Section 6.5. Least-Squares Approximations

In this section, we consider linear systems

$$A\vec{x} = \vec{b} \tag{1}$$

when it is inconsistent. Recall that a system is inconsistent when the last row of its augmented matrix is a pivot column. In other words, its augmented matrix has an Echelon form

$$\left[A, \vec{b}\right] \longrightarrow \begin{bmatrix} * & * & ... & * & b_1 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 0 & b_p \\ 0 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... \end{bmatrix}, \quad b_p \neq 0,$$

where $\begin{bmatrix} 0 & 0 & ... & 0 & b_p \end{bmatrix}$ is the first non-zero row from the top. Note that as long as b_p is not zero, for instance, $b_p = 0.000001$, the system is inconsistent. From a practical point of view, due to data sampling errors or floating point errors within computers or programming, many supposedly consistent systems could turn out to be inconsistent. In this section, we shall learn a method that would produce the best possible approximation (in a certain sense) to the linear system. When a system is consistent, this method would provide solutions to the systems.

Definition Let $A_{m\times n}$ be a $m \times n$ matrix and $\vec{b} \in R^m$, a vector $\hat{y}_0 \in R^n$ is called a least-squares approximation solution for linear system (1) if it minimizes $\left\|\vec{b} - A\vec{x}\right\|^2$ among all $\vec{x}$ in R^n, i.e.,

$$\left\|\vec{b} - A\hat{y}_0\right\|^2 \leq \left\|\vec{b} - A\vec{x}\right\|^2 \quad \text{for all } \vec{x} \in R^n.$$

Note that if (1) is consistent, then a least-squares approximation solution is a solution $\vec{x}_0$ since

$$\left\|\vec{b} - A\vec{x}_0\right\|^2 = 0 \leq \left\|\vec{b} - A\vec{x}\right\|^2.$$

We also note that, since

$$dist\left(\vec{b}, Col\,(A)\right) = \min\left\{\left\|\vec{b} - \vec{z}\right\| : \vec{z} \in Col\,(A)\right\} = \min\left\{\left\|\vec{b} - A\vec{x}\right\| : \ \vec{x} \in R^n\right\},$$

a least-squares approximation solution is such that

$$\left\|\vec{b} - A\hat{y}_0\right\| = dist\left(\vec{b}, Col\,(A)\right).$$

This observation motivates us to find least-squares approximation solutions as follows.

Let $\hat{b} = \text{Proj}_{Col(A)}\left(\vec{b}\right) \in Col\,(A)$. By definition, this means that

$$A\vec{x} = \hat{b}$$

is consistent. Suppose that $\hat{y}_0$ is a solution: $A\hat{y}_0 = \hat{b}$.Then,

$$\begin{aligned} \left\| \vec{b} - A\hat{y}_0 \right\| &= \left\| \vec{b} - \hat{b} \right\| = \left\| \vec{b} - \text{Proj}_{Col(A)} \left(\vec{b} \right) \right\| \\ &\le \left\| \vec{b} - \vec{z} \right\| \quad (\text{for } \vec{z} \in Col\,(A)\,, \text{ since } \text{Proj}_{Col(A)} \left(\vec{b} \right) \text{ is closest to } \vec{b}) \\ &= \left\| \vec{b} - A\vec{x} \right\| \quad (\vec{z} = A\vec{x} \text{ for any } \vec{z} \in Col\,(A)\ ). \end{aligned}$$

This means that the solution $\hat{y}_0$ is a least-squares approximation solution.

