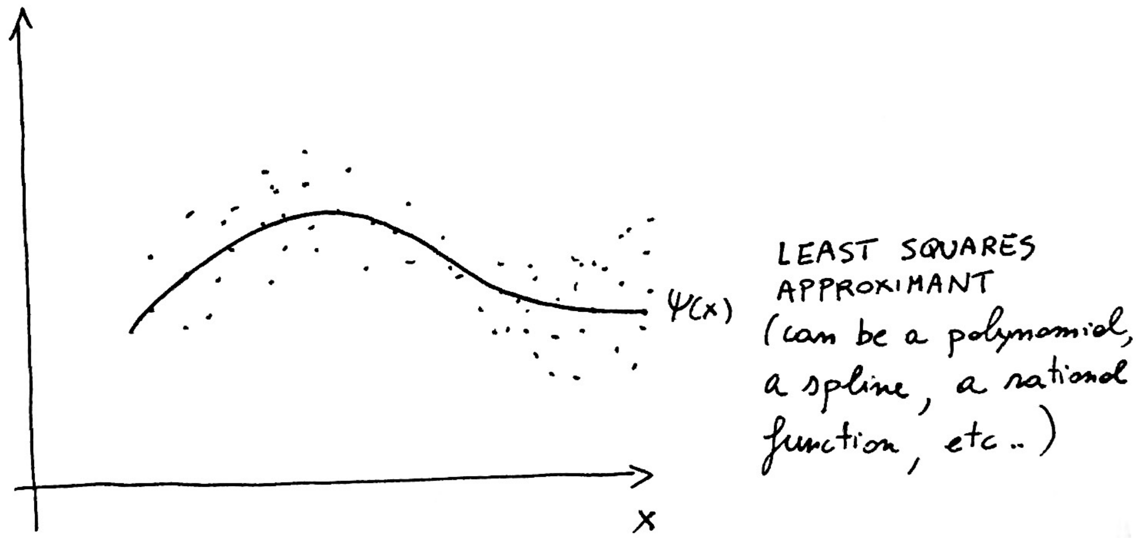
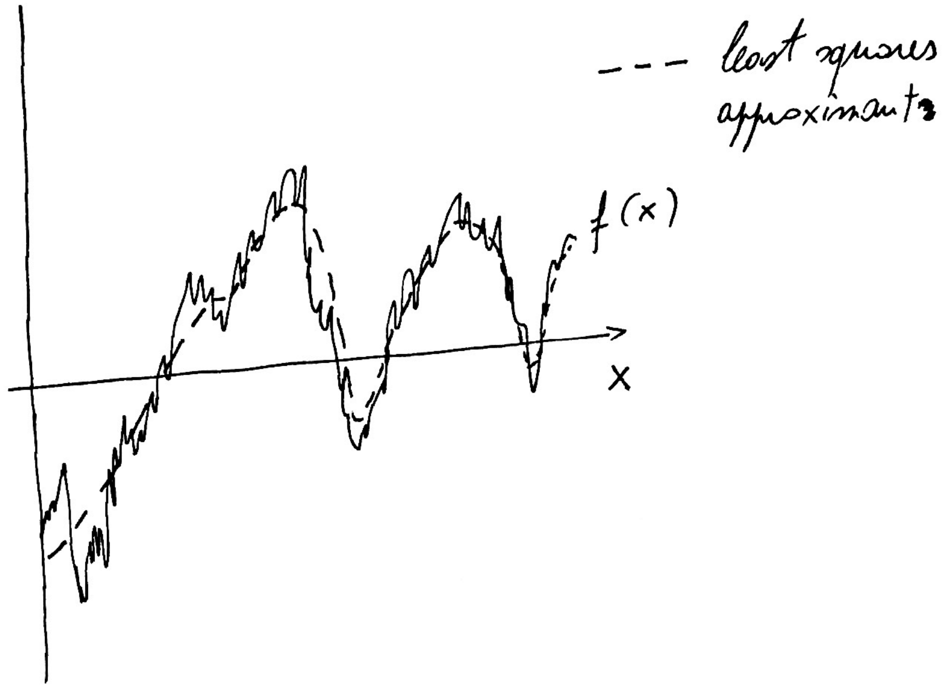
The least-squares method

Suppose we are interested in approximating (or modeling) the following data set



Clearly, it is not a good idea to compute an interpolant in this case as the interpolating function can be highly oscillatory. It may not even possible to compute such interpolant, e.g. if we have multiple data points at the same interpolation node.

