Data Analysis and Fitting: Modeling Data and Fitting

Ashton S. Reimer

[1]Center for Geospace Studies
SRI International

July 2020

Topics

1. Forward Models
2. Inverse Problems
3. Least-Squares Technique

Forward Models

A forward model, f, predicts observables, $\mathbf{y}$, given parameters, $\mathbf{p}$:

$$\mathbf{y} = f(\mathbf{p})$$

$$\mathbf{p} = p_1, p_2, ..., p_N \qquad \mathbf{y} = y_1, y_2, ..., y_M$$

For example, a forward model can be:

- a polynomial that we want to fit to data
- a physical model relating observations to physical parameters
- ...

For ISR data analysis:

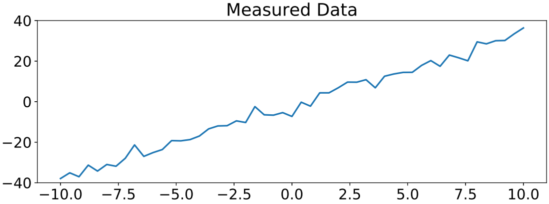
- Given physical parameters (N_e, T_e, T_i, v_{LOS}), ISR theory is part of the forward model we use to predict what the radar will observe.

Parameter Estimation and Inverse Problems

Inverse problem:

- What if we have a forward model, f, and observations, $\mathbf{y}$? How do we get the parameters ($\mathbf{p}$)?

We can try $f^{-1}(\mathbf{y}) = \mathbf{p}$, but what if measurements are noisey:



So a better forward model is $\mathbf{z} = f(\mathbf{p}) + \mathbf{e}$, but how do we invert this?

Least-Squares Estimation

For M data points, z_m, with independent measurement errors, σ_m, compute the "chi-square"; an error weighted difference between the data and the model, f_m:

$$\chi^2(\mathbf{p}) = \sum_{m=1}^{M} \frac{[z_m - f_m(\mathbf{p})]^2}{\sigma_m^2}$$

the model parameters that provide the "best fit" of the model to the data, $\hat{\mathbf{p}}_{LS}$, are those that minimizes $\chi^2(\mathbf{p})$: $\underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \sum_{m=1}^{M} \frac{[z_m - f_m(\mathbf{p})]^2}{\sigma_m^2} \right\}$

In general, measurements $\mathbf{z}$ may not be independent. The generalized least-squares estimate is:

$$\chi^2(\mathbf{p}) = [\mathbf{z} - f(\mathbf{p})]^T \, \mathbf{\Sigma_e}^{-1} \, [\mathbf{z} - f(\mathbf{p})]$$

where $\mathbf{\Sigma_e}$ is the covariance matrix of measurements $\mathbf{z}$.

Least-Squares Estimation from Maximum Likelihood

What is the **likelihood** of obtaining the data given the parameters?

$$P(\mathbf{z}|\mathbf{p}) \propto \prod_{m=1}^{M} \exp\left(-\frac{1}{2}\frac{[z_m - f_m(\mathbf{p})]^2}{\sigma_m^2}\right)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{m=1}^{M}\frac{[z_m - f_m(\mathbf{p})]^2}{\sigma_m^2}\right)$$

The "most likely" parameters $\hat{\mathbf{p}}_{ML}$ maximize $P(\mathbf{z}|\mathbf{p})$:

$$\hat{\mathbf{p}}_{ML} : \quad \underset{\mathbf{p}}{\operatorname{argmax}} \left\{P(\mathbf{z}|\mathbf{p})\right\}$$

$$\hat{\mathbf{p}}_{ML} : \quad \underset{\mathbf{p}}{\operatorname{argmax}} \left\{\log P(\mathbf{z}|\mathbf{p})\right\}$$

$$\hat{\mathbf{p}}_{ML} : \quad \underset{\mathbf{p}}{\operatorname{argmin}} \left\{\sum_{m=1}^{M}\frac{[z_m - f_m(\mathbf{p})]^2}{\sigma_m^2}\right\}$$

