ECE 4960

Spring 2017

Lecture 16

Nonlinear Equations and Optimization:
Model Parameter Extraction

Edwin C. Kan
School of Electrical and Computer Engineering
Cornell University

Models for the Real World

- To describe our complex world, often a model is created with the independent variables (x_1, x_2, … x_n) and meaningful model parameters (a_1, a_2, …, a_m).
- A "scalar" model S can then be generally expressed as:

$$S_{model} = S(x_1, x_2, \ldots x_n; a_1, a_2, \ldots, a_m)$$

- S can be the current into an electronic block, where x_1, x_2, … are the control input voltages, and a_1, a_2, … are the parameters describing the blocks (e.x., the resistance and capacitance values that determines the time constants how x_1, x_2, … will affect S).
- S can be your stock portfolio value, where x_1, x_2, … are the individual stock prices, and a_1, a_2, … are the shares you have.

Measurements of the Real World

- We can often make measurements of S with given values of the independent variables $(x_1, x_2, \ldots x_n)$, including **noises** and **uncertainties**:

$$S_{measured} = S(x_1, x_2, \ldots x_n)$$

- Notice that for every legitimate $(x_1, x_2, \ldots x_n)$, there can be a $S_{measured}$. We can have thousands of $S_{measured}$ when we sweep the values of $(x_1, x_2, \ldots x_n)$.
- S can be a vector of rank l as well, but we will restrict here for S to be a scalar and we can generalize the case of the S vectors by taking $||S||_2$.

Least-Square Model Parameter Extraction

- The **least-square parameter extraction** to construct the model parameters ($a_1, a_2, \ldots, a_m$) is an optimization problem to minimize the following function *V* that describe the least-square fitting:

$$V = \sum_i \left(S_{i,\mathrm{model}} - S_{i,measured}\right)^2$$

for all *i*-th instances with the same ($x_1, x_2, \ldots x_n$)

- To minimize the error *V* for the model fitting, it is equivalent to solve the nonlinear equations:

$$\frac{\partial V}{\partial a_j} = 0; \qquad \forall j = 1 \ldots m$$

Properties of Least-Square Extraction

- As V is always positive, the Hessian matrix is positive definite and we will always reach a minimum (least in "least square").
- We will have the exact m equations for the m parameters (a_1, a_2, …, a_m), so no **under-determined** or **over-determined** problems.
- As V is nonlinear, even when S is linear, there may be more than one set of solutions.
- The purpose of the model parameter extraction can be:
  - For estimation of S at an instance of (x_1, x_2, … x_n) that we do not have direct measurements.
  - $S_{measurement}$ may not be smooth, which makes taking the numerical derivative extremely noisy. However, S_{model} can be often made C-∞, meaning all derivatives of S are continuous.
  - For example, in power network, we can ask what x can make S to be unstable (by definition, you can not measure the unstable situation.

Least-Square Fitting of a Linear Line

- The model will now be linear, i.e., $y = ax+b$, where x and y are random variables (but they will have noises during measurements) and a and b are parameters.
- Four measurements available (containing errors): (x_0, y_0), (x_1, y_1), (x_2, y_2), and (x_3, y_3)

$$V(a,b) = (ax_0 + b - y_0)^2 + (ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + (ax_3 + b - y_3)^2$$

$$\frac{\partial V(a,b)}{\partial a} = 2x_0(ax_0 + b - y_0) + 2x_1(ax_1 + b - y_1) + 2x_2(ax_2 + b - y_2) + 2x_0(ax_3 + b - y_3) = 0$$

$$\frac{\partial V(a,b)}{\partial b} = 2(ax_0 + b - y_0) + 2(ax_1 + b - y_1) + 2(ax_2 + b - y_2) + 2(ax_3 + b - y_3) = 0$$

Reorganized:

$$f_1 = (x_0^2 + x_1^2 + x_2^2 + x_3^2)a + (x_0 + x_1 + x_2 + x_3)b - (x_0 y_0 + x_1 y_1 + x_2 y_2 + x_3 y_3) = 0$$

$$f_2 = (x_0 + x_1 + x_2 + x_3)a + 4b - (y_0 + y_1 + y_2 + y_3) = 0$$

Least-Square Fitting of a Linear Line

$$[H]_V = [J]_{f1,f2} = \begin{bmatrix} \frac{\partial f_1}{\partial a} & \frac{\partial f_1}{\partial b} \\ \frac{\partial f_2}{\partial a} & \frac{\partial f_2}{\partial b} \end{bmatrix} = \begin{bmatrix} x_0^2 + x_1^2 + x_2^2 + x_3^2 & x_0 + x_1 + x_2 + x_3 \\ x_0 + x_1 + x_2 + x_3 & 4 \end{bmatrix}$$

$$[RHS] = \begin{bmatrix} x_0 y_0 + x_1 y_1 + x_2 y_2 + x_3 y_3 \\ y_0 + y_1 + y_2 + y_3 \end{bmatrix}$$

