This handout supplements the text's approach to least square problems via the normal equations (Thm. 0.35) by an approach via orthogonality.

Solve

\[ x = \text{argmin} \| Ax - b \|_2^2 \]

where \( A \) is \( m \times n \) with \( m \geq n \) and \( A \) has rank \( n \). Let \( P \) be the orthonormal projection onto the range of \( A \)

\[ \text{R}(A) = \{ Ax \mid x \in \mathbb{R}^n \} \subseteq \mathbb{R}^m \]

Write

\[ b = Pb + (I - P)b = b_0 + b_1 \]

where \( b_0 \in \text{R}(A) \) and \( b_1 \perp \text{R}(A) \).
By the Pythagorean theorem, since
\[ Ax - b_0 \perp b_1 \quad \text{for any } x \]
(since \( Ax \in \mathbb{R}(A) \) and \( b_0 \in \mathbb{R}(A) \)),
so
\[ \| Ax - b \|_2^2 = \| Ax - b_0 \|_2^2 + \| b_1 \|_2^2, \]
where the second term is independent of \( x \). Hence
\[ \arg \min \| Ax - b \|_2^2 = \arg \min \| Ax - b_0 \|_2^2. \]
Since \( b_0 \in \mathbb{R}(A) \), there is some \( x_0 \in \mathbb{R}^n \) with \( b_0 = Ax_0 \), and we can minimize \( \| Ax - b_0 \|_2^2 = 0 \) by choosing \( x = x_0 \).

In practice, we follow two steps:

1. Compute the projection \( P \) by creating an ON basis for \( \mathbb{R}(A) \),
and
(2) Solving \( Ax = b_0 = \Pi b \) where \( A \) is \( mxn \) and \( b_0 \in \text{R}(A) \).

For (1), we view \( \text{R}(A) \) as the column space spanned by the columns \( a_1, \ldots, a_n \) of \( A \). Indeed,
\[
Ax = \sum_{i=1}^{n} x_i a_i \quad \forall x \in \mathbb{R}^n
\]

according to one interpretation of matrix multiplication. Note that since \( A \) has rank \( n \), its columns are linearly independent hence form a basis for \( \text{R}(A) \).

Thus we can ON i.e. them by Gram-Schmidt: \( e_1 = a_1/\|a_1\| \),
\[
e_2 = (a_2 - \langle a_2, e_1 \rangle e_1)/\|a_2 - \langle a_2, e_1 \rangle e_1\|_2
\]
and so forth.
In matrix terms, we seek a QR factorization

\[ A = QR = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} R_{11} & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_{nn} \end{bmatrix} \]

where \( Q \) has ON columns so
\( Q^*Q = I \) and \( R \) is upper triangular with positive diagonal entries

\[ R_{ij} = \|a_j - \langle a_j, e_i \rangle e_i\|_2 = \cdots = \|a_j\|_2. \]

Conversely, any such factorization produces an ON basis for \( \mathbb{R}(A) \), composed of the columns of \( Q \).

Once we have a QR factorization of \( A \), \( Q \) becomes much easier as well:
Since $Q^*Q = I$, solving
$$Ax = b_0$$
or
$$QRx = b_0$$
is equivalent to solving
$$Rx = Q^*b_0,$$
where $R$ is upper triangular and invertible.

Given the QR factorization of $A$, projection onto $R(A)$ is implemented by
$$b \mapsto b_0 = \sum_{j=1}^n \langle b, e_j \rangle e_j$$
$$= \left( \sum_{j=1}^n e_j e_j^* \right) b$$
$$= [e_1 \ldots e_n] \left[ \begin{array}{c} e_1^* \\ \vdots \\ e_n^* \end{array} \right] b$$
$$= QA^*b.$$
Thus \( P = QQ^* \). Note that
\[
Q^*Q = I 
\]
but \( QQ^* = P \neq I \) in general, especially if \( m > n \).

Since then \( QQ^* \) doesn't have enough rank to be an \( m \times m \) identity matrix. Thus \( 2 \) simplifies to
\[
QRx = QQ^*b \quad \text{or} \quad Rx = Q^*b,
\]
since we can cancel \( Q \) (why?).

Remark: We saw in class that
\[
A = QR \implies A^*A = R^*R,
\]
so \( R \) is the "Cholesky factor" of \( A \). Given \( R \),
\( Q = AR^{-1} \) is often convenient for hand calculations. In numerical analysis, there are better ways
to compute \( Q \) and \( R \).
Example. Approximate function values \( b_1 \ldots b_m \) at evaluation points \( t_1 \ldots t_m \) by a linear function \( x_1 + x_2 t \).

\[
AX - b = \begin{bmatrix} x_1 + x_2 t_1 - b_1 \\ x_1 + x_2 t_2 - b_2 \\ \vdots \\ x_1 + x_2 t_m - b_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}
\]

Onize columns of \( A \) to get

\[
e_1 = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad e_2 = \frac{t - \langle t, e_1 \rangle e_1}{\| t - \langle t, e_1 \rangle e_1 \|}
\]

Defining \( \bar{t} = \frac{1}{m} \sum_{i=1}^{m} t_i \) and

\[
\sigma = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (t_i - \bar{t})^2}
\]

gives

\[
e_2 = \frac{1}{\sigma} (t - \bar{t}).
\]
In matrix factorization language, we have

$$Q = \begin{bmatrix} e_1 & e_2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{m}} & - \frac{t}{\sigma} \\ 0 & \frac{1}{\sigma} \end{bmatrix}$$

since

$$e_1 = \frac{a_1}{\sqrt{m}} \quad e_2 = \frac{1}{\sigma} (a_2 - \frac{t}{\sigma} a_1).$$

Hence

$$A = QR$$

where

$$R = \begin{bmatrix} \frac{1}{\sqrt{m}} & - \frac{t}{\sigma} \\ 0 & \frac{1}{\sigma} \end{bmatrix} / \sigma \sqrt{m},$$

and

$$x = R^{-1} Q^* b$$

$$= \begin{bmatrix} \frac{1}{\sqrt{m}} & - \frac{t}{\sigma} \\ 0 & \frac{1}{\sigma} \end{bmatrix} \begin{bmatrix} \sqrt{m} \overline{b} \\ \frac{1}{\sigma} (t \overline{b} - \overline{c}) \end{bmatrix}$$

$$= \begin{bmatrix} \overline{b} - \frac{t}{\sigma} (t \overline{b} - \overline{c}) \frac{1}{\sigma} \\ \frac{1}{\sigma^2} (t \overline{b} - \overline{c}) \frac{1}{\sigma} \end{bmatrix}.$$