

Nonlinear Least Squares

Stephen Boyd

EE103
Stanford University

December 6, 2016

Location from range measurements

- ▶ 3-vector x is position in 3-D, which we will estimate
- ▶ *range* measurements give (noisy) distance to known locations

$$\rho_i = \|x - a_i\| + v_i, \quad i = 1, \dots, m$$

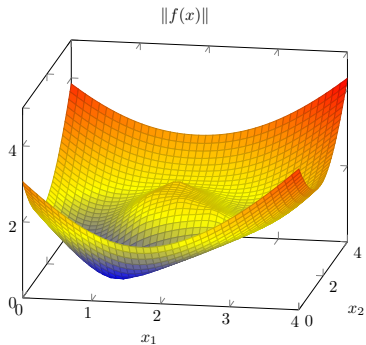
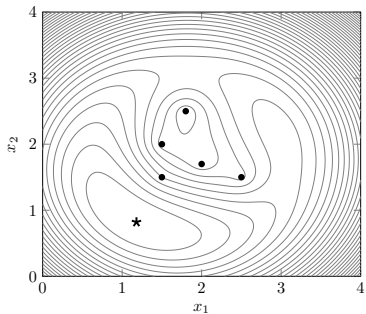
- a_i are known locations
 - v_i are noises
- ▶ least squares location estimation: choose \hat{x} that minimizes

$$\sum_{i=1}^m (\|x - a_i\| - \rho_i)^2$$

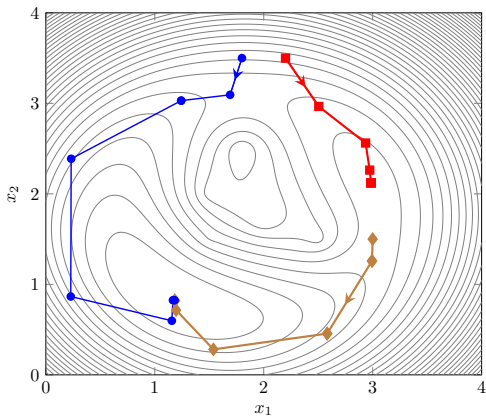
- ▶ GPS works like this

Example: Location from range measurements

range to 5 points (circles)



