Data Analysis and Fitting: Modeling Data and Fitting

Ashton S. Reimer

[1]Center for Geospace Studies
SRI International

July 2021

Topics

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

Important Concepts

- **Measurements**: directly recorded by an instrument, e.g.:
  - mercury increasing in volume within a thermometer
  - voltage samples from an analog to digital converter
- **Observations**: estimated from measurements using physics:
  - thermal expansion
  - IS Radar theory
- A **forward model**, f, contains the physics to predict measurements, y, given observations, p:
  - $\mathrm{y} = f(\mathrm{p})$
  - $\mathrm{p} = p_1, p_2, ..., p_N \qquad \mathrm{y} = y_1, y_2, ..., y_M$
- The **inverse problem**: With f and y, how do we obtain p?
  - It is rarely as simple as: $f^{-1}(\mathrm{y}) = \mathrm{p}$
  - Measurements are noisy: $\mathrm{y} = f(\mathrm{p}) + \mathrm{e}$
  - This usually requires making assumptions!

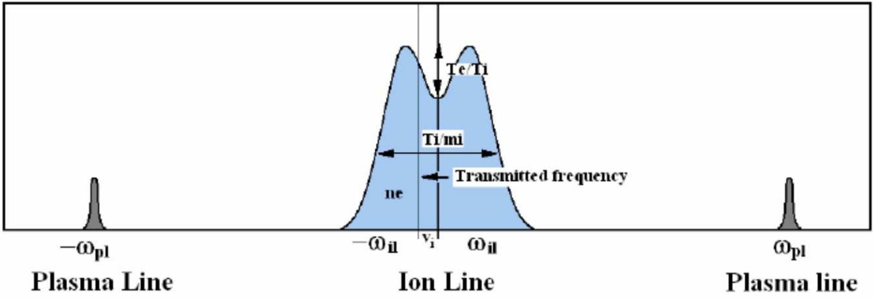
IS Radar Forward Model: Theory

IS Radar theory predicts the statistical properties of the Power Spectrum of the scattered signal:

$$\langle |n_e(\mathrm{k},\omega)|^2 \rangle \overset{\mathcal{F}}{\longleftrightarrow} \langle E_s(t) E_s^*(t-\tau) \rangle$$

as a function of plasma parameters: N_e, T_e, T_i, v_{LOS}.

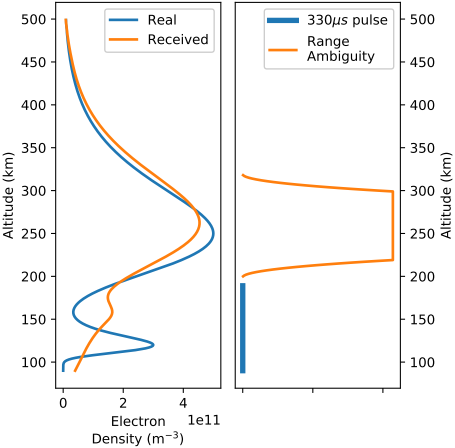
Forward Model? $f(N_e, T_e, T_i, v_{LOS}) = \langle E_s(t) E_s^*(t-\tau) \rangle$



IS Radar Forward Model: Ambiguity

Ambiguity:
How we measure causes "blurring" in space & time

- $W_{t\tau}$: "range-lag ambiguity function"

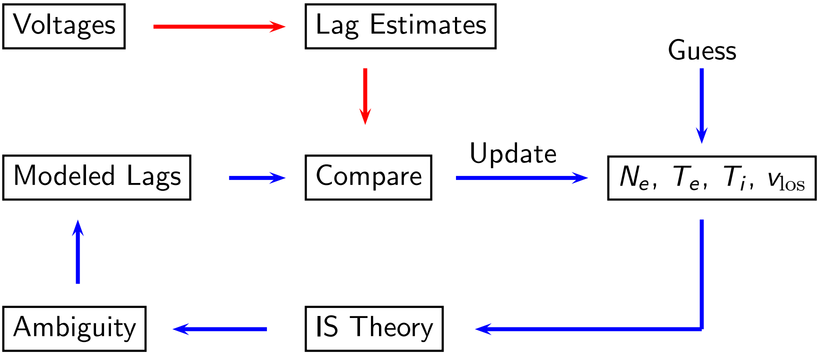


Forward model:

$$f(N_e, T_e, T_i, v_{LOS}, W_{t\tau}) = \langle E_s(t) E_s^*(t-\tau) \rangle$$

IS Radar Inverse Problem

Compare measurements and modeled measurements:

