1 The Discrete Least Squares Problem (3.1)

Given points $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$, we want to find an *approximate function* to fit those points; that is,

$$p(t_i) = y_i$$

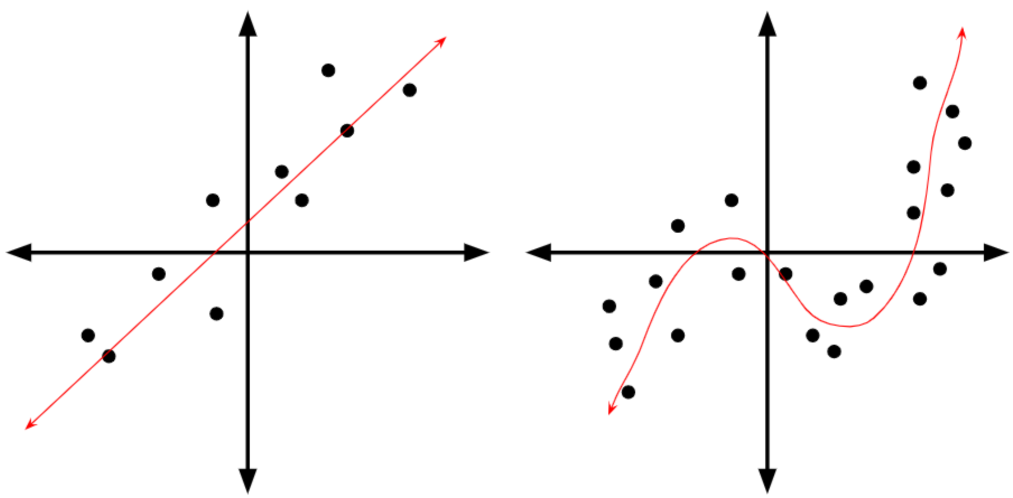
for $i = 1, 2, \ldots, n$. For example, suppose we want to find

$$p(t) = a_0 + a_1 t$$

or

$$p(t) = a_0 + a_1 t + a_2 t^2.$$

The idea is that we're just trying to find the *line of best fit* (a line with the smallest margin of error). We're not trying to find a line that passes through all the points, just the one of best fit.



The error, as mentioned, is defined by $r_i = y_i - p(t_i)$, where y_i is the exact value and $p(t_i)$ is the value of the function (the estimate). Then, the r_i's generate a vector known as the **residual** vector.

1.1 Problem Statement

Our goal is to find a $p(t) = a_0 + a_1 t$ such that $r \in \mathbb{R}^a$ is as small as possible. Hence, it turns into finding a_0, a_1 to minimize $||r||_2$.

1.1.1 Matrix Notation

Let's begin by converting

$$r_i = y_i - p(t_i) = y_i - (a_0 + a_1 t_i) = y_i - a_0 - a_1 t_i \quad i = 1, 2, \ldots, n$$

into matrix notation. This gives us

$$\underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}}_{\mathbf{r} \in \mathbb{R}^n} = \begin{bmatrix} y_1 - a_0 - a_1 t_1 \\ y_2 - a_0 - a_1 t_2 \\ \vdots \\ y_n - a_0 - a_1 t_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \begin{bmatrix} a_0 + a_1 t_1 \\ a_0 + a_1 t_2 \\ \vdots \\ a_0 + a_1 t_n \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y} \in \mathbb{R}^n} - \underbrace{\begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}}_{A \in \mathbb{R}^{n \times m}} \underbrace{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}}_{\mathbf{x} \in \mathbb{R}^m}$$

So, with the above matrix notation, also represented by

$$\mathbf{r} = \mathbf{y} - A\mathbf{x},$$

the goal is to find an $\mathbf{x} \in \mathbb{R}^m$ such that $||\mathbf{r}||_2$ is minimized. Essentially,

$$\min_{x \in \mathbb{R}^m} ||\mathbf{r}||_2 = \min_{x \in \mathbb{R}^m} ||\mathbf{y} - A\mathbf{x}||_2.$$

Here,

- n is the number of datapoints, and
- m is the number of unknowns from the function.

Remarks:

- A is not squared.
- $n > m$.

