Orthogonal Subspaces

D. H. Luecking

MASC

01 April 2024

The following is sometimes called the Fredholm Alternative.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^{\perp}$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^{\perp}$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

Suppose $\mathbf{b}$ is not in $\mathcal{R}(A)$ so that $A\mathbf{x} = \mathbf{b}$ has no solution.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

Suppose $\mathbf{b}$ is not in $\mathcal{R}(A)$ so that $A\mathbf{x} = \mathbf{b}$ has no solution. In many problems, when this happens we would like to find a vector $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

Suppose $\mathbf{b}$ is not in $\mathcal{R}(A)$ so that $A\mathbf{x} = \mathbf{b}$ has no solution. In many problems, when this happens we would like to find a vector $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

We can write $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

Suppose $\mathbf{b}$ is not in $\mathcal{R}(A)$ so that $A\mathbf{x} = \mathbf{b}$ has no solution. In many problems, when this happens we would like to find a vector $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

We can write $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.

It turns out that solving $A\mathbf{x} = \mathbf{c}$ gives us the $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

The following is sometimes called the Fredholm Alternative.

Theorem

If A is an $n \times k$ matrix and $\mathbf{b} \in \mathbb{R}^n$, then either there is a vector $\mathbf{x} \in \mathbb{R}^k$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbb{R}^n$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The first condition is the statement that $\mathbf{b} \in \mathcal{R}(A)$. The second condition is that $\mathbf{b}$ is not in $\mathcal{N}(A^T)^\perp$. Since $\mathcal{R}(A) = \mathcal{N}(A^T)^\perp$, this theorem is equivalent to "either $\mathbf{b}$ is in $\mathcal{R}(A)$ or $\mathbf{b}$ is not in $\mathcal{R}(A)$", which is obvious.

Suppose $\mathbf{b}$ is not in $\mathcal{R}(A)$ so that $A\mathbf{x} = \mathbf{b}$ has no solution. In many problems, when this happens we would like to find a vector $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

We can write $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.

It turns out that solving $A\mathbf{x} = \mathbf{c}$ gives us the $\hat{\mathbf{x}}$ that makes $\|A\hat{\mathbf{x}} - \mathbf{b}\|$ as small as possible.

Because $\mathbf{d} \in \mathcal{N}(A^T)$, if we apply A^T to $\mathbf{b} = \mathbf{c} + \mathbf{d}$ we get

$$A^T(\mathbf{b}) = A^T(\mathbf{c} + \mathbf{d}) = A^T\mathbf{c}$$

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^T A\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$.

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^T A\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$. This actually has a solution, as any solution of $A\mathbf{x} = \mathbf{c}$ is also a solution of $A^T A\mathbf{x} = A^T\mathbf{c}$.

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^T A\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$. This actually has a solution, as any solution of $A\mathbf{x} = \mathbf{c}$ is also a solution of $A^T A\mathbf{x} = A^T\mathbf{c}$.

The reverse is also true: Any solution of $A^T A\mathbf{x} = A^T\mathbf{c}$ is a solution of $A\mathbf{x} = \mathbf{c}$.

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^T A\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$. This actually has a solution, as any solution of $A\mathbf{x} = \mathbf{c}$ is also a solution of $A^T A\mathbf{x} = A^T\mathbf{c}$.

The reverse is also true: Any solution of $A^T A\mathbf{x} = A^T\mathbf{c}$ is a solution of $A\mathbf{x} = \mathbf{c}$.

To see this, we can rewrite $A^T A\mathbf{x} = A^T\mathbf{c}$ as $A^T(A\mathbf{x} - \mathbf{c}) = \mathbf{0}$.

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^TA\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$. This actually has a solution, as any solution of $A\mathbf{x} = \mathbf{c}$ is also a solution of $A^TA\mathbf{x} = A^T\mathbf{c}$.

The reverse is also true: Any solution of $A^TA\mathbf{x} = A^T\mathbf{c}$ is a solution of $A\mathbf{x} = \mathbf{c}$.

To see this, we can rewrite $A^TA\mathbf{x} = A^T\mathbf{c}$ as $A^T(A\mathbf{x} - \mathbf{c}) = \mathbf{0}$. This means $A\mathbf{x} - \mathbf{c} \in \mathcal{N}(A^T)$.

