S. Boyd and S. Lall

Least-squares data fitting

Least-squares data fitting

we are given:

- functions $f_1, \ldots, f_n : S \to \mathbb{R}$, called *regressors* or *basis functions*
- *data* or *measurements* (s_i, g_i), $i = 1, \ldots, m$, where $s_i \in S$ and (usually) $m \gg n$

problem: find coefficients $x_1, \ldots, x_n \in \mathbb{R}$ so that

$$x_1 f_1(s_i) + \cdots + x_n f_n(s_i) \approx g_i, \quad i = 1, \ldots, m$$

$i.e.$, find linear combination of functions that fits data

least-squares fit: choose x to minimize total square fitting error:

$$\sum_{i=1}^{m} (x_1 f_1(s_i) + \cdots + x_n f_n(s_i) - g_i)^2$$

Least-squares data fitting

- total square fitting error is $\|Ax - g\|^2$, where $A_{ij} = f_j(s_i)$
- hence, least-squares fit is given by

$$x = (A^\mathsf{T} A)^{-1} A^\mathsf{T} g$$

  (assuming A is skinny, full rank)
- corresponding function is

$$f_{\text{lsfit}}(s) = x_1 f_1(s) + \cdots + x_n f_n(s)$$

- applications:
  - interpolation, extrapolation, smoothing of data
  - developing simple, approximate model of data

Least-squares polynomial fitting

problem: fit polynomial of degree $< n$,

$$p(t) = a_0 + a_1 t + \cdots + a_{n-1} t^{n-1},$$

to data (t_i, y_i), $i = 1, \ldots, m$

- basis functions are $f_j(t) = t^{j-1}$, $j = 1, \ldots, n$
- matrix A has form $A_{ij} = t_i^{j-1}$

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ & \vdots & & & \vdots \\ 1 & t_m & t_m^2 & \cdots & t_m^{n-1} \end{bmatrix}$$

(called a *Vandermonde matrix*)

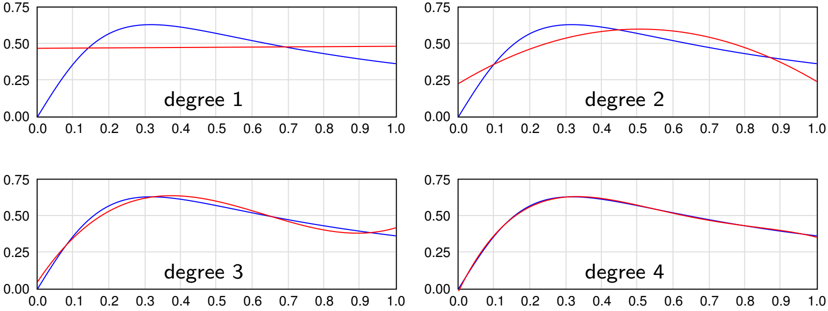
Vandermonde matrices

assuming $t_k \neq t_l$ for $k \neq l$ and $m \geq n$, A is full rank:

- suppose $Aa = 0$
- corresponding polynomial $p(t) = a_0 + \cdots + a_{n-1}t^{n-1}$ vanishes at m points $t_1, \ldots, t_m$
- by fundamental theorem of algebra p can have no more than $n-1$ zeros, so p is identically zero, and $a = 0$
- columns of A are independent, *i.e.*, A full rank

Example

- fit $g(t) = 4t/(1+10t^2)$ with polynomial
- $m = 100$ points between $t = 0$ & $t = 1$
- fits for degrees $1, 2, 3, 4$ have RMS errors $.135, .076, .025, .005$, respectively



Growing sets of regressors

consider *family* of least-squares problems

$$\text{minimize} \quad \left\| \sum_{i=1}^{p} x_i a_i - y \right\|$$

for $p = 1, \ldots, n$

($a_1, \ldots, a_p$ are called *regressors*)

