Math 407: Linear Optimization

Lecture 15: The Linear Least Squares Problem

Math Dept, University of Washington

December 5, 2016

Linear Least Squares

A linear least squares problem is one of the form

$$\mathcal{LLS} \quad \underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|_2^2\,,$$

where

$$A\in\mathbb{R}^{m\times n},\ \ b\in\mathbb{R}^m, \quad \text{and} \quad \|y\|_2^2 := y_1^2+y_2^2+\cdots+y_m^2\ .$$

This problem formulation is usually credited to Legendre and Gauss who made careful studies of the method around 1800. But others applied the basic approach in an ad hoc manner in the previous 50 years to observational data and, in particular, to studying the motion of the planets

Applications

Celestial Mechanics

Polynomial fitting

Approximation by basis functions

Linear regression and maximum likelihood

System identification

Kalman filtering/smoothing

Tikhonov regularization

Regularization, ridge regression

Total least squares

Optimality Conditions

We say that $\overline{x}$ is a solution to $\mathcal{LLS}$ if

$$\|A\overline{x} - b\|_2 \leq \|Ax - b\|_2 \quad \forall\, x \in \mathbb{R}^n.$$

Optimality Conditions

We say that $\overline{x}$ is a solution to $\mathcal{LLS}$ if

$$\|A\overline{x} - b\|_2 \leq \|Ax - b\|_2 \quad \forall\ x \in \mathbb{R}^n.$$

Basic identity:

$$\|u + v\|_2^2 = (u+v)^T(u+v) = u^T u + 2u^T v + v^T v = \|u\|_2^2 + 2u^T v + \|v\|_2^2\,.$$

Optimality Conditions

We say that $\overline{x}$ is a solution to $\mathcal{LLS}$ if

$$\|A\overline{x} - b\|_2 \leq \|Ax - b\|_2 \quad \forall\ x \in \mathbb{R}^n.$$

Basic identity:

$$\|u + v\|_2^2 = (u+v)^T(u+v) = u^T u + 2u^T v + v^T v = \|u\|_2^2 + 2u^T v + \|v\|_2^2 .$$

If $x \in \mathbb{R}^n$, then with $u = A(\overline{x} - x)$ and $v = Ax - b$ we obtain

$$\begin{aligned}
\|A\overline{x} - b\|_2^2 &= \|A(\overline{x} - x) + (Ax - b)\|_2^2 \\
&= \|A(\overline{x} - x)\|_2^2 + 2(A(\overline{x} - x))^T(Ax - b) + \|Ax - b\|_2^2 \\
&\geq \|A(\overline{x} - x)\|_2^2 + 2(A(\overline{x} - x))^T(Ax - b) + \|A\overline{x} - b\|_2^2 .
\end{aligned}$$

Optimality Conditions

Canceling $\|A\bar{x} - b\|_2^2$ from both sides gives

$$\begin{aligned} 0 &\geq \|A(\bar{x} - x)\|_2^2 + 2(A(\bar{x} - x))^T(Ax - b) \\ &= 2(A(\bar{x} - x))^T(A\bar{x} - b) - \|A(\bar{x} - x)\|_2^2 . \end{aligned}$$

Setting $x = \bar{x} + tw$ for $(t, w) \in \mathbb{R} \times \mathbb{R}^n$, gives

$$\frac{t^2}{2}\|Aw\|_2^2 \geq tw^T A^T(A\bar{x} - b) \quad \forall\ t \in \mathbb{R} \quad \text{and} \quad w \in \mathbb{R}^n.$$

Dividing by $t > 0$, gives

$$\frac{t}{2}\|Aw\|_2^2 \geq w^T A^T(A\bar{x} - b) \quad \forall\ t \in \mathbb{R} \setminus \{0\} \quad \text{and} \quad w \in \mathbb{R}^n.$$

Sending t down to zero gives

$$0 \geq w^T A^T(A\bar{x} - b) \quad \forall\ w \in \mathbb{R}^n,$$

which implies that $A^T(A\bar{x} - b) = 0$ (why?), or equivalently,

$$A^T A\bar{x} = A^T b. \tag{1}$$

Linear Least Squares and the Normal Equations

Theorem:
The vector $\overline{x}$ satisfies

$$\|A\overline{x} - b\|_2 \leq \|Ax - b\|_2 \quad \forall\, x \in \mathbb{R}^n,$$

if and only if $A^T A\overline{x} = A^T b$.

Linear Least Squares and the Normal Equations

Theorem:
The vector $\overline{x}$ satisfies

$$\|A\overline{x} - b\|_2 \leq \|Ax - b\|_2 \quad \forall\, x \in \mathbb{R}^n,$$

if and only if $A^T A\overline{x} = A^T b$.