So, if we multiply the equation $A\mathbf{x} = \mathbf{b}$ (which has no solution) by A^T we get $A^T A\mathbf{x} = A^T\mathbf{b} = A^T\mathbf{c}$. This actually has a solution, as any solution of $A\mathbf{x} = \mathbf{c}$ is also a solution of $A^T A\mathbf{x} = A^T\mathbf{c}$.

The reverse is also true: Any solution of $A^T A\mathbf{x} = A^T\mathbf{c}$ is a solution of $A\mathbf{x} = \mathbf{c}$.

To see this, we can rewrite $A^T A\mathbf{x} = A^T\mathbf{c}$ as $A^T(A\mathbf{x} - \mathbf{c}) = \mathbf{0}$. This means $A\mathbf{x} - \mathbf{c} \in \mathcal{N}(A^T)$. But, since also $A\mathbf{x} - \mathbf{c} \in \mathcal{R}(A)$ and this is orthogonal to $\mathcal{N}(A^T)$, we get $A\mathbf{x} - \mathbf{c} = \mathbf{0}$.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned}
(\mathbf{b} - \mathbf{v})^T(\mathbf{b} - \mathbf{v}) &= (\mathbf{d} + \mathbf{c} - \mathbf{v})^T(\mathbf{d} + \mathbf{c} - \mathbf{v}) \\
&= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c} - \mathbf{v}) + (\mathbf{c} - \mathbf{v})^T\mathbf{d} + (\mathbf{c} - \mathbf{v})^T(\mathbf{c} - \mathbf{v}) \\
\|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c} - \mathbf{v}\|^2
\end{aligned}$$

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned}
(\mathbf{b} - \mathbf{v})^T(\mathbf{b} - \mathbf{v}) &= (\mathbf{d} + \mathbf{c} - \mathbf{v})^T(\mathbf{d} + \mathbf{c} - \mathbf{v}) \\
&= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c} - \mathbf{v}) + (\mathbf{c} - \mathbf{v})^T\mathbf{d} + (\mathbf{c} - \mathbf{v})^T(\mathbf{c} - \mathbf{v}) \\
\|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c} - \mathbf{v}\|^2
\end{aligned}$$

This means that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned}
(\mathbf{b} - \mathbf{v})^T(\mathbf{b} - \mathbf{v}) &= (\mathbf{d} + \mathbf{c} - \mathbf{v})^T(\mathbf{d} + \mathbf{c} - \mathbf{v}) \\
&= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c} - \mathbf{v}) + (\mathbf{c} - \mathbf{v})^T\mathbf{d} + (\mathbf{c} - \mathbf{v})^T(\mathbf{c} - \mathbf{v}) \\
\|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c} - \mathbf{v}\|^2
\end{aligned}$$

This means that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$. If we solve $A\mathbf{x} = \mathbf{c}$ and get $\hat{\mathbf{x}}$, we get the best "solution" to $A\mathbf{x} = \mathbf{b}$ in the sense that it as close as possible.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned}
(\mathbf{b} - \mathbf{v})^T(\mathbf{b} - \mathbf{v}) &= (\mathbf{d} + \mathbf{c} - \mathbf{v})^T(\mathbf{d} + \mathbf{c} - \mathbf{v}) \\
&= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c} - \mathbf{v}) + (\mathbf{c} - \mathbf{v})^T\mathbf{d} + (\mathbf{c} - \mathbf{v})^T(\mathbf{c} - \mathbf{v}) \\
\|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c} - \mathbf{v}\|^2
\end{aligned}$$

This means that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$. If we solve $A\mathbf{x} = \mathbf{c}$ and get $\hat{\mathbf{x}}$, we get the best "solution" to $A\mathbf{x} = \mathbf{b}$ in the sense that it as close as possible. Finally, we get all the solutions of $A\mathbf{x} = \mathbf{c}$ by solving $A^TA\mathbf{x} = A^T\mathbf{b}$.