There will be no solutions[1] to $A\mathbf{x} = \mathbf{y}$. Instead, we want to find $\mathbf{x}$ that minimizes the overall error.

1.2 Other Basis Functions

Instead of linear functions $p(t) = a_0 + a_1 t$, we an try other types of functions. For example, consider polynomials, or

$$p(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k = \sum_{i=0}^{k} a_i t^i.$$

Here, $t^0 = 1$ and we have $(k+1)$ unknowns. The matrix formulation is given by

$$\underbrace{\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}}_{r \in \mathbb{R}^n} = \begin{bmatrix} y_1 - p(t_1) \\ y_2 - p(t_2) \\ \vdots \\ y_n - p(t_n) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} p(t_1) \\ p(t_2) \\ \vdots \\ p(t_n) \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y} \in \mathbb{R}^n} - \underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^k \\ 1 & t_2 & t_2^2 & \ldots & t_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \ldots & t_n^k \end{bmatrix}}_{A \in \mathbb{R}^{nn \times (k+1)}} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix}}_{\mathbf{x} \in \mathbb{R}^{k+1}}.$$

Here, $\mathbf{r} = \mathbf{y} - A\mathbf{x}$ is called the *residual*. To solve this, we will use the QR decomposition.

1.3 QR Decomposition

In essence, QR decomposition states that we can decompose a matrix A into two matrices, Q and R, such that

$$A = QR,$$

where Q is orthogonal and R is upper-triangular.

1.3.1 A Brief Review

Definition 1.1: Orthogonal Matrix

$Q \in \mathbb{R}^{n \times n}$ is called **orthogonal** if $Q^T = Q^{-1}$, i.e., $Q^T Q = QQ^T = I$.

[1] A has more rows than columns.

(Example.) Consider the rotations in $\mathbb{R}^n$. Note that

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

It follows that

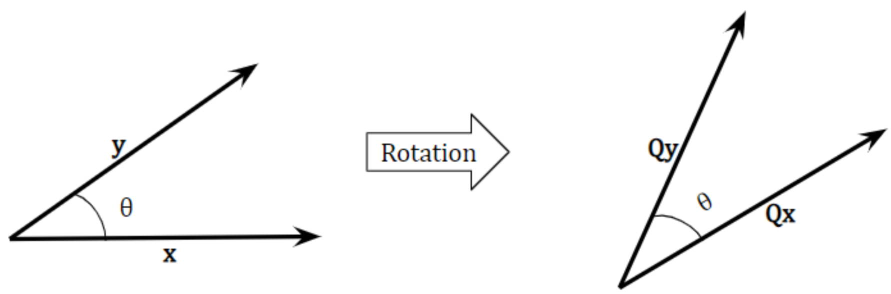
$$QQ^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Theorem 1.1

If $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, then

1. $\langle Q\mathbf{x}, Q\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$.
2. $||Q\mathbf{x}||_2 = ||\mathbf{x}||_2$.

It might be useful to consider this visualization:



Proof. We'll prove each part.

1. $\langle Q\mathbf{x}, Q\mathbf{y}\rangle = (Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T\mathbf{y} = \langle \mathbf{x}, \mathbf{y}\rangle$.
2. $||Q\mathbf{x}||_2 = \sqrt{\langle Q\mathbf{x}, Q\mathbf{y}\rangle} = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} = ||\mathbf{x}||_2$.

Thus, we're done. $\square$

1.3.2 QR Decomposition

Theorem 1.2

Let $A \in \mathbb{R}^{n\times m}$ such that $n \geq m$. Then, there exists an orthogonal $Q \in \mathbb{R}^{n\times m}$ and $R \in \mathbb{R}^{n\times m}$ with

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ 0 & r_{22} & \dots & r_{2m} \\ 0 & 0 & \dots & r_{3m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nm} \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \hat{\mathbb{R}} \\ 0 \end{bmatrix},$$

with $\hat{\mathbb{R}} \in \mathbb{R}^{m\times m}$ being an upper-triangular matrix and the 0 being the zero-matrix.

Later on, we'll talk more about QR decomposition and solving the Least Squares Problem.