- approximate y by linear combination of $a_1, \ldots, a_p$
- project y onto **span**$\{a_1, \ldots, a_p\}$
- regress y on $a_1, \ldots, a_p$
- as p increases, get better fit, so optimal residual decreases

Growing sets of regressors

solution for each $p \leq n$ is given by

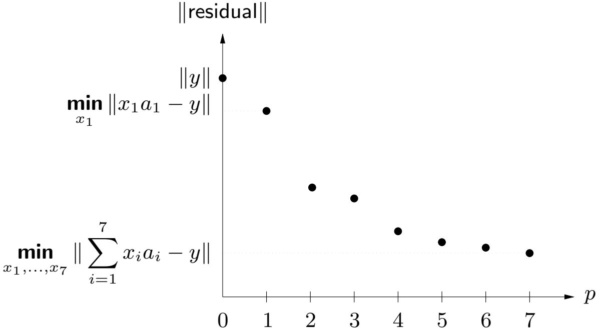
$$x_{\mathrm{ls}}^{(p)} = (A_p^\mathsf{T} A_p)^{-1} A_p^\mathsf{T} y = R_p^{-1} Q_p^\mathsf{T} y$$

where

- $A_p = [a_1 \cdots a_p] \in \mathbb{R}^{m \times p}$ is the first p columns of A
- $A_p = Q_p R_p$ is the QR factorization of A_p
- $R_p \in \mathbb{R}^{p \times p}$ is the leading $p \times p$ submatrix of R
- $Q_p = [q_1 \cdots q_p]$ is the first p columns of Q

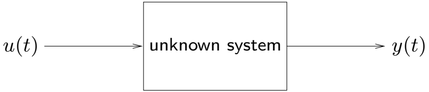
Norm of optimal residual versus p

plot of optimal residual versus p shows how well y can be matched by linear combination of $a_1, \ldots, a_p$, as function of p



Least-squares system identification

we measure input $u(t)$ and output $y(t)$ for $t = 0, \ldots, N$ of unknown system



system identification problem: find reasonable model for system based on measured I/O data u, y

example with scalar u, y (vector u, y readily handled): fit I/O data with moving-average (MA) model with n delays

$$\hat{y}(t) = h_0 u(t) + h_1 u(t-1) + \cdots + h_n u(t-n)$$

where $h_0, \ldots, h_n \in \mathbb{R}$

System identification

we can write model or predicted output as

$$\begin{bmatrix} \hat{y}(n) \\ \hat{y}(n+1) \\ \vdots \\ \hat{y}(N) \end{bmatrix} = \begin{bmatrix} u(n) & u(n-1) & \cdots & u(0) \\ u(n+1) & u(n) & \cdots & u(1) \\ \vdots & \vdots & & \vdots \\ u(N) & u(N-1) & \cdots & u(N-n) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix}$$

model prediction error is

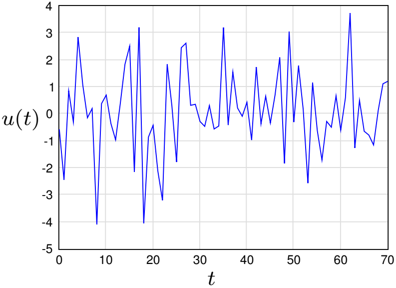
$$e = (y(n) - \hat{y}(n), \ldots, y(N) - \hat{y}(N))$$

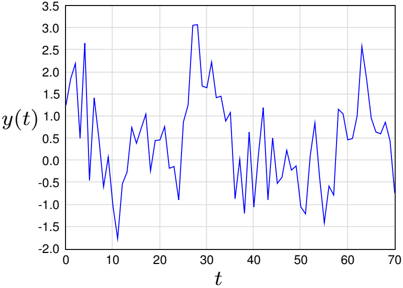
least-squares identification: choose model ($i.e.$, h) that minimizes norm of model prediction error $\|e\|$