- The Jacobian matrix does not contain *a* or *b*: one step linear solution!
- Try this out:

x	0.0	1.5	2.2	3.0	4.3	5.2
y	1.1	3.9	5.5	7.5	10.0	11.2

Least-Square Fitting of Power Law

- $y = ax^m$, where x and y are random variables (but they will have noises during measurements) and a and m are parameters.
- Four measurements available (containing errors): (x_0, y_0), (x_1, y_1), (x_2, y_2), and (x_3, y_3)

$$V(a,m) = \left(ax_0^m - y_0\right)^2 + \left(ax_1^m - y_1\right)^2 + \left(ax_2^m - y_2\right)^2 + \left(ax_3^m - y_3\right)^2$$

$$\frac{\partial V(a,m)}{\partial a} = 2x_0^m\left(ax_0^m - y_0\right) + 2x_1^m\left(ax_1^m - y_1\right) + 2x_2^m\left(ax_2^m - y_2\right) + 2x_3^m\left(ax_3^m - y_3\right) = 0$$

$$\frac{\partial V(a,m)}{\partial m} = 2a \cdot \left[\ln x_0 \cdot x_0^m\left(ax_0^m - y_0\right) + \ln x_1 \cdot x_1^m\left(ax_1^m - y_1\right) + \ln x_2 \cdot x_2^m\left(ax_2^m - y_2\right) + \ln x_3 \cdot x_3^m\left(ax_3^m - y_3\right)\right] = 0$$

Reorganized:

$$f_1 = \left(x_0^{2m} + x_1^{2m} + x_2^{2m} + x_3^{2m}\right)a - \left(x_0^m y_0 + x_1^m y_1 + x_2^m y_2 + x_3^m y_3\right) = 0$$

$$f_2 = \left(\ln x_0 \cdot x_0^{2m} + \ln x_1 \cdot x_1^{2m} + \ln x_2 \cdot x_2^{2m} + \ln x_3 \cdot x_3^{2m}\right)a^2 - \left(\ln x_0 \cdot x_0^m y_0 + \ln x_1 \cdot x_1^m y_1 + \ln x_2 \cdot x_2^m y_2 + \ln x_3 \cdot x_0^m y_3\right)a = 0$$

Least-Square Fitting of Power Law

$$[H]_V = [J]_{f1,f2} = \begin{bmatrix} \frac{\partial f_1}{\partial a} & \frac{\partial f_1}{\partial m} \\ \frac{\partial f_2}{\partial a} & \frac{\partial f_2}{\partial m} \end{bmatrix} = \begin{bmatrix} x_0^{2m} + x_1^{2m} + x_2^{2m} + x_3^{2m} & \frac{\partial f_1}{\partial m} \\ 2a\left(\ln x_0 \cdot x_0^{2m} + \ln x_1 \cdot x_1^{2m} + \ln x_2 \cdot x_2^{2m} + \ln x_3 \cdot x_3^{2m}\right) - & \frac{\partial f_2}{\partial m} \\ \left(\ln x_0 \cdot x_0^{m} y_0 + \ln x_1 \cdot x_1^{m} y_1 + \ln x_2 \cdot x_2^{m} y_2 + \ln x_3 \cdot x_0^{m} y_3\right) & \end{bmatrix}$$

- The Jacobian matrix contains *a* or *b*: nonlinear and initial guess is needed!
- Try this out:

x	1.0	4.5	9.0	20	74	181
y	3.0	49.4	245	1808	2.2×10^4	7.3×10^4

A Preliminary Summary of Least Square

- *V* will almost NEVER be 0 or even close to 0, but the value of the eventual V carries the information of how **consistent** the model and the measurement are, and the noises/uncertainties in measurement.
- *V* naturally contains the "weighting", which can be seen in the *x* and log(*x*) examples.
- *V* can be readily normalized in different ways.
- The model can be thought as a "**rule**" or a "**component**" in the real world. Testing of the magnitude of *V* is in different model is an important statistical analysis tool, for example: principal component analysis (PCA) and machine learning.