Proof: We just showed that if $\overline{x}$ is a solution to $\mathcal{LLS}$, then the normal equations are satisfied. We now show the reverse implication.
Assume that $A^T(A\overline{x} - b) = 0$. Then, for all $x \in \mathbb{R}^n$,

$$\begin{aligned}
\|Ax - b\|_2^2 &= \|(Ax - A\overline{x}) + (A\overline{x} - b)\|_2^2 \\
&= \|A(x - \overline{x})\|_2^2 + 2(A(x - \overline{x}))^T(A\overline{x} - b) + \|A\overline{x} - b\|_2^2 \\
&\geq 2(x - \overline{x})^T A^T (A\overline{x} - b) + \|A\overline{x} - b\|_2^2 \\
&= \|A\overline{x} - b\|_2^2
\end{aligned}$$

or equivalently, $\overline{x}$ solves $\mathcal{LLS}$.

$\text{Ran}(A^T A) = \text{Ran}(A^T)$

Lemma:

For every matrix $A \in \mathbb{R}^{m\times n}$ we have

$$\text{Null}(A^T A) = \text{Null}(A) \quad \text{and} \quad \text{Ran}(A^T A) = \text{Ran}(A^T)\ .$$

$\mathrm{Ran}(A^T A) = \mathrm{Ran}(A^T)$

Lemma:
For every matrix $A \in \mathbb{R}^{m \times n}$ we have

$$\mathrm{Null}(A^T A) = \mathrm{Null}(A) \quad \text{and} \quad \mathrm{Ran}(A^T A) = \mathrm{Ran}(A^T)\ .$$

Proof: Note that if $x \in \mathrm{Null}(A)$, then $Ax = 0$ and so $A^T Ax = 0$, that is, $x \in \mathrm{Null}(A^T A)$. Therefore, $\mathrm{Null}(A) \subset \mathrm{Null}(A^T A)$.

$\text{Ran}(A^T A) = \text{Ran}(A^T)$

Lemma:
For every matrix $A \in \mathbb{R}^{m \times n}$ we have

$$\text{Null}(A^T A) = \text{Null}(A) \quad \text{and} \quad \text{Ran}(A^T A) = \text{Ran}(A^T) \ .$$

Proof: Note that if $x \in \text{Null}(A)$, then $Ax = 0$ and so $A^T Ax = 0$, that is, $x \in \text{Null}(A^T A)$. Therefore, $\text{Null}(A) \subset \text{Null}(A^T A)$.

Conversely, if $x \in \text{Null}(A^T A)$, then

$$A^T Ax = 0 \Longrightarrow x^T A^T Ax = 0 \Longrightarrow (Ax)^T (Ax) = 0 \Longrightarrow \|Ax\|_2^2 = 0 \Longrightarrow Ax = 0,$$

or equivalently, $x \in \text{Null}(A)$. Therefore, $\text{Null}(A^T A) \subset \text{Null}(A)$, and so $\text{Null}(A^T A) = \text{Null}(A)$.

$\text{Ran}(A^T A) = \text{Ran}(A^T)$

Lemma:
For every matrix $A \in \mathbb{R}^{m \times n}$ we have

$$\text{Null}(A^T A) = \text{Null}(A) \quad \text{and} \quad \text{Ran}(A^T A) = \text{Ran}(A^T) \ .$$

Proof: Note that if $x \in \text{Null}(A)$, then $Ax = 0$ and so $A^T Ax = 0$, that is, $x \in \text{Null}(A^T A)$. Therefore, $\text{Null}(A) \subset \text{Null}(A^T A)$.

Conversely, if $x \in \text{Null}(A^T A)$, then
$$A^T Ax = 0 \Longrightarrow x^T A^T Ax = 0 \Longrightarrow (Ax)^T (Ax) = 0 \Longrightarrow \|Ax\|_2^2 = 0 \Longrightarrow Ax = 0,$$
or equivalently, $x \in \text{Null}(A)$. Therefore, $\text{Null}(A^T A) \subset \text{Null}(A)$, and so $\text{Null}(A^T A) = \text{Null}(A)$.

Since $\text{Null}(A^T A) = \text{Null}(A)$, the Fundamental Theorem of the Alternative tells us that
$$\text{Ran}(A^T A) = \text{Ran}((A^T A)^T) = \text{Null}(A^T A)^\perp = \text{Null}(A)^\perp = \text{Ran}(A^T),$$
which proves the lemma.

Existence and Uniqueness of Solutions to $\mathcal{LLS}$

Theorem:
Consider the linear least squares problem $\mathcal{LLS}$.

1. A solution to the normal equations $A^T Ax = A^T b$ always exists.
2. A solution to $\mathcal{LLS}$ always exists.
3. The linear least squares problem $\mathcal{LLS}$ has a unique solution if and only if $\mathrm{Null}(A) = \{0\}$ in which case $(A^T A)^{-1}$ exists and the unique solution is given by $\overline{x} = (A^T A)^{-1} A^T b$.
4. If $\mathrm{Ran}(A) = \mathbb{R}^m$, then $(AA^T)^{-1}$ exists and $\overline{x} = A^T (AA^T)^{-1} b$ solves $\mathcal{LLS}$, indeed, $A\overline{x} = b$.