Remark (Approximation of rough functions)



The key idea of the least squares method is to compute an approximant of a data set or a function $f(x)$ by minimizing a suitable error norm, usually the L_2 norm (discrete or continuous).

Least - squares approximation of data

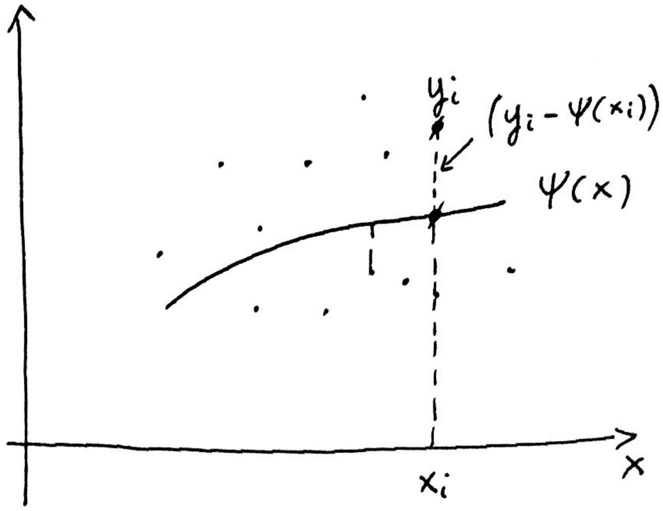
Suppose we have available $(n+1)$ data points $\{x_i, y_i\}_{i=0,\dots,n}$, where x_i are not necessarily distinct. We look for an approximant of such data set in the form:

$$\psi(x) = \sum_{k=0}^{m} a_k \underbrace{\phi_k(x)}_{\substack{\text{known} \\ \text{basis functions}}} \qquad (m+1) \leq (n+1)$$

To determine the unknown coefficients defining $\psi(x)$, i.e., $\{a_0, \dots, a_m\}$ we minimize the distance between $\psi(x)$ and the available data points:

$$\min_{a_0,\dots,a_m} \underbrace{\sum_{p=0}^{n} \left(y_p - \psi(x_p)\right)^2}_{E(a_0,\dots,a_m)} \qquad \text{(DISCRETE } L_2 \text{ NORM)}$$

ERROR FUNCTION



The minimum of the error function $E(a_0, \dots, a_m)$ can be determined by setting equal to zero the gradient, i.e.,

$$\frac{\partial E}{\partial a_K} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial a_K} \sum_{p=0}^{n} \left(y_p - \sum_{J=0}^{m} a_J \phi_J(x_p) \right)^2 = 0$$

$$\Rightarrow 2 \sum_{p=0}^{n} \left(y_p - \sum_{J=0}^{m} a_J \phi_J(x_p) \right) \phi_K(x_p) = 0$$

$$\Rightarrow \sum_{J=0}^{m} \left(\sum_{p=0}^{n} \phi_K(x_p)\, \phi_J(x_p) \right) a_J = \sum_{p=0}^{n} y_p\, \phi_K(x_p)$$

$K = 0, \dots, m$

(LINEAR SYSTEM OF EQUATIONS IN $a_0, \dots, a_m$)

The last system is known as system of NORMAL EQUATIONS and it can be written in the following matrix-vector form:

$$B^T B\, a = B^T y$$

$$B = {}_{n}\begin{bmatrix} \phi_0(x_0) & \cdots & \phi_m(x_0) \\ \vdots & & \vdots \\ \phi_0(x_n) & \cdots & \phi_m(x_n) \end{bmatrix}^{m} \qquad a = \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix}_m \qquad y = \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix}_n$$

Remark: The matrix $B^T B$ is positive semi-definite and symmetric. If the colums of B are linearly independent $B^T B$ is INVERTIBLE. (and positive definite)

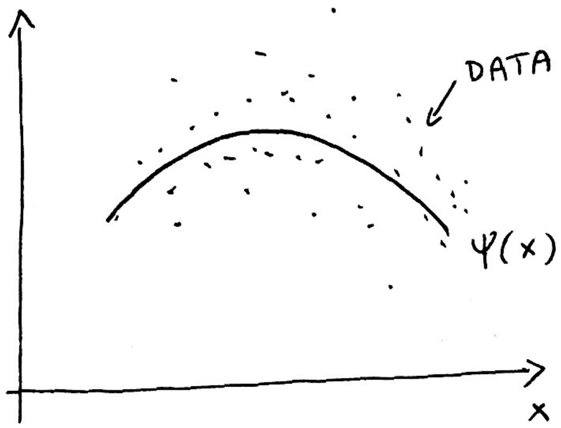
Remark: If $n = m$ and the nodes x_i are distinct then the least-squares approximant is an interpolant. (provided ~~the~~ B is invertible)