Least-Squares Estimation

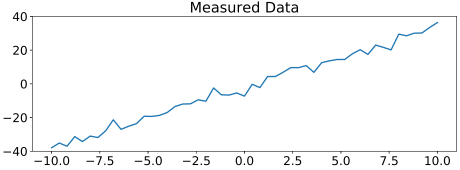
Some important terminology/concepts:

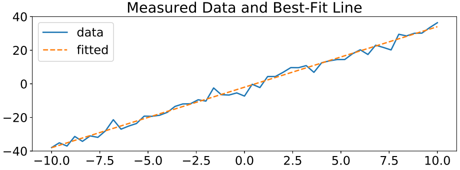
- $\frac{z_m - f_m(\mathbf{p})}{\sigma_m}$ are known as the "normalized errors". The numerator is called the "residual".
- assumed that each "normalized error" is normally distributed with zero mean and unit variance: $\mathcal{N}(0, 1)$
- the chi-square, χ^2, is the sum of the square of $\mathcal{N}(0, 1)$ random variables, so by definition, χ^2 is a chi-squared distributed random variable, with $M - N$ degrees of freedom.

Example: Fitting a Linear Model to Data

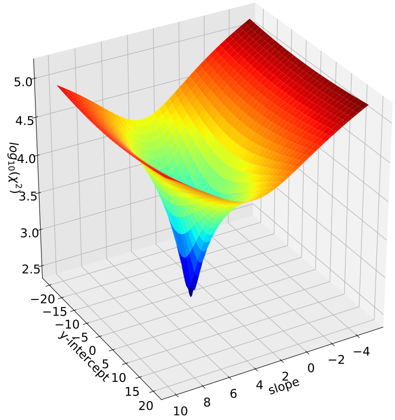
Given a Model:

$$y = mx + b$$





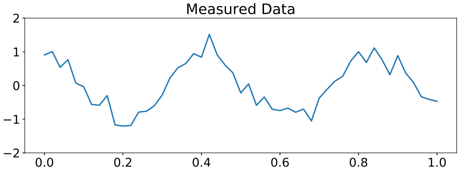
Calculate $\chi^2(m, b)$:

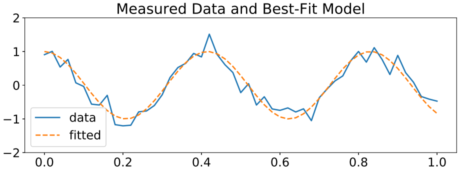


Example: Fitting a Non-Linear Model to Data

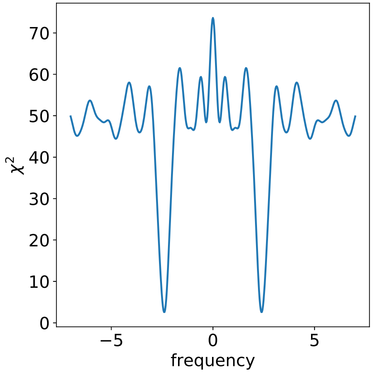
Given a Model:

$$y = \cos(2\pi f t)$$





Calculate $\chi^2(f)$:



Least-Squares Estimation

For some simple models (linear, quadratic, etc):

- analytic solutions exist for the minimum χ^2

In general, non-linear least squares algorithms are required:

- Levenberg-Marquardt (LM) algorithm is most commonly used
- LM requires a good initial guess
- Standard LM packages:
  - FORTRAN: MINPACK lmdif.f and lmder.f
  - Python: scipy.optimize.leastsq (wrapper around lmdif and lmder)
  - Matlab: Optimization Toolbox lsqnonlin
  - IDL: LMFIT

Summary

- A forward model is a function that predicts measurements given input (physical) parameters
- Solving the inverse problem:
  - Given: measurements, measurement errors, and forward model
  - How do we solve for the model parameters?
- Least-squares can be used to solve inverse problems:
  - chi-squared: sum of the error weighted differences between the data and the forward model
  - "best-fit" model parameters are those that minimize chi-squared

Next topic:

- Is a fit meaningful?
- What is the confidence in the fit (uncertainty)?