Example: Suppose we perform an experiment where we measure the output y for various inputs x.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned} (\mathbf{b} - \mathbf{v})^T(\mathbf{b} - \mathbf{v}) &= (\mathbf{d} + \mathbf{c} - \mathbf{v})^T(\mathbf{d} + \mathbf{c} - \mathbf{v}) \\ &= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c} - \mathbf{v}) + (\mathbf{c} - \mathbf{v})^T\mathbf{d} + (\mathbf{c} - \mathbf{v})^T(\mathbf{c} - \mathbf{v}) \\ \|\mathbf{b} - \mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c} - \mathbf{v}\|^2 \end{aligned}$$

This means that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$. If we solve $A\mathbf{x} = \mathbf{c}$ and get $\hat{\mathbf{x}}$, we get the best "solution" to $A\mathbf{x} = \mathbf{b}$ in the sense that it as close as possible. Finally, we get all the solutions of $A\mathbf{x} = \mathbf{c}$ by solving $A^TA\mathbf{x} = A^T\mathbf{b}$.

Example: Suppose we perform an experiment where we measure the output y for various inputs x. That is we get a table of data of the form $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$.

Recall that we have written $\mathbf{b} = \mathbf{c} + \mathbf{d}$ where $\mathbf{c} \in \mathcal{R}(A)$ and $\mathbf{d} \perp \mathcal{R}(A)$. We will now see that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$ and the distance from $\mathbf{b}$ to $\mathbf{c}$ is $\|\mathbf{d}\|$.

Suppose we take any vector $\mathbf{v} = A\mathbf{x} \in \mathcal{R}(A)$ and consider $\|\mathbf{b} - \mathbf{v}\|^2$. This equals

$$\begin{aligned}(\mathbf{b}-\mathbf{v})^T(\mathbf{b}-\mathbf{v}) &= (\mathbf{d}+\mathbf{c}-\mathbf{v})^T(\mathbf{d}+\mathbf{c}-\mathbf{v}) \\ &= \mathbf{d}^T\mathbf{d} + \mathbf{d}^T(\mathbf{c}-\mathbf{v}) + (\mathbf{c}-\mathbf{v})^T\mathbf{d} + (\mathbf{c}-\mathbf{v})^T(\mathbf{c}-\mathbf{v}) \\ \|\mathbf{b}-\mathbf{v}\|^2 &= \|\mathbf{d}\|^2 + \|\mathbf{c}-\mathbf{v}\|^2\end{aligned}$$

This means that the closest vector to $\mathbf{b}$ in $\mathcal{R}(A)$ is $\mathbf{c}$. If we solve $A\mathbf{x} = \mathbf{c}$ and get $\hat{\mathbf{x}}$, we get the best "solution" to $A\mathbf{x} = \mathbf{b}$ in the sense that it as close as possible. Finally, we get all the solutions of $A\mathbf{x} = \mathbf{c}$ by solving $A^T A\mathbf{x} = A^T\mathbf{b}$.

Example: Suppose we perform an experiment where we measure the output y for various inputs x. That is we get a table of data of the form $\{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$. Then we plot these points and try to find the best curve of some sort that matches this data.

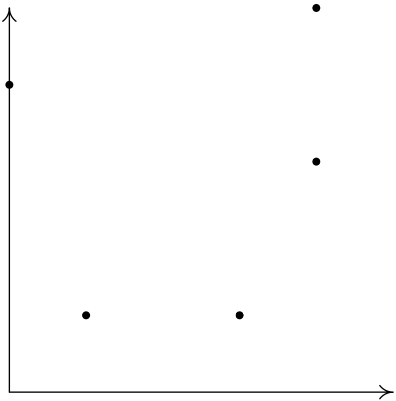
Lets suppose we get the data, listed and plotted below

Lets suppose we get the data, listed and plotted below

The data: $(0,4),(1,1),(3,1),(4,3),(4,5)$

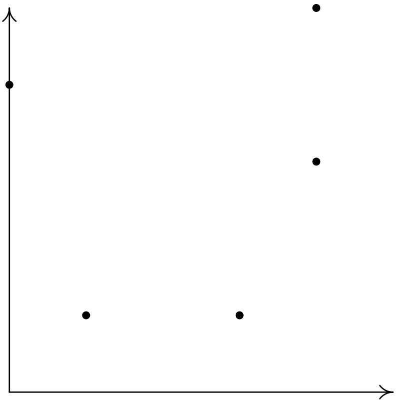
Lets suppose we get the data, listed and plotted below

The data: $(0,4),(1,1),(3,1),(4,3),(4,5)$



Lets suppose we get the data, listed and plotted below

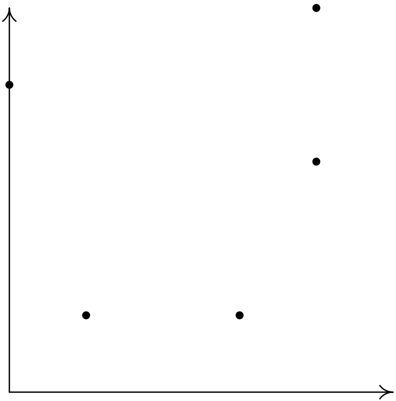
The data: $(0, 4), (1, 1), (3, 1), (4, 3), (4, 5)$



This looks vaguely like a parabola, so we might conjecture that (apart from small random deviations) the relationship between x and y has the form $y = ax^2 + bx + c$.