Levenberg-Marquardt from 3 initial points



Model fitting

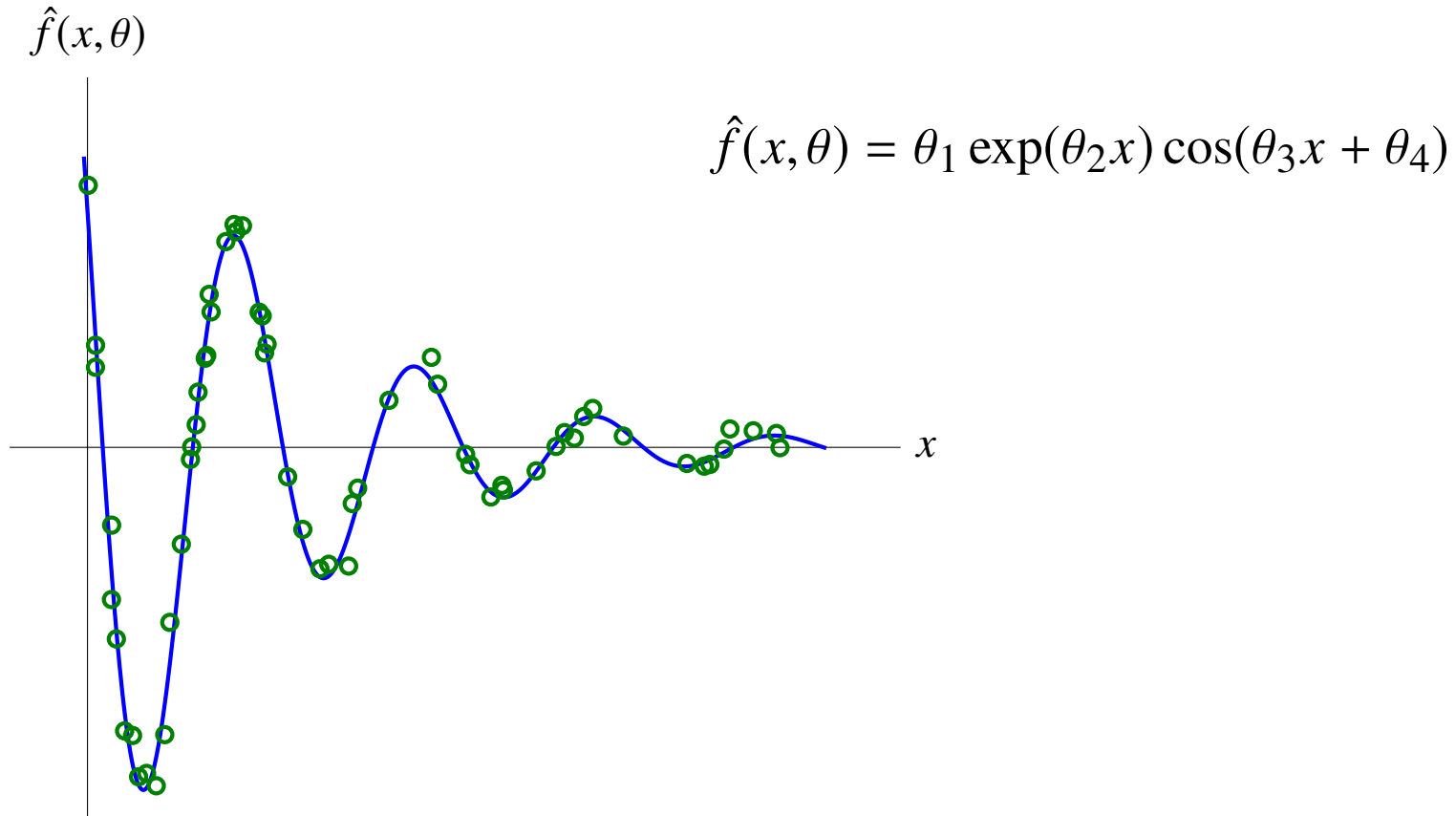
$$\text{minimize } \sum_{i=1}^N (\hat{f}(x^{(i)}, \theta) - y^{(i)})^2$$

- model $\hat{f}(x, \theta)$ is parameterized by parameters $\theta_1, \dots, \theta_p$
- $(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})$ are data points
- the minimization is over the model parameters θ
- on page 9.9 we considered models that are linear in the parameters θ :

$$\hat{f}(x, \theta) = \theta_1 f_1(x) + \dots + \theta_p f_p(x)$$

here we allow $\hat{f}(x, \theta)$ to be a nonlinear function of θ

Example



a nonlinear least squares problem with four variables $\theta_1, \theta_2, \theta_3, \theta_4$:

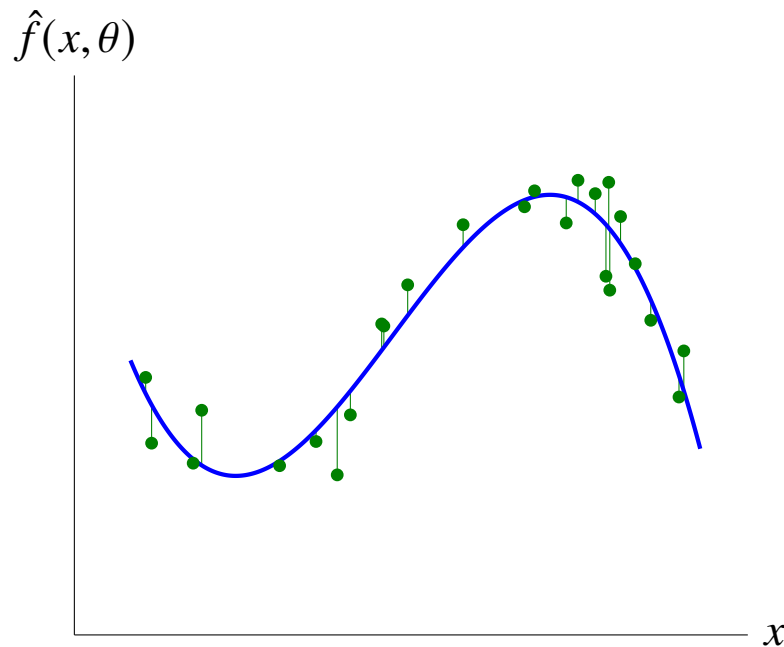
$$\text{minimize } \sum_{i=1}^N \left(\theta_1 e^{\theta_2 x^{(i)}} \cos(\theta_3 x^{(i)} + \theta_4) - y^{(i)} \right)^2$$