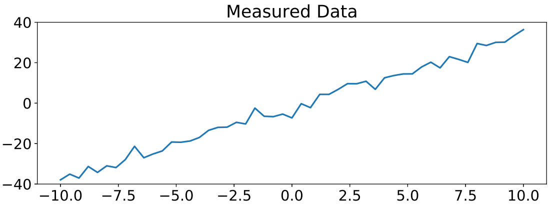


Measurement Uncertainty

Inverse problem:

- What if we have a forward model, f, and measurements, y? How do we get the observables, p?

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



Perhaps a better forward model is $\mathrm{y} = f(\mathrm{p}, \mathrm{e}) = g(\mathrm{p}) + \mathrm{e}$, but how do we invert this?

Least-Squares Estimation

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

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

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

In general, measurements y_m may not be independent. The generalized least-squares estimate is:

$$\chi^2(\mathrm{p}) = [\mathrm{y} - f(\mathrm{p})]^T \, \Sigma_{\mathrm{e}}{}^{-1} \, [\mathrm{y} - f(\mathrm{p})]$$

where Σ_{e} is the covariance matrix of measurements y.

Least-Squares Estimation

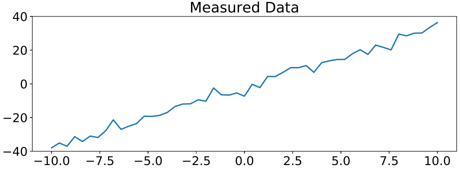
Important terminology and concepts:

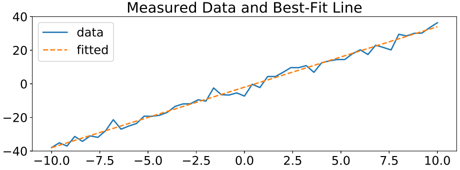
- $\frac{y_m - f_m(\mathrm{p})}{\sigma_m}$ are known as the "normalized errors". The numerator is called the "residual".
- assumed that each "normalized error" is Gaussian distributed with zero mean and unit variance: $\mathcal{N}(0, 1)$
- the chi-squared, χ^2, is the sum of the square of $\mathcal{N}(0, 1)$ 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

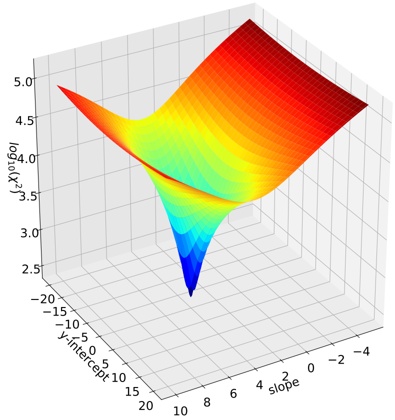
Given a Model:

$$y = mx + b$$





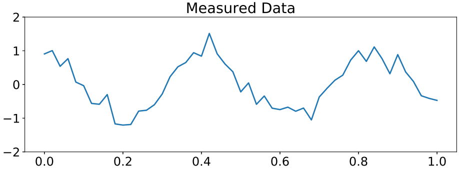
Calculate $\chi^2(m, b)$:

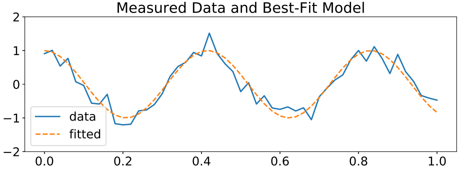


Example: Fitting a Non-Linear Model to Data

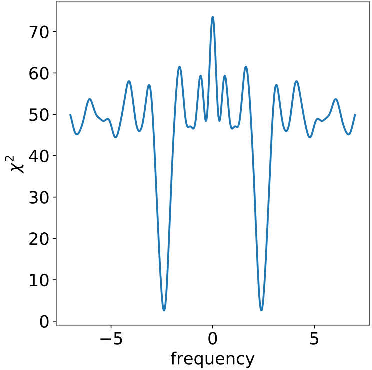
Given a Model:

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





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/observable) 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 measurement and the forward model predictions
  - "best-fit" model parameters are those that minimize chi-squared

Next topic:

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