Math 407: Linear Optimization

Lecture 16: The Linear Least Squares Problem II

Math Dept, University of Washington

February 28, 2018

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

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

Linear Least Squares

A linear least squares problem is one of the form

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

where

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

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, $A\bar{x} = b$.

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

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

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

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

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 $\mathcal{LLS}$.

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 !

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.

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

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

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, $\text{Ran}(A) = \mathbb{R}^m$. 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 having least norm:

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

This problem is of the form $\mathcal{D}$ $\min \left\{ \|z - b\|_2 \,\middle|\, z \in S \right\}$ 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 $P_{\mathrm{Null}(A)}$

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

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 $P_{\text{Ran}(M)} = M(M^T M)^{-1} M^T$.
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 $\mathcal{D}$ $\min\left\{\|z-b\|_2 \,\middle|\, z \in S\right\}$

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

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

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

That is, $x^0 = A^T(AA^T)^{-1}b$ is the least norm solution to $Ax = b$.

Theorem:
Let $A \in \mathbb{R}^{m\times n}$ be such that $m \le n$ 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

We now define the Gram-Schmidt orthogonalization process for a set of linearly independent vectors $a_1, \dots, a_n \in \mathbb{R}^m$.
(This implies that $n <\leq m$ (why?))

The orthogonal vectors $q_1, \dots, q_n$ are defined inductively, as follows:

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

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

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

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

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

Again suppose $a_1, \ldots, a_n \in \mathbb{R}^m$ are linearly independent and set

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

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

$$A = QR \ ,$$

where Q has orthonormal columns 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

$$AP = QR = [Q_1 \;\; Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1. \tag{1}$$

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 form an orthonormal basis for $\mathrm{Ran}(A)$.

(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 and R_{11} is invertible.

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 $k \le \min\{m,n\}$. 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,

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

Solving the Normal Equations with the QR Factorization

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ with $k := \mathrm{rank}(A) \leq \min\{m, n\}$.

The normal equations $A^T Ax = A^T b$ yield solutions to the LLS problem

$$\min \tfrac{1}{2} \|Ax - b\|_2^2 \ .$$

We show how the QR factorization of A is used to solve these equations.

Solving the Normal Equations with the QR Factorization

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ with $k := \mathrm{rank}\,(A) \leq \min\{m, n\}$.

The normal equations $A^T Ax = A^T b$ yield solutions to the LLS problem

$$\min \tfrac{1}{2} \|Ax - b\|_2^2 \ .$$

We show how the QR factorization of A is used to solve these equations.

Suppose A has condensed QR factorization

$$AP = Q[R_1 \ R_2]$$

where

$P \in \mathbb{R}^{n \times n}$ is a permutation matrix,
$Q \in \mathbb{R}^{m \times k}$ has orthonormal columns,
$R_1 \in \mathbb{R}^{k \times k}$ is nonsingular and upper triangular, and
$R_2 \in \mathbb{R}^{k \times (n-k)}$ with $k = \mathrm{rank}\,(A) \leq \min\{n, m\}$.

Solving the Normal Equations with the QR Factorization

Since

$$A = Q[R_1 \;\; R_2]P^T$$

the normal equations $A^T Ax = A^T b$ become

$$P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T Q \begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T x = A^T Ax = A^T b = P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T b,$$

or equivalently

$$P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} \begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T x = P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T b.$$

since $Q^T Q = I_k$.

Solving the Normal Equations with the QR Factorization

Multiply

$$P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} \begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T x = P \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T b$$

on the left by P^T to obtain

$$\begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} \begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T x = \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T b$$

Defining $w := \begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T x$ and setting $\hat{b} := Q^T b$ gives

$$\begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} w = \begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} \hat{b}.$$

Take $w = \hat{b}$ and define $\hat{x} := P^T x$. Then

$$\begin{bmatrix} R_1 & R_2 \end{bmatrix} \hat{x} = \hat{b}.$$

Solving the Normal Equations with the QR Factorization

We now need to solve the following equation for $\hat{x}$:

$$\begin{bmatrix} R_1 & R_2 \end{bmatrix} \hat{x} = \hat{b}.$$

To do this write

$$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix}$$

with $\hat{x}_1 \in \mathbb{R}^k$. If we assume that $\hat{x}_2 = 0$, then

$$R_1 \hat{x}_1 = \begin{bmatrix} R_1 & R_2 \end{bmatrix} \begin{pmatrix} \hat{x}_1 \\ 0 \end{pmatrix} = \hat{b}$$

which we can solve by taking

$$\hat{x}_1 = R_1^{-1} \hat{b}$$

since R_1 is an invertible upper triangular $k \times k$ matrix by construction.

Solving the Normal Equations with the QR Factorization

Then

$$x = P\hat{x} = P\begin{pmatrix} R_1^{-1}\hat{b} \\ 0 \end{pmatrix}.$$

Does this x solve the normal equations?

Solving the Normal Equations with the QR Factorization

Then

$$x = P\hat{x} = P\begin{pmatrix} R_1^{-1}\hat{b} \\ 0 \end{pmatrix}.$$

Does this x solve the normal equations?

$$\begin{aligned}
A^T A\overline{x} &= A^T AP\begin{bmatrix} R_1^{-1}\hat{b} \\ 0 \end{bmatrix} \\
&= A^T Q\begin{bmatrix} R_1 & R_2 \end{bmatrix} P^T P\begin{bmatrix} R_1^{-1}\hat{b} \\ 0 \end{bmatrix} \\
&= A^T QR_1R_1^{-1}\hat{b} && \text{(since } P^TP = I\text{)} \\
&= A^T Q\hat{b} \\
&= A^T QQ^T b \\
&= P\begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T QQ^T b && \text{(since } A^T = P\begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T\text{)} \\
&= P\begin{bmatrix} R_1^T \\ R_2^T \end{bmatrix} Q^T b && \text{(since } Q^TQ = I\text{)} \\
&= A^T b \quad !
\end{aligned}$$

Solving the Normal Equations with the QR Factorization

This derivation yields the following recipe for solving the normal equations:

$AP = Q[R_1 \ R_2]$	the general condensed QR factorization	$o(m^2 n)$
$\hat{b} = Q^T b$	a matrix-vector product	$o(km)$
$\overline{w}_1 = R_1^{-1}\hat{b}$	a back solve	$o(k^2)$
$\overline{x} = P \begin{bmatrix} R_1^{-1}\hat{b} \\ 0 \end{bmatrix}$	a matrix-vector product	$o(kn)$.