Orthogonal distance regression

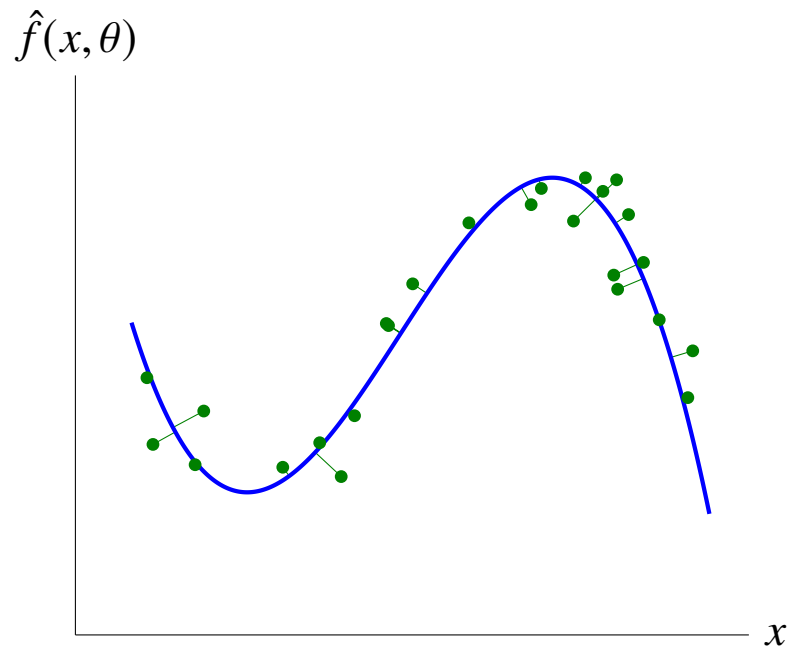
minimize the mean square distance of data points to graph of $\hat{f}(x, \theta)$

Example: orthogonal distance regression with cubic polynomial

$$\hat{f}(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$$



standard least squares fit

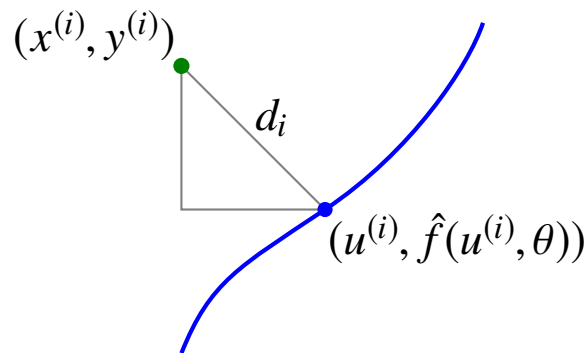


orthogonal distance fit

Nonlinear least squares formulation

$$\text{minimize} \quad \sum_{i=1}^N \left((\hat{f}(u^{(i)}, \theta) - y^{(i)})^2 + \|u^{(i)} - x^{(i)}\|^2 \right)$$

- optimization variables are model parameters θ and N points $u^{(i)}$
- i th term is squared distance of data point $(x^{(i)}, y^{(i)})$ to point $(u^{(i)}, \hat{f}(u^{(i)}, \theta))$



$$d_i^2 = (\hat{f}(u^{(i)}, \theta) - y^{(i)})^2 + \|u^{(i)} - x^{(i)}\|^2$$

- minimizing d_i^2 over $u^{(i)}$ gives squared distance of $(x^{(i)}, y^{(i)})$ to graph
- minimizing $\sum_i d_i^2$ over $u^{(1)}, \dots, u^{(N)}$ and θ minimizes mean squared distance