Remark: (Example of a low-order polynomial model)

Consider $\phi_0(x) = 1$ $\phi_1(x) = x$ and $\phi_2(x) = x^2$

$$\psi(x) = \sum_{k=0}^{2} a_k \phi_k(x)$$

$$\psi(x) = a_0 + a_1 x + a_2 x^2$$



Note that if we have avalable only 3 data points (x_0, y_0) (x_1, y_1) and (x_2, y_2) (x_i distinct)

then

$$B = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix}$$

(Vandermonde matrix ($\det B \neq 0$))

$$\Rightarrow \quad B^T B a = B^T y \overset{B \text{ invertible}}{\Longleftrightarrow} B a = y \quad \Rightarrow \psi(x) \text{ interpolates the data}$$

(normal equations for least-squares approximation) (interpolation conditions)

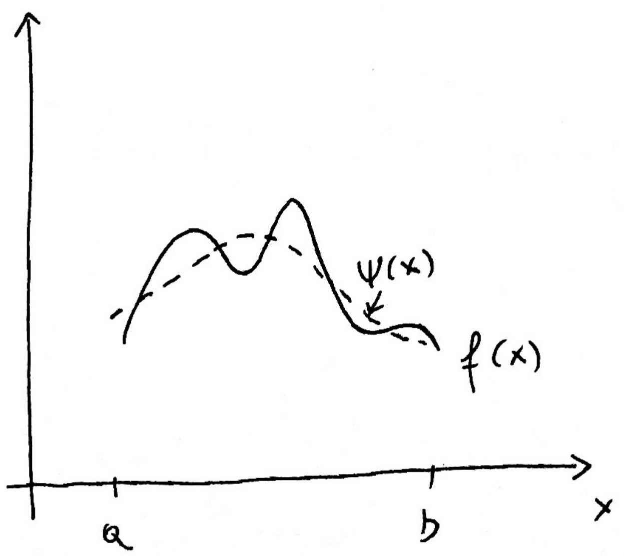
In other words, unisolvent interpolation problems can be computed by least-squares. (if B is invertible)

Least-squares approximation of functions

Consider a function $f(x)$ and an approximant of f in the form

$$\psi(x) = \sum_{k=0}^{m} a_k \phi_k(x)$$

We would like to construct $\psi(x)$ by minimizing the distance between $f(x)$ and $\psi(x)$ in the L_2 norm. To this end, suppose $f(x)$ and $\psi(x)$ are defined in $x \in [a, b]$.



$$(h(x), g(x))_{L_2} = \int_a^b f(x) g(x)\, dx$$

(inner product)

$$\| h(x) \|_{L_2}^2 = \int_a^b h(x)^2\, dx$$

Define the error norm:

$$E(a_0, \dots, a_m) = \| f(x) - \psi(x) \|_{L_2}^2$$

$$= \left\| f(x) - \sum_{K=0}^{m} a_K \phi_K(x) \right\|_{L_2}^2$$

$$= \int_a^b \left(f(x) - \sum_{K=0}^{m} a_K \phi_K(x) \right)^2 dx$$

A necessary condition for a global minimum is:

$$\frac{\partial E}{\partial a_p} = 0 \quad p=0,\dots,m \quad \Rightarrow \quad \sum_{K=0}^{m} (\phi_p, \phi_K)_{L_2} a_K = (f, \phi_p) \quad p=0,\dots,m$$

This yields:

$$\underbrace{\begin{bmatrix} (\phi_0, \phi_0)_{L_2} & \cdots & (\phi_0, \phi_m)_{L_2} \\ & & \\ (\phi_m, \phi_0)_{L_2} & \cdots & (\phi_m, \phi_m)_{L_2} \end{bmatrix}}_{\text{MASS MATRIX}} \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} (f, \phi_0)_{L_2} \\ \vdots \\ (f, \phi_m)_{L_2} \end{bmatrix}$$

(symmetric and positive definite)