Existence and Uniqueness of Solutions to $\mathcal{LLS}$

Theorem:
Consider the linear least squares problem $\mathcal{LLS}$.

1. A solution to the normal equations $A^TAx = A^Tb$ always exists.
2. A solution to $\mathcal{LLS}$ always exists.
3. The linear least squares problem $\mathcal{LLS}$ has a unique solution if and only if $\text{Null}(A) = \{0\}$ in which case $(A^TA)^{-1}$ exists and the unique solution is given by $\bar{x} = (A^TA)^{-1}A^Tb$.
4. If $\text{Ran}(A) = \mathbb{R}^m$, then $(AA^T)^{-1}$ exists and $\bar{x} = A^T(AA^T)^{-1}b$ solves $\mathcal{LLS}$, indeed, $A\bar{x} = b$.

Proof: $(1 \iff 2)$ Since $\text{Ran}(A^TA) = \text{Ran}(A^T)$.

Existence and Uniqueness of Solutions to $\mathcal{LLS}$

Theorem:
Consider the linear least squares problem $\mathcal{LLS}$.

1. A solution to the normal equations $A^T Ax = A^T b$ always exists.
2. A solution to $\mathcal{LLS}$ always exists.
3. The linear least squares problem $\mathcal{LLS}$ has a unique solution if and only if $\mathrm{Null}(A) = \{0\}$ in which case $(A^T A)^{-1}$ exists and the unique solution is given by $\overline{x} = (A^T A)^{-1} A^T b$.
4. If $\mathrm{Ran}(A) = \mathbb{R}^m$, then $(AA^T)^{-1}$ exists and $\overline{x} = A^T (AA^T)^{-1} b$ solves $\mathcal{LLS}$, indeed, $A\overline{x} = b$.

Proof: $(1 \iff 2)$ Since $\mathrm{Ran}(A^T A) = \mathrm{Ran}(A^T)$.
(3) Normal equations have a unique solution $\iff$ $(A^T A)^{-1}$ exists $\iff$ $\mathrm{Null}(A^T A) = \{0\}$ $\iff$ $\mathrm{Null}(A) = \{0\}$.

Existence and Uniqueness of Solutions to $\mathcal{LLS}$

Theorem:
Consider the linear least squares problem $\mathcal{LLS}$.

1. A solution to the normal equations $A^TAx = A^Tb$ always exists.
2. A solution to $\mathcal{LLS}$ always exists.
3. The linear least squares problem $\mathcal{LLS}$ has a unique solution if and only if $\mathrm{Null}(A) = \{0\}$ in which case $(A^TA)^{-1}$ exists and the unique solution is given by $\overline{x} = (A^TA)^{-1}A^Tb$.
4. If $\mathrm{Ran}(A) = \mathbb{R}^m$, then $(AA^T)^{-1}$ exists and $\overline{x} = A^T(AA^T)^{-1}b$ solves $\mathcal{LLS}$, indeed, $A\overline{x} = b$.

Proof: ($1 \iff 2$) Since $\mathrm{Ran}(A^TA) = \mathrm{Ran}(A^T)$.
(3) Normal equations have a unique solution $\iff$ $(A^TA)^{-1}$ exists $\iff$ $\mathrm{Null}(A^TA) = \{0\}$ $\iff$ $\mathrm{Null}(A) = \{0\}$.
(4) Since $\mathrm{Ran}(A) = \mathrm{Nul}\left(A^T\right)^{\perp}$.

Distance to a Subspace

Let $S \subset \mathbb{R}^m$ be a subspace and suppose $b \in \mathbb{R}^m$ is not in S.
Find the point $\bar{z} \in S$ such that

$$\|\bar{z} - b\|_2 \leq \|z - b\|_2 \qquad \forall\ z \in S,$$

or equivalently, solve

$$\mathcal{D} \quad \min_{z \in S} \frac{1}{2}\|z - b\|_2^2 \ .$$

Distance to a Subspace

Let $S \subset \mathbb{R}^m$ be a subspace and suppose $b \in \mathbb{R}^m$ is not in S.
Find the point $\bar{z} \in S$ such that

$$\|\bar{z} - b\|_2 \leq \|z - b\|_2 \qquad \forall\ z \in S,$$

or equivalently, solve

$$\mathcal{D} \quad \min_{z \in S} \tfrac{1}{2} \|z - b\|_2^2 \ .$$

Least Squares Connection:
Suppose $S = \mathrm{Ran}(A)$.
Then $\bar{z} \in \mathbb{R}^m$ solves $\mathcal{D}$ if and only if there is an $\bar{x} \in \mathbb{R}^n$ with $\bar{z} = A\bar{x}$ such that $\bar{x}$ solves $\mathcal{LLS}$.

Projections and Complementary Subspaces

- $P \in \mathbb{R}^{m\times m}$ is said to be a *projection* if and only if $P^2 = P$.

