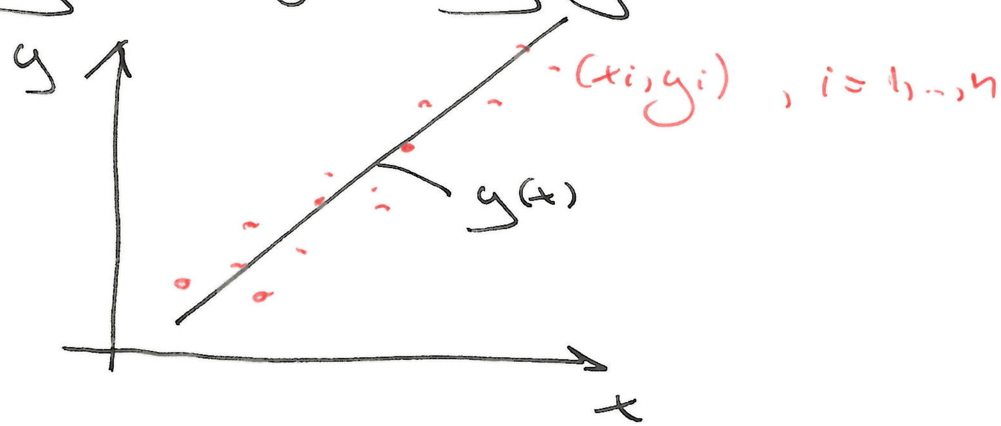
LEAST-SQUARES APPROXIMATION

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

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

* 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 $\min_{a,b} S(a,b)$

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

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

$$Ax = b$$

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

$m+1$: degree of the polynomial

n: number of data points

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

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

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

The problem $Vx = y$ has dimension $(m+1)\times n$ and is therefore <u>overdetermined</u> when $n > m+1$. When $n = m+1$, a standard interpolation is recovered

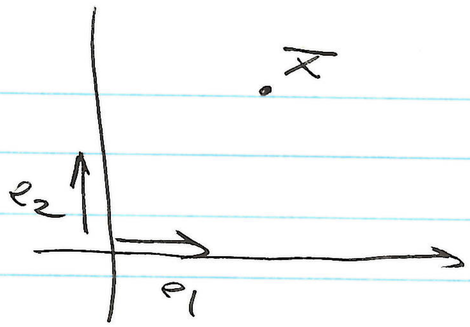
Remarks

* 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., $y(x) = a x^b$,)

Then

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



$$\bar{x} = \alpha_1 \bar{e}_1 + \alpha_2 e_2$$

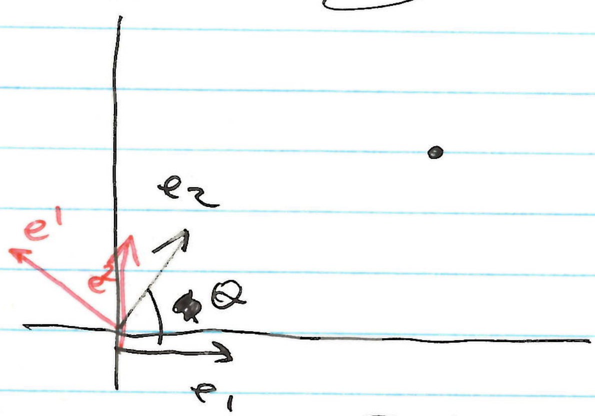
$$\langle \bar{x}, \bar{e}_1 \rangle = \alpha_1 \underbrace{\langle \bar{e}_1, \bar{e}_1 \rangle}_{1} + \alpha_2 \underbrace{\langle \bar{e}_1, \bar{e}_2 \rangle}_{0}$$

$$\alpha_1 = \langle \bar{x}, \bar{e}_1 \rangle, \quad \alpha_2 = \langle \bar{x}, e_2 \rangle$$

$$\|\bar{e}_1\| = \|\bar{e}_2\| = 1$$

$$\|\cdot\| = \|\cdot\|_2$$

Non-orthogonal basis



$$\bar{x} = \beta_1 \bar{e}_1 + \beta_2 \bar{e}_2$$

Dual basis $\bar{e}^j$, $j = 1, 2$

$$\langle \bar{e}_i, \bar{e}^j \rangle = \delta_{ij}$$

$\cos(\pi + \theta) =$

$\langle e^1, e_1 \rangle = \|e^1\|$

$\|e^1\| \sin\theta = 1$

$$\langle e^1, e_1 \rangle = 1, \quad \langle e^1, e_2 \rangle = 0$$

$$\langle e^2, e_1 \rangle = 0, \quad \langle e^2, e_2 \rangle = 1$$

Then $\langle e^1, \bar{x} \rangle = \beta_1 \underbrace{\langle e^1, e_1 \rangle}_{1} + \beta_2 \underbrace{\langle e^1, e_2 \rangle}_{0}$

$$\beta_1 = \langle e^1, \bar{x} \rangle$$

$$\underbrace{\angle(e_1, e_2)}_{\theta} \to 0 \implies |\beta_1|, |\beta_2| \to \infty$$

* poor conditioning of the normal matrix can be ~~[illegible]~~ eliminated by using combinations of orthogonal polynomials as approximating functions

ORTHOGONAL POLYNOMIALS (Trefethen, Ch 17)

Consider definition of a weighted L_2 inner product 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 orthogonal 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 orthogonal polynomials $p_0, p_1, p_2, \dots$ by performing the Gram-Schmidt orthogonalization procedure. They satisfy the relations

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