We can further simplify the above two-step procedure. Let $\hat{b}$ $=\text{Proj}_{Col(A)} \left(\vec{b} \right)$. Then, since $Col\,(A) = \{A\vec{x} : \text{all } \vec{x} \in R^n\}$

$$\vec{b} - \hat{b} \perp Col\,(A) \Longrightarrow \left(\vec{b} - \hat{b} \right) \cdot A\vec{x} = 0.$$

Suppose $\hat{y}_0$ solves $A\hat{y}_0 = \hat{b}$. The above equation becomes

$$\left(\vec{b} - A\hat{y}_0 \right) \cdot A\vec{x} = 0.$$

Using the fact that, for any two vectors $\vec{u}$ and $\vec{v}$,

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} \qquad (\text{ one-row matrix times one-column matrix }),$$

we see, for any $\vec{x} \in R^n$,

$$\left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) \cdot \vec{x} = \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right)^T \vec{x} = \left(\vec{b} - A\hat{y}_0 \right)^T A\vec{x} = \left(\vec{b} - A\hat{y}_0 \right) \cdot A\vec{x} = 0.$$

In the above equation, if we take

$$\vec{x} = \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right),$$

then

$$\begin{aligned} 0 &= \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) \cdot \vec{x} = \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) \cdot \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) \\ &= \left\| \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) \right\|^2 \Longrightarrow \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) = 0 \end{aligned}$$

This leads to the following linear system

$$0 = \left(A^T \left(\vec{b} - A\hat{y}_0 \right) \right) = A^T\vec{b} - A^T A\hat{y}_0 \Longrightarrow A^T\vec{b} = A^T A\hat{y}_0.$$

Theorem The least-squares approximation solutions coincide with the solution set for

$$A^T A\hat{y} = A^T\vec{b}.$$

Any solution is a least-squares approximation.

Example 6.5.1. Find least-squares approximation solutions for

$$(a) \quad A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

$$(b) \quad A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{6\times 4}, \quad \vec{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

Solution. (a) We computer A^TA and $A^T\vec{b}$:

$$A^TA = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

We then solve

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \vec{y} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

by performing row operations as follows:

$$\begin{bmatrix} 17 & 1 & 19 \\ 1 & 5 & 11 \end{bmatrix} \to \begin{bmatrix} 0 & -84 & -168 \\ 1 & 5 & 11 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 2 \\ 1 & 5 & 11 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}.$$

So,

$$\begin{aligned} y_2 &= 2 \\ y_1 &= 1. \end{aligned}$$

Ans for (a): Least squares solution is $\vec{y_0} = [1, 2]^T$. One can verify that

$$\left\| A\vec{y_0} - \vec{b} \right\|^2 = \left\| \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \right\|^2 = 84.$$

This means that it is impossible to solve

$$A\vec{y} = \vec{b} \quad \text{or} \quad \left\| A\vec{y} - \vec{b} \right\| = 0.$$

The best one can do is to make

$$\left\| A\vec{y} - \vec{b} \right\| = \sqrt{84} = 9.165\,2.$$

Solution: (b)

$$(b) \qquad A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}_{6\times 4} , \quad \vec{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

Then

$$A^T A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}^T \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}.$$

We then solve

$$A^T A\hat{y} = A^T\vec{b}$$

i.e.,

$$\begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \hat{y} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}.$$

We solve it by row operations on augmented matrix:

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \to \begin{bmatrix} 0 & 2 & 2 & -4 & -14 \\ 0 & 2 & 0 & -2 & -10 \\ 0 & 0 & 2 & -2 & -4 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix}$$

$$\to \begin{bmatrix} 0 & 0 & 2 & -2 & -4 \\ 0 & 2 & 0 & -2 & -10 \\ 0 & 0 & 2 & -2 & -4 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 1 & 0 & 0 & 1 & 3 \end{bmatrix}.$$

y_4 is the free variable, and

$$\begin{aligned} y_1 &= 3 - y_4 \\ y_2 &= -5 + y_4 \\ y_3 &= -2 + y_4. \end{aligned}$$

There are infinite many least squares approximations. For instance, letting $y_4 = 0$, we have

$$\hat{y} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix};$$

or $y_4 = 1$,

$$\vec{z} = \begin{bmatrix} 2 \\ -4 \\ -1 \\ 1 \end{bmatrix}.$$

One can verify that

$$\left\| A\hat{y} - \vec{b} \right\| = \left\| \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} - \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ -2 \\ 2 \end{bmatrix} \right\| = 2\sqrt{3}.$$

For the second solution $\vec{z}$,

$$\left\| A\vec{z} - \vec{b} \right\| = \left\| \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ -2 \\ 2 \end{bmatrix} \right\| = 2\sqrt{3}.$$

Section 6.4. The Gram-Schmidt Algorithm

The Gram-Schmidt Algorithm is a process for producing an orthonormal basis for any subspace in R^n. The idea of this algorithm is demonstrated below.