We say that P is a projection onto the subspace $S = \text{Ran}(P)$.

Projections and Complementary Subspaces

- $P \in \mathbb{R}^{m \times m}$ is said to be a *projection* if and only if $P^2 = P$.

We say that P is a projection onto the subspace $S = \text{Ran}(P)$.

- If $x \in \text{Ran}(P)$, there is $w \in \mathbb{R}^m$ s.t. $x = Pw$, so

$Px = P(Pw) = P^2w = Pw = x$.

Projections and Complementary Subspaces

- $P \in \mathbb{R}^{m\times m}$ is said to be a *projection* if and only if $P^2 = P$.

We say that P is a projection onto the subspace $S = \text{Ran}(P)$.

- If $x \in \text{Ran}(P)$, there is $w \in \mathbb{R}^m$ s.t. $x = Pw$, so

$Px = P(Pw) = P^2 w = Pw = x$.

- If P is a projection, then

$$(I-P)^2 = I - P - P + P^2 = I - P,$$

and so $(I - P)$ is also a projection.

Projections and Complementary Subspaces

- $P \in \mathbb{R}^{m\times m}$ is said to be a *projection* if and only if $P^2 = P$.

We say that P is a projection onto the subspace $S = \text{Ran}(P)$.

- If $x \in \text{Ran}(P)$, there is $w \in \mathbb{R}^m$ s.t. $x = Pw$, so

$Px = P(Pw) = P^2 w = Pw = x$.

- If P is a projection, then

$$(I-P)^2 = I - P - P + P^2 = I - P,$$

and so $(I-P)$ is also a projection.

- Since for all $w \in \mathbb{R}^m$,

$$w = Pw + (I-P)w,$$

we have

$$\mathbb{R}^m = \text{Ran}(P) + \text{Ran}(I-P).$$

We say that the subspaces $\text{Ran}(P)$ and $\text{Ran}(I-P)$ are *complementary subspaces*, i.e., $\text{Ran}(P) \cap \text{Ran}(I-P) = \{0\}$ and $\text{Ran}(P) + \text{Ran}(I-P) = \mathbb{R}^m$.

Projections and Complementary Subspaces

- $P \in \mathbb{R}^{m\times m}$ is said to be a *projection* if and only if $P^2 = P$. We say that P is a projection onto the subspace $S = \text{Ran}(P)$.
- If $x \in \text{Ran}(P)$, there is $w \in \mathbb{R}^m$ s.t. $x = Pw$, so $Px = P(Pw) = P^2w = Pw = x$.
- If P is a projection, then

$$(I - P)^2 = I - P - P + P^2 = I - P,$$

and so $(I - P)$ is also a projection.
- Since for all $w \in \mathbb{R}^m$,

$$w = Pw + (I - P)w,$$

we have

$$\mathbb{R}^m = \text{Ran}(P) + \text{Ran}(I - P).$$

We say that the subspaces $\text{Ran}(P)$ and $\text{Ran}(I - P)$ are *complementary subspaces*, i.e., $\text{Ran}(P) \cap \text{Ran}(I - P) = \{0\}$ and $\text{Ran}(P) + \text{Ran}(I - P) = \mathbb{R}^m$.
- Conversely, given complementary subspaces S_1 and S_2 ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$) there is a projection P such that $S_1 = \text{Ran}(P)$ and $S_2 = \text{Ran}(I - P)$.

Orthogonal Projections

- The subspace orthogonal to the subspace $S \subset \mathbb{R}^m$ is

$$S^{\perp} := \left\{ y \,\middle|\, y^T x = 0 \ \forall \ x \in S \right\}.$$

S and $S^{\perp}$ are complementary ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$).

Orthogonal Projections

• The subspace orthogonal to the subspace $S \subset \mathbb{R}^m$ is

$$S^\perp := \left\{ y \,\middle|\, y^T x = 0 \;\forall\; x \in S \right\}.$$

S and $S^\perp$ are complementary ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$).

• Let P_s be the projection onto S for which $(I - P_s)$ is the projection onto $S^\perp$. We call P_s the orthogonal projection onto S. Then

$$0 = ((I - P_s)y)^T (P_s w) = y^T [(I - P_s)^T P_s] w \qquad \forall\; y, w \in \mathbb{R}^m$$

Orthogonal Projections

- The subspace orthogonal to the subspace $S \subset \mathbb{R}^m$ is

$$S^\perp := \left\{ y \,\middle|\, y^T x = 0 \;\forall\; x \in S \right\}.$$

S and $S^\perp$ are complementary ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$).

