pb = b_1$. Therefore in summary,

> *the orthogonal projection b_1 of b onto the column space of A is the nearest vector to b in the column space of A with respect to $\|\cdot\|_2$ and the solution of the LSP associated with the system is a column vector x_0 of scalars which produce the linear combination of the columns of A to form b_1.*

Hence the solution of an LSP problem always exists!

How many solutions are there?

The solution is unique if and only if $\operatorname{rank} A = m$.
If $\operatorname{rank} A < m$, there are infinitely many solutions!

Exercise: Prove the above statements!

Normal Equations Method

How to get x_0?

Since $R(A)^{\perp} = N(A^T)$, where $N(A^T)$ is the null space of the linear map $y \mapsto A^T y$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, $A^T r = 0$.

Normal Equations Method

How to get x_0?

Since $R(A)^\perp = N(A^T)$, where $N(A^T)$ is the null space of the linear map $y \mapsto A^T y$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, $A^T r = 0$. Hence x_0 is a solution of

$$A^T A x = A^T b \tag{2}$$

The systems of equations (2) are called the *Normal Equations* of the Least Squares Problem associated with $Ax = b$.

Normal Equations Method

How to get x_0?

Since $R(A)^{\perp} = N(A^T)$, where $N(A^T)$ is the null space of the linear map $y \mapsto A^T y$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, $A^T r = 0$. Hence x_0 is a solution of

$$A^T A x = A^T b \tag{2}$$

The systems of equations (2) are called the *Normal Equations* of the Least Squares Problem associated with $Ax = b$.

Exercise: Let $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$, where $n \geq m$. Prove that the following statements are equivalent.

1. $\operatorname{rank} A = m$.
2. $A^T A$ is nonsingular.
3. $A^T A$ is positive definite.

Normal Equations Method

How to get x_0?

Since $R(A)^{\perp} = N(A^T)$, where $N(A^T)$ is the null space of the linear map $y \mapsto A^T y$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, $A^T r = 0$. Hence x_0 is a solution of

$$A^T A x = A^T b \tag{2}$$

The systems of equations (2) are called the *Normal Equations* of the Least Squares Problem associated with $Ax = b$.

Exercise: Let $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$, where $n \geq m$. Prove that the following statements are equivalent.

1. $\operatorname{rank} A = m$.
2. $A^T A$ is nonsingular.
3. $A^T A$ is positive definite.

Pseudocode for solving the Least Squares Problem via Normal Equations when $\operatorname{rank} A = m$:

1. Form the Normal Equations. $(2nm^2 + O(nm)$ flops)
2. Solve them via the Cholesky Method. $(m^3/3 + O(m^2)$ flops)
3. Compute two norm squared of the residual vector. $(2nm + 3n$ flops)