Example 6.4.1. Let $W = Span\{\vec{v}_1, \vec{v}_2\}$, where

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \ \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Construct an orthonormal basis for W.

Solution. We start with

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{1}{\sqrt{9+36}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{45}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

Next, we need to construct a vector $\vec{u}_2$ satisfying

$$\begin{aligned} &(1)\ \vec{u}_2 \text{ is orthogonal to } \vec{u}_1 \ \text{ and is a unit vector} \\ &(2)\ \vec{u}_2 \text{ is in } W = Span\{\vec{v}_1, \vec{v}_2\}. \end{aligned}$$

Suppose we already find this vector $\vec{u}_2$. Then, since both $\vec{u}_1$ and $\vec{u}_2$ are in W, we see

$$Span\{\vec{u}_1, \vec{u}_2\} \subset W.$$

Note that Condition (1) implies that $\vec{u}_1$ and $\vec{u}_2$ are linearly independent, $Span\{\vec{u}_1, \vec{u}_2\}$ has dimension 2. So

$$Span\{\vec{u}_1, \vec{u}_2\} = W.$$

Thus, $\vec{u}_1, \vec{u}_2$ form an orthonormal basis. How to find such a vector $\vec{u}_2$? Recall the concept of orthogonal decomposition:

$$\vec{v}_2 = \text{Proj}_{\vec{u}_1}\vec{v}_2 + \vec{z}, \quad z \perp \vec{u}_1.$$

Since $\vec{u}_1$ is in W, we see that

$$\begin{aligned} \vec{z} &= \vec{v}_2 - Proj_{\vec{u}_1}\vec{v}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\,\vec{u}_1 \\ &= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{5}{\sqrt{5}}\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \end{aligned}$$

is in W. Thus, the vector

$$\vec{u}_2 = \frac{\vec{z}}{\|\vec{z}\|} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

is exactly what we want. So

$$\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

forms an orthonormal basis.

The Gram-Schmidt Algorithm: Suppose that $\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_p\}$ forms a basis for the subspace $W = Span\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_p\}$ in R^n. Define

$$\vec{u}_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|^2}$$

$$\vec{v}_2 = \vec{x}_2 - \vec{x}_2 \cdot \vec{v}_1 \vec{v}_1 \qquad (= \vec{x}_2 - \text{Proj}_{Span\{\vec{v}_1\}} \vec{x}_2$$

$$\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|^2}$$

$$\vec{v}_3 = \vec{x}_3 - (\vec{x}_3 \cdot \vec{u}_1)\,\vec{u}_1 - (\vec{x}_3 \cdot \vec{u}_2)\,\vec{u}_2 \qquad (= \vec{x}_3 - \text{Proj}_{Span\{\vec{u}_1, \vec{u}_1\}} \vec{x}_3)$$

$$\vec{u}_3 = \frac{\vec{v}_3}{\|\vec{v}_3\|^2}$$

$$...$$

$$\vec{v}_i = \vec{x}_i - \text{Proj}_{Span\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_{i-1}\}} \vec{x}_i$$

$$\vec{u}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|^2}$$

$$...$$

$$\vec{v}_p = \vec{x}_p - \text{Proj}_{Span\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_{p-1}\}} \vec{x}_p = \vec{x}_p - [(\vec{x}_p \cdot \vec{u}_1)\,\vec{u}_1 + ... + (\vec{x}_p \cdot \vec{u}_{p-1})\,\vec{u}_{p-1}]$$

$$\vec{u}_p = \frac{\vec{v}_p}{\|\vec{v}_p\|^2}.$$

Then

$$\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\} \text{ forms an orthogonal basis, and}$$
$$\{\vec{u}_1,\ \vec{u}_2, ..., \vec{u}_p\} \text{ forms an orthogonal basis for } W.$$

QR Factorization of Matrices.