- Let P_s be the projection onto S for which $(I - P_s)$ is the projection onto $S^\perp$. We call P_s the orthogonal projection onto S. Then

$$0 = ((I - P_s)y)^T (P_s w) = y^T [(I - P_s)^T P_s] w \qquad \forall\; y, w \in \mathbb{R}^m$$

But this holds if and only if

$$(I - P_s)^T P_s = 0 \;!$$

Orthogonal Projections

• The subspace orthogonal to the subspace $S \subset \mathbb{R}^m$ is

$$S^\perp := \left\{ y \,\middle|\, y^T x = 0 \ \forall\ x \in S \right\}.$$

S and $S^\perp$ are complementary ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$).

• Let P_s be the projection onto S for which $(I - P_s)$ is the projection onto $S^\perp$. We call P_s the orthogonal projection onto S. Then

$$0 = ((I - P_s)y)^T (P_s w) = y^T [(I - P_s)^T P_s] w \qquad \forall\ y, w \in \mathbb{R}^m$$

But this holds if and only if

$$(I - P_s)^T P_s = 0\ !$$

So $P_s = P_s^T P_s = (P_s^T P_s)^T = P_s^T$.

Orthogonal Projections

- The subspace orthogonal to the subspace $S \subset \mathbb{R}^m$ is

$$S^\perp := \left\{ y \,\middle|\, y^T x = 0 \; \forall \; x \in S \right\}.$$

S and $S^\perp$ are complementary ($S_1 \cap S_2 = \{0\}$ and $S_1 + S_2 = \mathbb{R}^m$).

- Let P_s be the projection onto S for which $(I - P_s)$ is the projection onto $S^\perp$. We call P_s the orthogonal projection onto S. Then

$$0 = ((I - P_s)y)^T (P_s w) = y^T [(I - P_s)^T P_s] w \qquad \forall \; y, w \in \mathbb{R}^m$$

But this holds if and only if

$$(I - P_s)^T P_s = 0 \; !$$

So $P_s = P_s^T P_s = (P_s^T P_s)^T = P_s^T$.

- Conversely, if $P = P^T$ and $P^2 = P$, then $(I - P)^T P = 0$. Therefore, a matrix P is an orthogonal projection if and only if $P^2 = P$ and $P = P^T$.

Orthogonal Projections and Orthonormal Bases

Let $Q \in \mathbb{R}^{n \times k}$ be a matrix whose columns form an orthonormal basis for the subspace S, so that $k = \dim S$.
Set $P = QQ^T$ and note that $Q^T Q = I_k$ the $k \times k$ identity matrix.
Then

$$P^2 = QQ^T QQ^T = QI_k Q^T = QQ^T = P \quad \text{and} \quad P^T = (QQ^T)^T = QQ^T = P,$$

so that $P = P_S$ the orthogonal projection onto S !

Orthogonal Projections and Orthonormal Bases

Let $Q \in \mathbb{R}^{n\times k}$ be a matrix whose columns form an orthonormal basis for the subspace S, so that $k = \dim S$.
Set $P = QQ^T$ and note that $Q^TQ = I_k$ the $k \times k$ identity matrix.
Then

$$P^2 = QQ^TQQ^T = QI_kQ^T = QQ^T = P \quad \text{and} \quad P^T = (QQ^T)^T = QQ^T = P,$$

so that $P = P_S$ the orthogonal projection onto S !

Lemma:

(1) The projection $P \in \mathbb{R}^{n\times n}$ is orthogonal if and only if $P = P^T$.

(2) If the columns of the matrix $Q \in \mathbb{R}^{n\times k}$ form an orthonormal basis for the subspace $S \subset \mathbb{R}^n$, then $P := QQ^T$ is the orthogonal projection onto S.

Orthogonal Projections and Distance to a Subspace

Theorem: Let $S \subset \mathbb{R}^m$ be a subspace and let $b \in \mathbb{R}^m \setminus S$. Then the unique solution to the least distance problem

$$\mathcal{D} \qquad \underset{z \in S}{\text{minimize}} \, \|z - b\|_2$$

is $\bar{z} := P_S b$, where P_S is the orthogonal projector onto S.

Orthogonal Projections and Distance to a Subspace

Theorem: Let $S \subset \mathbb{R}^m$ be a subspace and let $b \in \mathbb{R}^m \setminus S$. Then the unique solution to the least distance problem

$$\mathcal{D} \qquad \underset{z \in S}{\text{minimize}} \, \|z - b\|_2$$

is $\bar{z} := P_S b$, where P_S is the orthogonal projector onto S.

Proof: Let $\bar{z} := P_S b$. Then, for every $z \in S$,

$$\begin{aligned}
\|z - b\|_2^2 &= \|P_S z - P_S b - (I - P_S)b\|_2^2 \\
&= \|P_S(z - b) - (I - P_S)b\|_2^2 \\
&= \|P_S(z - b)\|_2^2 - 2(z - b)^T P_S^T (I - P_S)b + \|(I - P_S)b\|_2^2 \\
&= \|P_S(z - b)\|_2^2 + \|(I - P_S)b\|_2^2 \\
&\geq \|(P_S - I)b\|_2^2 = \|\bar{z} - b\|_2^2 .
\end{aligned}$$

Consequently, $\|z - b\|_2 \geq \|\bar{z} - b\|_2$ for all $z \in S$, that is, $\bar{z} = P_S b$ solves $\mathcal{D}$.

