Statistical Analysis of an Experiment in Physical Chemistry

Ralph Mansson, Alan Welsh, Jeremy Frey, Lefteris Danos
Departments of Mathematics and Chemistry, University of Southampton

YSM2003 - April 15, 2003

Summary

- Introduction to Second Harmonic Generation experiments.
- Theoretical models for SHG data.
- Questions of interest for the experimenters.
- Conclusions and future directions of work.

Second Harmonic Generation Experiments I

Second Harmonic Generation is a technique to study behaviour at interfaces. Polarised light is bounced off the surface of a solution, and the intensity of the second harmonic is measured in volts. The aim is to study the dependence of intensity on polarisation angles.

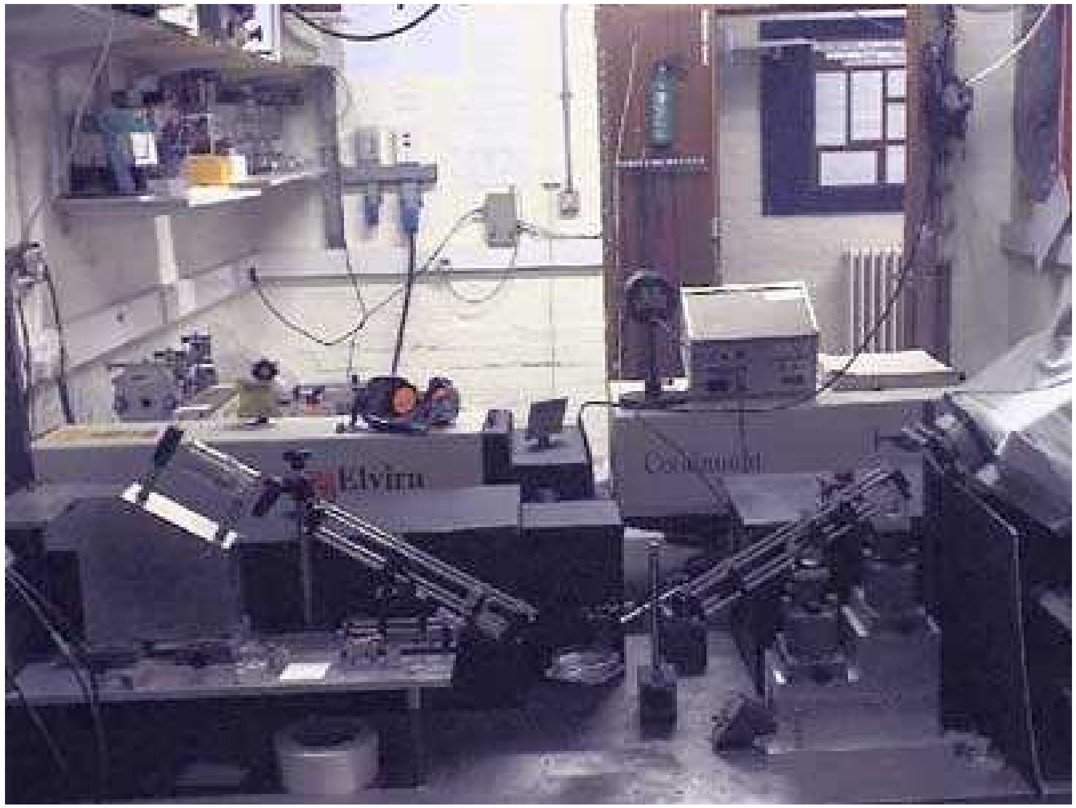
Response: Intensity (voltage).

Factors: Input (γ) and Output (Γ) polarisation angles.

Model: Nonlinear model based on three coefficients A, B, and C which may be complex. Polarisation angles occur in the model in trigonometric functions.

Constraint: Overall phase of the experiment is not determined, so one of A, B, and C is real. The choice of constraint will affect the parameters, but not the ratios of these parameters.

