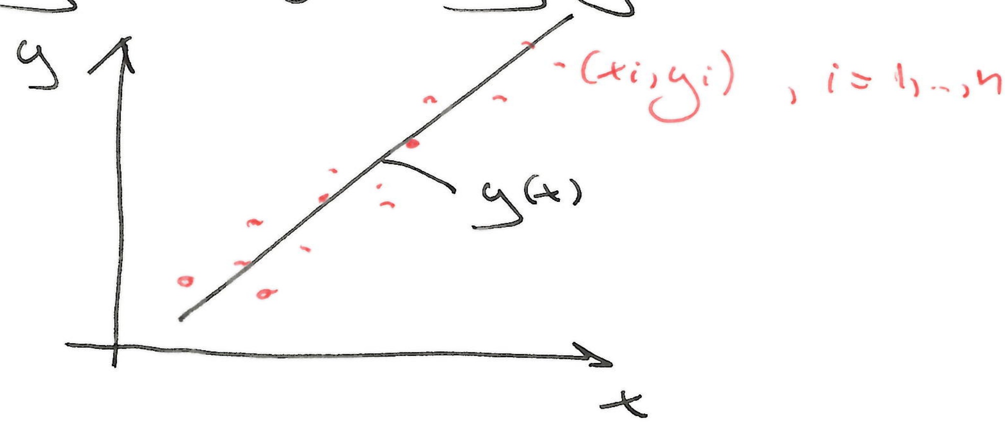
LEAST-SQUARES APPROXIMATION

Applicable when:

* the number of samples is much larger than the number of parameters in the approximating function

* the data is approximate only, so it is not necessary for the approximating function to go through every point



the approximating function $y(t)$ should be chosen to <u>minimize</u> the deviation from the data in some suitable sense (expressed using different vector norms)

$$S_p = \| \bar{e} \|_p \quad , \quad e_i = y_i - y(t_i)$$

* when p is odd $\| \cdot \|_p$ is <u>not differentiable</u> (problem hard to solve)

* $p = 2$ – the most common choice (least-squares approximation)

Ex Given data $\{x_i, y_i\}_{i=1}^{n}$

Approximating function $y = ax + b$

$$S(a,b) = \|e\|_2^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

Want to find $\underline{\min_{a,b} S(a,b)}$

$$\left.\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0 \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0 \end{aligned}\right\} \Rightarrow \underbrace{\begin{cases} a\sum_i x_i^2 + b\sum_i x_i = \sum_i x_i y_i \\ a\sum_i x_i + bn = \sum_i y_i \end{cases}}_{\text{system of normal equations easily solved for } (a,b)}$$

The approach can be generalized to higher-order approximating polynomials, e.g.,

$$y(x) = a_0 + a_1 x + \dots + a_m x^m \quad (\text{degree } m, \ m \ll n)$$

Then

$$S(a_0, a_1, \dots, a_m) = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - \dots - a_m x_n^m)^2$$

Conditions $\frac{\partial S}{\partial a_j} = 0, \ j = 0, \dots, m$ give rise to the normal system

$$\underbrace{\begin{bmatrix} N & \sum x_i & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \cdots & \sum x_i^{m+1} \\ \vdots & & & \\ \sum x_i^m & \sum x_i^{m+1} & \cdots & \sum x_i^{2m} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^n y_i \end{bmatrix}}_{b}$$

Structure of the normal system

$$Ax = b$$

$A = V^T V$, where $V : \mathbb{R}^{m+1} \to \mathbb{R}^n$

degree of the polynomial

number of data points

V – rectangular $n \times (m+1)$ ([illegible]) Vandermonde matrix

$$V = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ \vdots & \vdots & & \\ 1 & x_2 & \cdots & x_n^m \end{bmatrix} \quad \text{and} \quad b = V^T b$$

The normal system $V^T V x = V^T b$ is thus $(m+1) \times (m+1)$

The problem $Vx = b$ has dimension ~~[illegible]~~ $(m+1) \times n$ and is therefore overdetermined when $n > m+1$. When $n = m+1$, a standard interpolation is recovered

Remarks

* The matrix $V^T V$ of the normal system is symmetric (good!) but ~~to~~ ill-conditioned (bad!); due to poor conditioning m should not exceed ~ 10

* When the data exhibits special trends, other (non-polynomial) approximating functions can be used, e.g., $y(x) = ax^b$, ~~[illegible]~~)

Then

$$e_i = \log y_i - \underbrace{\log a}_{c} + b \log x_i = \log y_i - c - b \log x_i$$

* good conditioning of the normal matrix can be ~~...~~ eliminated by using combinations of <u>orthogonal polynomials</u> as approximating functions

ORTHOGONAL POLYNOMIALS (Trefethen, Ch 17)

Consider definition of a <u>weighted L_2 inner product</u> for functions $f, g : [a,b] \to \mathbb{R}$.

the weight function : $w \in C^1(a,b)$, $w(x) > 0 \quad \forall x \in (a,b)$

$$\int_{-1}^{1} w(x)\,dx < \infty$$

$$(f,g)_w = \int_a^b f(x)\,g(x)\,w(x)\,dx$$

the functions are f and g are <u>orthogonal</u> on (a,b) w.r.t the weight $w(x)$ iff

$$(f,g)_w = 0$$

Consider a family of degree-n polynomials $p_n(x)$ defined on $[a,b]$. For a given weight $w(x)$, one can obtain a family of <u>orthogonal</u> polynomials $p_0, p_1, p_2, \ldots$ <u>by performing</u> the Gram-Schmidt orthogonalization procedure. They satisfy the relations

$$(p_j, p_k)_w = 0 \quad k \neq j$$