Orthogonal Projections and Distance to a Subspace

It remains only to establish the uniqueness of the solution. Suppose there is another $\hat{z} \in S$ achieving the minimal distance $\|\bar{z} - b\|_2$. Then

$$\begin{aligned}
\|\bar{z} - b\|_2^2 &= \|\hat{z} - b\|_2^2 \\
&= \|(\hat{z} - \bar{z}) + (\bar{z} - b)\|_2^2 \\
&= \|P_S(\hat{z} - \bar{z}) + (P_S b - b)\|_2^2 \\
&= \|P_S(\hat{z} - \bar{z}) - (I - P_S)b\|_2^2 \\
&= \|P_S(\hat{z} - \bar{z})\|_2^2 - 2\,\langle P_S(\hat{z} - \bar{z}),\, (I - P_S)b\rangle + \|(I - P_S)b\|_2^2 \\
&= \|\hat{z} - \bar{z}\|_2^2 - 2(\hat{z} - \bar{z})^T P_S(I - P_S)b + \|\bar{z} - b\|_2^2 \\
&= \|\hat{z} - \bar{z}\|_2^2 + \|\bar{z} - b\|_2^2\,,
\end{aligned}$$

so that $\|\hat{z} - \bar{z}\|_2 = 0$ and $\hat{z} = \bar{z}$.

The Orthogonal Projection onto Ran(A)

Reconsider the linear least-squares problem $\mathcal{LLS}$ as it relates to our new found knowledge about orthogonal projections and their relationship to least distance problems for subspaces.

Consider the case where $m >> n$ and $\text{Null}(A) = \{0\}$. In this case, the theorem above tells us that $\bar{x} = (A^T A)^{-1} A^T b$ solves $\mathcal{LLS}$, and $\bar{z} = P_S b$ solves $\mathcal{D}$ where P_S is the orthogonal projector onto $S = \text{Ran}(A)$. Hence,

$$P_{\text{Ran}(A)} b = \bar{z} = A\bar{x} = A(A^T A)^{-1} A^T b.$$

Since this is true for all possible choices of the vector b, we have

$$P_{\text{Ran}(A)} = A(A^T A)^{-1} A^T \quad !$$

The Orthogonal Projection onto Ran(A)

That is, the matrix $A(A^TA)^{-1}A^T$ is the orthogonal projector onto the range of A.

This can also be verified by showing that the matrix $M = A(A^TA)^{-1}A^T$ satisfies $M^2 = M$, $M^T = M$, and $\text{Ran}(M) = \text{Ran}(A)$.

Proposition: Let $A \in \mathbb{R}^{m\times n}$ with $m \leq n$ and $\text{Null}(A) = \{0\}$. Then the orthogonal projector onto $\text{Ran}(A)$ is given by

$$P_{\text{Ran}(A)} = A(A^TA)^{-1}A^T.$$

Minimal Norm Solutions to $Ax = b$

Let $A \in \mathbb{R}^{m \times n}$ and suppose that $m < n$. Then A is short and fat so A most likely has rank m, or equivalently, $\mathrm{Ran}(A) = \mathbb{R}^m$. So for the purposes of this discussion we assume that $\mathrm{rank}(A) = m$.

Minimal Norm Solutions to $Ax = b$

Let $A \in \mathbb{R}^{m \times n}$ and suppose that $m < n$. Then A is short and fat so A most likely has rank m, or equivalently, $\mathrm{Ran}(A) = \mathbb{R}^m$. So for the purposes of this discussion we assume that $\mathrm{rank}(A) = m$.

Since $m < n$, the set of solutions to $Ax = b$ will be infinite since the nullity of A is $n - m$. Indeed, if x^0 is any particular solution to $Ax = b$, then the set of solutions is given by

$$x^0 + \mathrm{Null}(A) := \left\{ x^0 + z \,\middle|\, z \in \mathrm{Null}(A) \right\}.$$

Minimal Norm Solutions to $Ax = b$

Let $A \in \mathbb{R}^{m\times n}$ and suppose that $m < n$. Then A is short and fat so A most likely has rank m, or equivalently, $\mathrm{Ran}(A) = \mathbb{R}^m$. So for the purposes of this discussion we assume that $\mathrm{rank}(A) = m$.

Since $m < n$, the set of solutions to $Ax = b$ will be infinite since the nullity of A is $n - m$. Indeed, if x^0 is any particular solution to $Ax = b$, then the set of solutions is given by

$$x^0 + \mathrm{Null}(A) := \left\{x^0 + z \,\middle|\, z \in \mathrm{Null}(A)\right\}.$$

In this setting, one might prefer the solution to the system having least norm:

$$\min_{z\in\mathrm{Null}(A)} \frac{1}{2}\left\|z + x^0\right\|_2^2 \ .$$

This problem is of the form $\mathcal{D}$ $\min\left\{\|z - b\|_2 \,|\, z \in S\right\}$ whose solution is $\bar{z} = P_S b$ when S is a subspace.
Consequently, the solution is given by $\bar{z} = -P_S x^0$ where P_s is now the orthogonal projection onto $S := \mathrm{Null}(A)$.