Lets suppose we get the data, listed and plotted below

The data: $(0, 4), (1, 1), (3, 1), (4, 3), (4, 5)$



This looks vaguely like a parabola, so we might conjecture that (apart from small random deviations) the relationship between x and y has the form $y = ax^2 + bx + c$.

The data can be used to determine what a, b, c must be. Ideally, we want the equations to $ax_i^2 + bx_i + c = y_i$ to hold for each data point.

We cannot expect there to be values of a, b, c that make this true for every data point, so we look for values that make it as close as possible.

We cannot expect there to be values of a, b, c that make this true for every data point, so we look for values that make it as close as possible.

The equations for a, b, c are

$$\begin{aligned} c &= 4 \\ 1a + 1b + c &= 2 \\ 9a + 3b + c &= 1 \\ 16a + 4b + c &= 3 \\ 16a + 4b + c &= 5 \end{aligned}$$

We cannot expect there to be values of a, b, c that make this true for every data point, so we look for values that make it as close as possible.

The equations for a, b, c are

$$\begin{aligned} c &= 4 \\ 1a + 1b + c &= 2 \\ 9a + 3b + c &= 1 \\ 16a + 4b + c &= 3 \\ 16a + 4b + c &= 5 \end{aligned}$$

or $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \\ 16 & 4 & 1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ 3 \\ 5 \end{pmatrix},$$

By the previous argument, we get the vector $\mathbf{x}$ that makes $A\mathbf{x}$ closest to $\mathbf{b}$ by solving $A^T A\mathbf{x} = A^T\mathbf{b}$.

So we compute

$$A^T A = \begin{pmatrix} 594 & 156 & 42 \\ 156 & 42 & 12 \\ 42 & 12 & 5 \end{pmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{pmatrix} 139 \\ 37 \\ 15 \end{pmatrix}$$

and row-reduce

$$\left(\begin{array}{ccc|c} 594 & 156 & 42 & 139 \\ 156 & 42 & 12 & 37 \\ 42 & 12 & 5 & 15 \end{array}\right).$$

to get (approximately) $a = 1.35$, $b = -5.5$ and $c = 4.65$.

By the previous argument, we get the vector $\mathbf{x}$ that makes $A\mathbf{x}$ closest to $\mathbf{b}$ by solving $A^T A\mathbf{x} = A^T\mathbf{b}$.

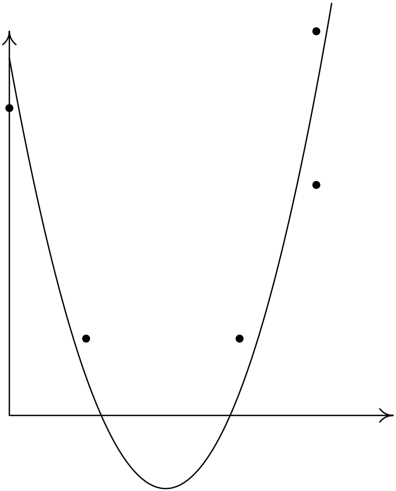
So we compute

$$A^T A = \begin{pmatrix} 594 & 156 & 42 \\ 156 & 42 & 12 \\ 42 & 12 & 5 \end{pmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{pmatrix} 139 \\ 37 \\ 15 \end{pmatrix}$$

and row-reduce

$$\left(\begin{array}{ccc|c} 594 & 156 & 42 & 139 \\ 156 & 42 & 12 & 37 \\ 42 & 12 & 5 & 15 \end{array}\right).$$

to get (approximately) $a = 1.35$, $b = -5.5$ and $c = 4.65$. I have graphed the quadratic $1.35x^2 - 5.5x + 4.65$ on the same plot as the data points on the next slide.