. . . a least-squares problem (with variables h)

Example

data used to fit model



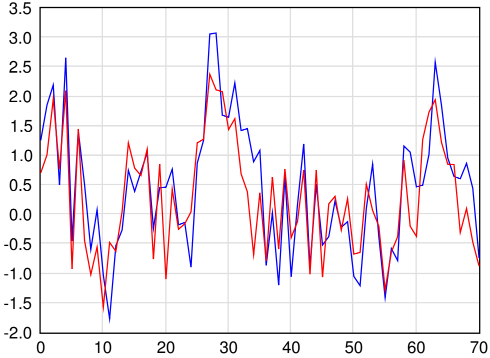


Example

for $n = 7$ we obtain MA model with

$$(h_0, \ldots, h_7) = (.024,\ .282,\ .418,\ .354,\ .243,\ .487,\ .208,\ .441)$$

with relative prediction error $\|e\|/\|y\| = 0.37$



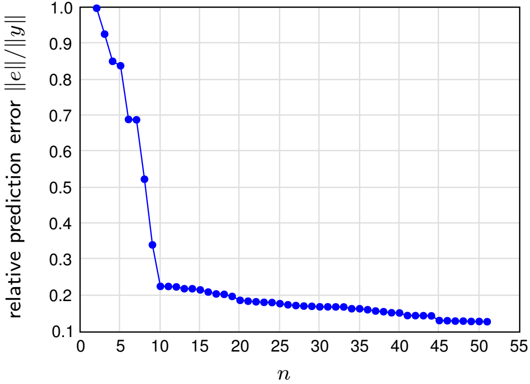
$y(t)$ actual output, $\hat{y}(t)$ predicted from model

Model order selection

question: how large should n be?

- obviously the larger n, the smaller the prediction error *on the data used to form the model*
- suggests using largest possible model order for smallest prediction error

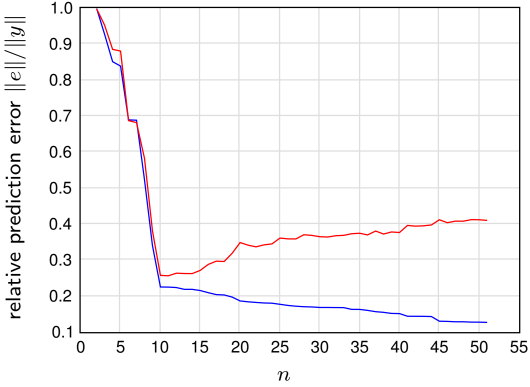
Model order selection



difficulty: for n too large the *predictive ability* of the model on *other I/O data* (from the same system) becomes worse

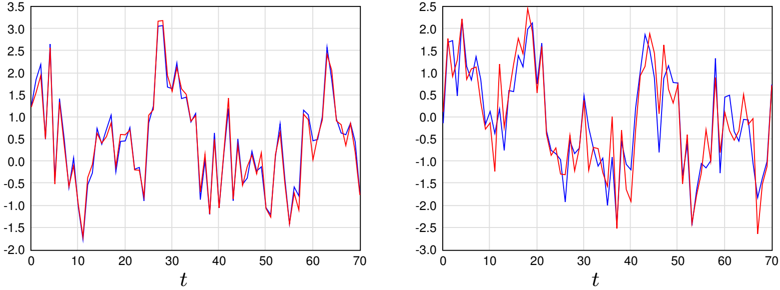
Out of sample validation

- evaluate model predictive performance on another I/O data set *not used to develop model* model validation data set
- check prediction error of models (developed using *modeling data*) on *validation data*
- plot suggests $n = 10$ is a good choice



Validation

for $n = 50$ the actual and predicted outputs on system identification and model validation data are:



- $y(t)$ actual output, $\hat{y}(t)$ predicted from model
- loss of predictive ability when n too large called *model overfit* or *overmodeling*