A Formula for $P_{\text{Null}(A)}$

Observe that $\quad P_{\text{Null}(A)} = P_{\text{Ran}(A^T)^\perp} = I - P_{\text{Ran}(A^T)}$.

A Formula for $P_{\text{Null}(A)}$

Observe that $\quad P_{\text{Null}(A)} = P_{\text{Ran}(A^T)^\perp} = I - P_{\text{Ran}(A^T)}$.

We have already shown that if $M \in \mathbb{R}^{p\times q}$ satisfies $\text{Null}(M) = \{0\}$, then $P_{\text{Ran}(M)} = M(M^TM)^{-1}M^T$.
If we take $M = A^T$, then our assumption that $\text{Ran}(A) = \mathbb{R}^m$ gives

$$\text{Null}(M) = \text{Null}(A^T) = \text{Ran}(A)^\perp = (\mathbb{R}^m)^\perp = \{0\}.$$

Hence,

$$P_{\text{Ran}(A^T)} = A^T(AA^T)^{-1}A \qquad \text{and}$$

$$P_{\text{Null}(A)} = P_{\text{Ran}(A^T)^\perp} = I - P_{\text{Ran}(A^T)} = I - A^T(AA^T)^{-1}A \ .$$

Solution to $\mathcal{D}$ $\min\{\|z - b\|_2 \mid z \in S\}$

Note that we can take $x^0 := A^T(AA^T)^{-1}b$ as our particular solution to $Ax = b$ in $\mathcal{D}$ since $Ax^0 = AA^T(AA^T)^{-1}b = b$.

Solution to $\mathcal{D}$ $\min \{ \|z - b\|_2 \mid z \in S \}$

Note that we can take $x^0 := A^T(AA^T)^{-1}b$ as our particular solution to $Ax = b$ in $\mathcal{D}$ since $Ax^0 = AA^T(AA^T)^{-1}b = b$.

Hence, our solution to $\mathcal{D}$ is

$$\begin{aligned}
\overline{x} &= x^0 + P_{\text{Null}(A)}(-x^0) \\
&= x^0 + (I - A^T(AA^T)^{-1}A)(-x^0) \\
&= A^T(AA^T)^{-1}Ax^0 \\
&= A^T(AA^T)^{-1}AA^T(AA^T)^{-1}b \\
&= A^T(AA^T)^{-1}b.
\end{aligned}$$

That is, $A^T(AA^T)^{-1}b$ is the least norm solution to $Ax = b$.

Solution to $\mathcal{D}$ $\min\{\|z-b\|_2 \,|\, z \in S\}$

Theorem:
Let $A \in \mathbb{R}^{m\times n}$ be such that $m \leq n$ and $\mathrm{Ran}(A) = \mathbb{R}^m$.

(1) The matrix AA^T is invertible.

(2) The orthogonal projection onto $\mathrm{Null}(A)$ is given by

$$P_{\mathrm{Null}(A)} = I - A^T(AA^T)^{-1}A \ .$$

(3) For every $b \in \mathbb{R}^m$, the system $Ax = b$ is consistent, and the least norm solution to this system is uniquely given by

$$\bar{x} = A^T(AA^T)^{-1}b \ .$$

Gram-Schmidt Orthogonalization and the QR-Factorization

The Gram-Schmidt orthogonalization process for a sequence of linearly independent vectors $a_1, \ldots, a_k \in \mathbb{R}^m$ (note that this implies that $n <\leq m$ (why?)) is as follows:

Define vectors $q_1, \ldots, q_n$ inductively, as follows: set

$$p_1 = a_1, \qquad q_1 = p_1/\|p_1\|,$$

$$p_j = a_j - \sum_{i=1}^{j-1} \langle a_j, q_j \rangle q_i \quad \text{and} \quad q_j = p_j/\|p_j\| \quad \text{for} \quad 2 \leq j \leq n.$$

For $1 \leq j \leq n$, $q_j \in \mathrm{Span}\{a_1, \ldots, a_j\}$, so $p_j \neq 0$ by the linear independence of $a_1, \ldots, a_j$.

An elementary induction argument shows that the q_j's form an orthonormal basis for $\mathrm{span}\,(a_1, \ldots, a_n)$.