A more straightforward example:

A more straightforward example:

Find the least squares solution of the following system:

$$\begin{aligned} x_1 + x_2 &= 3 \\ -2x_1 + 3x_2 &= 1 \\ 2x_1 - x_2 &= 2 \end{aligned}$$

A more straightforward example:

Find the least squares solution of the following system:

$$\begin{aligned} x_1 + \ \ x_2 &= 3 \\ -2x_1 + 3x_2 &= 1 \\ 2x_1 - \ \ x_2 &= 2 \end{aligned}$$

Which is the same as

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

A more straightforward example:

Find the least squares solution of the following system:

$$\begin{aligned} x_1 + x_2 &= 3 \\ -2x_1 + 3x_2 &= 1 \\ 2x_1 - x_2 &= 2 \end{aligned}$$

Which is the same as

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Multiplying by the transpose gives

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

A more straightforward example:

Find the least squares solution of the following system:

$$\begin{aligned} x_1 + \quad x_2 &= 3 \\ -2x_1 + 3x_2 &= 1 \\ 2x_1 - \quad x_2 &= 2 \end{aligned}$$

Which is the same as

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Multiplying by the transpose gives

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

This has a solution $\hat{\mathbf{x}} = \begin{pmatrix} 83/50 \\ 71/50 \end{pmatrix} = \begin{pmatrix} 1.66 \\ 1.42 \end{pmatrix}$.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$. The least squares solution, is a vector $\hat{\mathbf{x}}$ that gives the residual vector the smallest possible norm.

Using the method of least squares to get vectors orthogonal to a subspace.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$)

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

But every subspace S of $\mathbb{R}^n$ is the column space of some matrix: take any basis of S (or any set of vectors whose span is S) and make them the columns of a matrix A. Then $S = \mathcal{R}(A)$.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

But every subspace S of $\mathbb{R}^n$ is the column space of some matrix: take any basis of S (or any set of vectors whose span is S) and make them the columns of a matrix A. Then $S = \mathcal{R}(A)$.

If we apply this to A^T in place of A we can split $\mathbf{b}$ into $\mathbf{u}$ in $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$ and $\mathbf{v}$ in $\mathcal{R}(A^T)^\perp = \mathcal{N}(A)$.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. Often we are required to find one or both of these. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

But every subspace S of $\mathbb{R}^n$ is the column space of some matrix: take any basis of S (or any set of vectors whose span is S) and make them the columns of a matrix A. Then $S = \mathcal{R}(A)$.

If we apply this to A^T in place of A we can split $\mathbf{b}$ into $\mathbf{u}$ in $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$ and $\mathbf{v}$ in $\mathcal{R}(A^T)^\perp = \mathcal{N}(A)$.

That is, we solve $AA^T\mathbf{x} = A\mathbf{b}$ to get $\hat{x}$ and then $\mathbf{u} = A^T\hat{x}$ is in $\mathcal{N}(A)^\perp$ and $\mathbf{v} = \mathbf{b} - A^T\hat{x}$ is in $\mathcal{N}(A)$.

Example: Let S be the span of $(1,1,2,0)^T$ and $(0,1,2,-2)^T$ and let $\mathbf{b}=(1,1,1,1)^T$. Find the vectors $\mathbf{u}\in S$ and $\mathbf{v}\in S^\perp$ such that $\mathbf{b}=\mathbf{u}+\mathbf{v}$.

Example: Let S be the span of $(1,1,2,0)^T$ and $(0,1,2,-2)^T$ and let $\mathbf{b} = (1,1,1,1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Now S is the column space of the 4×2 matrix A below and we need the least squares solution of

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Example: Let S be the span of $(1,1,2,0)^T$ and $(0,1,2,-2)^T$ and let $\mathbf{b} = (1,1,1,1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Now S is the column space of the 4×2 matrix A below and we need the least squares solution of

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$.

Example: Let S be the span of $(1, 1, 2, 0)^T$ and $(0, 1, 2, -2)^T$ and let $\mathbf{b} = (1, 1, 1, 1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Now S is the column space of the 4×2 matrix A below and we need the least squares solution of

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then $\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$

Example: Let S be the span of $(1,1,2,0)^T$ and $(0,1,2,-2)^T$ and let $\mathbf{b} = (1,1,1,1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Now S is the column space of the 4×2 matrix A below and we need the least squares solution of

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and
$\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.