4.5. Overdetermined Systems

So far, we considered linear systems $Ax = b$ with the *same* number of equations and unknowns (i.e., $A \in \mathbb{R}^{n\times n}$). In the case where $A \in \mathbb{R}^{m\times n}$, with $m > n$ (more equations than unknowns) the existence of a true solution is not guaranteed, in this case we look for the "best possible" substitute for a solution. Before analyzing what that means, let's look at how such problems arise.

As an example, in an experiment, we measure the pressure of a gas in a closed container as a function of the temperature. From Physics, we know that

$$\begin{aligned} pV &= nR\frac{5}{9}(T + 459.67) \\ \Rightarrow p &= \alpha T + \beta, \qquad \alpha = \frac{5nR}{9V}, \beta = \frac{5nR \cdot 459.67}{9V} \end{aligned}$$

What are α and β? Experimentally, the measurements should ideally lie on a straight line $y = c_1 x + c_0$, but do not, due to measurement error. If we have n measurement pairs $(x_1, y_1), \ldots, (x_n, y_n)$ we would have wanted:

$$\left.\begin{array}{cccc} y_1 & = & c_1 x_1 & + & c_0 \\ y_2 & = & c_1 x_2 & + & c_0 \\ & \vdots & & & \\ y_n & = & c_1 x_n & + & c_0 \end{array}\right\} \Rightarrow \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Here, $A_{n\times 2} x_{2\times 1} = b_{n\times 1}$ is a rectangular system. We cannot hope to find a true solution to this system. Instead, lets try to find an "approximate" solution, such that $Ax \approx b$. Lets look at the residual of this "interpolation". The residual of the approximation of each data point is:

$$r_i = y_i - f(x_i) = y_i - c_1 x_i - c_0$$

If we write the vector of all residuals:

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} = \begin{bmatrix} y_1 - c_1 x_1 - c_0 \\ y_2 - c_1 x_2 - c_0 \\ \vdots \\ y_n - c_1 x_n - c_0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_0 \end{bmatrix} = b - Ax$$

Although we can't find an x such that $Ax = b$ (thus, $r = 0$), we can at least try to make r *small*. As another example, consider the problem of finding the best parabola $f(x) = c_2 x^2 + c_1 x + c_0$ that fits measurements $(x_1, y_1), \ldots, (x_n, y_n)$. We would like

$$\left.\begin{array}{c} f(x_1) \approx y_1 \\ f(x_2) \approx y_2 \\ \vdots \\ f(x_n) \approx y_n \end{array}\right\} = \left.\begin{array}{c} c_2 x_1^2 + c_1 x_1 + c_0 \approx y_1 \\ c_2 x_2^2 + c_1 x_2 + c_0 \approx y_2 \\ \vdots \\ c_2 x_n^2 + c_1 x_n + c_0 \approx y_n \end{array}\right\} \Rightarrow \underbrace{\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ & \vdots & \\ x_n^2 & x_n & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix}}_{x} \approx \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{b}$$

Once again, we would like to make $r = b - Ax$ as small as possible.

How do we quantify r being small? $\Rightarrow$ using a norm! We could ask that $\|r\|_1, \|r\|_2$ or $\|r\|_\infty$ be as small as possible. Any of these norms would be intuitive to consider for minimization (especially 1- and ∞-norms are very intuitive). However, we typically use the 2-norm for this purpose, because its the easiest to work with in this problem.

Definition

The *least squares solution* of the overdetermined system $Ax \approx b$ is the vector x that minimizes $\|r\|_2 = \|b - Ax\|_2$.

Define $Q(x) = Q(x_1, x_2, \ldots, x_n) = \|b - Ax\|_2^2$ where $x = (x_1, \ldots, x_n)$ and $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ ($m > n$). The least squares solution is the set of values $x_1, \ldots, x_n$ that *minimize* $Q(x_1, x_2, \ldots, x_n)$.

$$Q(x_1, \ldots, x_n) = \|b - Ax\|_2^2 = \|r\|_2^2 = \sum_{i=1}^{m} r_i^2$$

$$r = b - Ax \Rightarrow r_i = b_i - (Ax)_i \Rightarrow r_i = b_i - \sum a_{ij} x_j$$

$$\Rightarrow Q(x_1, \ldots, x_n) = \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j\right)^2$$

If $x_1 \ldots, x_n$ are those that *minimize* Q, then:

$$\frac{\partial Q}{\partial x_1} = 0, \frac{\partial Q}{\partial x_2} = 0, \ldots, \frac{\partial Q}{\partial x_n} = 0$$

in order to guarantee a minimum.