Gram-Schmidt Orthogonalization and the QR-Factorization

If we now define

$$r_{jj} = \|p_j\| \neq 0 \quad \text{and} \quad r_{ij} = \langle a_j,\, q_i \rangle \quad \text{for} \quad 1 \leq i < j \leq n,$$

then

$$\begin{aligned}
a_1 &= r_{11}\, q_1, \\
a_2 &= r_{12}\, q_1 + r_{22}\, q_2, \\
a_3 &= r_{13}\, q_1 + r_{23}\, q_2 + r_{33}\, q_3, \\
&\vdots \\
a_n &= \sum_{i=1}^{n} r_{in}\, q_i .
\end{aligned}$$

Gram-Schmidt Orthogonalization and the QR-Factorization

Set

$$A = [a_1\ a_2\ \ \dots\ \ a_n] \in \mathbb{R}^{m\times n},\ R = [r_{ij}]\mathbb{R}^{n\times n},\ \text{and}\ Q = [q_1\ q_2\ \ \dots\ \ q_n] \in \mathbb{R}^{m\times n}$$

where $r_{ij} = 0,\ \ i > j$. Then

$$A = QR\ ,$$

where Q is unitary and R is an upper triangular $n \times n$ matrix.

In addition, R is invertible since the diagonal entries r_{jj} are non-zero.

This is called the *QR factorization* of the matrix A.

Full QR-Factorization

Suppose $A \in \mathbb{R}^{m \times n}$ with $m \geq n$. Then there exists a permutation matrix $P \in \mathbb{R}^{n \times n}$, a unitary matrix $Q \in \mathbb{R}^{m \times m}$, and an upper triangular matrix $R \in \mathbb{R}^{m \times n}$ such that $AP = QR$.

Let $Q_1 \in \mathbb{R}^{m \times n}$ denote the first n columns of Q, Q_2 the remaining $(m - n)$ columns of Q, and $R_1 \in \mathbb{R}^{n \times n}$ the first n rows of R, then

$$AP = QR = [Q_1 \ Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1. \tag{2}$$

Full QR-Factorization

Moreover, we have the following:

(a) We may choose R to have nonnegative diagonal entries.

(b) If A is of full rank, then we can choose R with positive diagonal entries, in which case we obtain the condensed factorization $A = Q_1 R_1$, where $R_1 \in \mathbb{R}^{n \times n}$ invertible and the columns of Q_1 forming an orthonormal basis for Ran(A).

(c) If $\mathrm{rank}\,(A) = k < n$, then

$$R_1 = \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix},$$

where R_{11} is a $k \times k$ invertible upper triangular matrix and $R_{12} \in \mathbb{R}^{k \times (n-k)}$. In particular, this implies that $AP = Q_{11}[R_{11} \;\; R_{12}]$, where Q_{11} are the first k columns of Q. In this case, the columns of Q_{11} form an orthonormal basis for the range of A.

Condenced QR-Factorization

Let $A \in \mathbb{R}^{m \times n}$ have rank $k \leq \min\{m, n\}$. Then there exist

$$\begin{array}{rl} Q \in \mathbb{R}^{m \times k} & \text{with orthonormal columns,} \\ R \in \mathbb{R}^{k \times n} & \text{full rank upper triangular, and} \\ P \in \mathbb{R}^{n \times n} & \text{a permutation matrix} \end{array}$$

such that

$$AP = QR.$$

In particular, the columns of the matrix Q form a basis for the range of A. Moreover, the matrix R can be written in the form

$$R = [R_1 \; R_2],$$

where $R_1 \in \mathbb{R}^{k \times k}$ is nonsingular.

Orthogonal Projections onto the Four Fundamental Subspaces

Let $A \in \mathbb{R}^{m \times n}$ have rank $k \leq \min\{m, n\}$. Let A and A^T have generalized QR factorizations

$$AP = Q[R_1 \ R_2] \quad \text{and} \quad A^T \widetilde{P} = \widetilde{Q}[\widetilde{R}_1 \ \widetilde{R}_2].$$

Since row rank equals column rank, $P \in \mathbb{R}^{n \times n}$ is a permutation matrix, $\widetilde{P} \in \mathbb{R}^{m \times m}$ is a permutation matrix, $Q \in \mathbb{R}^{m \times k}$ and $\widetilde{Q} \in \mathbb{R}^{n \times k}$ have orthonormal columns, $R_1, \widetilde{R}_1 \in \mathbb{R}^{k \times k}$ are both upper triangular nonsingular matrices, $R_2 \in \mathbb{R}^{k \times (n-k)}$, and $\widetilde{R}_2 \in \mathbb{R}^{k \times (m-k)}$. Moreover,

$$\begin{aligned}
QQ^T &\quad \text{is the orthogonal projection onto } \mathrm{Ran}(A), \\
I - QQ^T &\quad \text{is the orthogonal projection onto } \mathrm{Null}(A^T), \\
\widetilde{Q}\widetilde{Q}^T &\quad \text{is the orthogonal projection onto } \mathrm{Ran}(A^T), \text{ and} \\
I - \widetilde{Q}\widetilde{Q}^T &\quad \text{is the orthogonal projection onto } \mathrm{Null}(A)^\perp.
\end{aligned}$$