Let $A = [\vec{x}_1, \vec{x}_2, ..., \vec{x}_p]_{n \times p}$ be a column matrix whose columns are linearly independent. Then A can be factored as

$$A = QR$$

where Q is an orthonormal matrix

$$Q \ = [\vec{u}_1,\ \vec{u}_2, ..., \vec{u}_p]_{n \times p}$$

whose columns $\{\vec{u}_1,\ \vec{u}_2, ..., \vec{u}_p\}$ forms an orthonormal basis for $Col\,(A)$, and

R is a $(p \times p)$ upper triangular invertible matrix with positive diagonal entries.

Proof. We use the Gram-Schmidt algorithm to produce $\{\vec{u}_1,\ \vec{u}_2, ..., \vec{u}_p\}$. One sees that

$$\begin{aligned}
\vec{v}_i &= \vec{x}_i - \text{Proj}_{Span\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_{i-1}\}} \vec{x}_i = \vec{x}_i - \left[(\vec{x}_i \cdot \vec{u}_1)\, \vec{u}_1 + ... + (\vec{x}_i \cdot \vec{u}_{i-1})\, \vec{u}_{i-1}\right] \\
\vec{u}_i &= \frac{\vec{v}_i}{\|\vec{v}_i\|^2}.
\end{aligned}$$

So $\vec{x}_i$ is a linear combination of $\{\vec{u}_1,\ \vec{u}_2, ..., \vec{u}_i\}$ as

$$\begin{aligned}
\vec{x}_i &= \vec{v}_i + (\vec{x}_i \cdot \vec{u}_1)\, \vec{u}_1 + ... + (\vec{x}_i \cdot \vec{u}_{i-1})\, \vec{u}_{i-1} \\
&= (\vec{x}_i \cdot \vec{u}_1)\, \vec{u}_1 + ... + (\vec{x}_i \cdot \vec{u}_{i-1})\, \vec{u}_{i-1} + \|\vec{v}_i\|\, \vec{u}_i.
\end{aligned}$$

In particular, we have

$$\begin{aligned}
\vec{x}_1 &= \|\vec{v}_1\|\, \vec{u}_1 \\
\vec{x}_2 &= (\vec{x}_2 \cdot \vec{u}_1)\, \vec{u}_1 + \|\vec{v}_2\|\, \vec{u}_2 \\
\vec{x}_3 &= (\vec{x}_3 \cdot \vec{u}_1)\, \vec{u}_1 + (\vec{x}_3 \cdot \vec{u}_2)\, \vec{u}_2 + \|\vec{v}_3\|\, \vec{u}_3.
\end{aligned}$$

Thus,

$$\begin{aligned}
A &= [\vec{x}_1, \vec{x}_2, ..., \vec{x}_p]_{n \times p} \\
&= [\vec{u}_1, \vec{u}_2, ..., \vec{u}_p]_{n \times p} \begin{bmatrix} \|\vec{v}_1\| & (\vec{x}_2 \cdot \vec{u}_1) & (\vec{x}_3 \cdot \vec{u}_1) & \cdots & (\vec{x}_p \cdot \vec{u}_1) \\ 0 & \|\vec{v}_2\| & (\vec{x}_3 \cdot \vec{u}_2) & \cdots & (\vec{x}_p \cdot \vec{u}_2) \\ 0 & 0 & \|\vec{v}_3\| & \cdots & (\vec{x}_p \cdot \vec{u}_3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \|\vec{v}_p\| \end{bmatrix}_{p \times p} \\
&= QR.
\end{aligned}$$