$$
\begin{aligned}
\frac{\partial Q}{\partial x_k} &= \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{m}\left(b_i - \sum_{j=1}^{n} a_{ij}x_j\right)^2\right) \\
&= \sum_{i=1}^{m} \frac{\partial}{\partial x_k}\left(b_i - \sum_{j=1}^{n} a_{ij}x_j\right)^2 \\
&= \sum_{i=1}^{m} 2\underbrace{\left(b_i - \sum_{j=1}^{n} a_{ij}x_j\right)}_{r_i} \frac{\partial}{\partial x_k}\left(b_i - \sum_{j=1}^{n} a_{ij}x_j\right) \\
&= \sum_{i=1}^{m} -2r_i a_{ik} = -2\sum_{i=1}^{m}[A^T]_{ki} r_i = -2[A^T r]_k = 0 \\
&\Rightarrow\ [A^T r]_k = 0
\end{aligned}
$$

Thus,

$$
\left.
\begin{array}{rclcrcl}
\partial Q/\partial x_1 & = & 0 & \Rightarrow & [A^T r]_1 & = & 0 \\
\partial Q/\partial x_2 & = & 0 & \Rightarrow & [A^T r]_2 & = & 0 \\
 & & & \vdots & & & \\
\partial Q/\partial x_n & = & 0 & \Rightarrow & [A^T r]_n & = & 0
\end{array}
\right\} \Rightarrow \boxed{A^T r = 0}
$$

Since $r = b - Ax$, we have:

$$
0 = A^T r = A^T(b - Ax) = A^T b - A^T Ax \Rightarrow \boxed{A^T Ax = A^T b}
$$

The system above is called the *normal equations system*; it is a *square* system that has as solution the least-squares approximation of $Ax \approx b$.

$$
\underbrace{A^T_{n\times m} A_{m\times n}}_{n\times n} \underbrace{x_{n\times 1}}_{n\times 1} = \underbrace{A^T_{n\times m} b_{m\times 1}}_{n\times 1}
$$

The normal equations *always* have a solution (with the simple condition that the columns of A have to be linearly independent, which is usually true).

4.5.1. QR factorization

While the normal equations can adequately compute the least squares solution, the condition number of $A^T A$ is the *square* of that of A (if A was a square matrix). An alternative method that does not suffer from this problematic conditioning is QR factorization.

Definition

An $n \times n$ matrix Q is called *orthonormal* if and only if

$$
Q^T Q = QQ^T = I
$$

Theorem

Let $A \in \mathbb{R}^{m \times n}$ $(m > n)$ have linearly independent columns. Then a decomposition $A = QR$ exists, such that $Q \in \mathbb{R}^{m \times m}$ is orthonormal and $R \in \mathbb{R}^{m \times n}$ is upper triangular, i.e.,

$$R = \begin{pmatrix} \hat{R} \\ O \end{pmatrix}$$

where $\hat{R}$ is an $n \times n$ upper triangular matrix. Additionally, given that A has linearly independent columns, all diagonal elements $r_{ii} \neq 0$.

Now, let us write

$$Q = \left[\begin{array}{c|c} \hat{Q} & Q^{\star} \end{array} \right]$$

where $\hat{Q} \in \mathbb{R}^{m \times n}$ contains the first n columns of Q and $Q^{\star} \in \mathbb{R}^{m \times (m-n)}$ contains the last $(m-n)$ columns. Respectively, we write:

$$R = \begin{pmatrix} \hat{R} \\ O \end{pmatrix}$$

where $\hat{R} \in \mathbb{R}^{n \times n}$ (and upper triangular) contains the first n rows of R. $\hat{R}$ is also *non-singular* because it has linearly independent columns. We can verify the following:

$$\hat{Q}^T \hat{Q} = I_{n \times n} \qquad \text{(although } \hat{Q}\hat{Q}^T \neq I_{m \times m}\text{!)}$$

The factorization $A = \hat{Q}\hat{R}$ is the so-called *economy size* QR factorization. Once we have $\hat{Q}$ and $\hat{R}$ computed, we observe that the normal equations can be written as:

$$\begin{aligned} A^T A x =& \ A^T b \\ \Rightarrow \hat{R}^T \underbrace{\hat{Q}^T \hat{Q}}_{=I_{m \times m}} \hat{R} x =& \ \hat{R}^T \hat{Q}^T b \\ \Rightarrow \hat{R}^T \hat{R} x =& \ \hat{R}^T \hat{Q}^T b \\ \Rightarrow \boxed{\hat{R} x = \hat{Q}^T b} \end{aligned}$$

The last equality follows because $\hat{R}$ is invertible. The benefit of using the QR factorization is that $\textbf{cond}(A^T A) = [\textbf{cond}(\hat{R})]^2$. Thus, the resulting equation is *much better* conditioned than the normal equations.