Second Harmonic Generation Experiments II



Modelling SHG data

A theoretical model relates the observed intensity of the second harmonic to the polarisation angle of the fundamental and harmonic beams. The observed intensity $\mathbf{Y_{kj}}$ is modelled by $\mathbf{E(Y_{kj}) = |E_{kj}|^2}$ where

$$\mathbf{E_{kj} = \{C\sin(2\boldsymbol{\gamma}_j)\}\sin(\Gamma_k) + \{A\cos^2(\boldsymbol{\gamma}_j) + B\sin^2(\gamma_j)\}\cos(\Gamma_k)}.$$

The unknown coefficients A, B, and C may be complex, and are estimated from experimental data. There are two parametrisation for this model that have been considered and will be compared later.

Analysis of SHG Experiments

The theoretical model has six parameters, but they are not all identifiable so a constraint is introduced (such as $\mathbf{A} \in \Re$) which leaves five parameters. The choice of parametrisation gives two sets of parameters,

$$(\mathbf{r_a}, \mathbf{r_b}, \mathbf{r_c}, \boldsymbol{\phi}_\mathbf{b}, \boldsymbol{\phi}_\mathbf{c}) \qquad \text{and} \qquad (\mathbf{x_a}, \mathbf{x_b}, \mathbf{x_c}, \mathbf{y_b}, \mathbf{y_c}).$$

Parameters are estimated using nonlinear least squares with sensible starting values. Profile plots can be constructed to investigate the linear approximation.

Diagnostic plots are used to investigate the fit of the model to a given set of data. Fitted versus residuals and normal probability plots to assess model assumptions.

Form of Complex Numbers

- Euler - $\mathbf{A} = \mathbf{r_a} \exp(\mathbf{i}\boldsymbol{\phi}_\mathbf{a})$
- Real and Imaginary - $\mathbf{A} = \mathbf{x_a} + \mathbf{i}\mathbf{y_a}$

When we map between the two forms, we get the same information - estimates of the parameters $(\mathbf{r_a}, \mathbf{r_b}, \mathbf{r_c}, \boldsymbol{\phi}_\mathbf{b}, \boldsymbol{\phi}_\mathbf{c})$ and their standard errors are the same.

The Euler representation is better for dealing with complex numbers in this experiment. There is less curvature in the profile plots of the parameter estimates (magnitudes and phase angles), and they are easier to use when considering ratios of the coefficients. For example,

$$\frac{A}{B} = \frac{r_a}{r_b} \exp(i(\phi_a - \phi_b)) \qquad or \qquad \frac{A}{B} = \frac{x_a + iy_a}{x_b + iy_b}$$

Parameter Estimation

There are three output angles, two of these simplify the model.

- S-polarised : one parameter, $\mathbf{f}(\mathbf{r_c})$.
- P-polarised : three parameters, $\mathbf{f}(\mathbf{r_a}, \mathbf{r_b}, \boldsymbol{\phi}_\mathbf{b})$.
- (often 45°) : all five parameters, $\mathbf{f}(\mathbf{r_a}, \mathbf{r_b}, \mathbf{r_c}, \boldsymbol{\phi}_\mathbf{b}, \boldsymbol{\phi}_\mathbf{c})$.

We can use all the data to estimate the parameters simultaneously, or we can use the data from different output angles to estimate $\{\mathbf{r_c}\}$, $\{\mathbf{r_a}, \mathbf{r_b}, \boldsymbol{\phi}_\mathbf{b}\}$, and $\{\boldsymbol{\phi}_\mathbf{c}\}$ separately. Our advice is that the simultaneous approach is better due to less bias (a better fit) and greater efficiency (standard errors are smaller).

Ratios and Inference

The main quantities of interest from an experiment are ratios of the coefficients, such as $\mathbf{A}/\mathbf{B}$. When the Euler representation of complex numbers is used for coefficients, the ratios can be divided into a ratio of the individual magnitudes and the difference between the phase angles.

To compare ratios estimated from different experiments (to investigate whether they are constant) we need to construct confidence intervals for the ratios. We can use Fieller's method for calculating intervals, or we can estimate standard errors by

- Taylor series approximation,
- Profile intervals based on the likelihood function, or
- Bootstrapping based on 1000 samples.

Example - Phenylalanine I

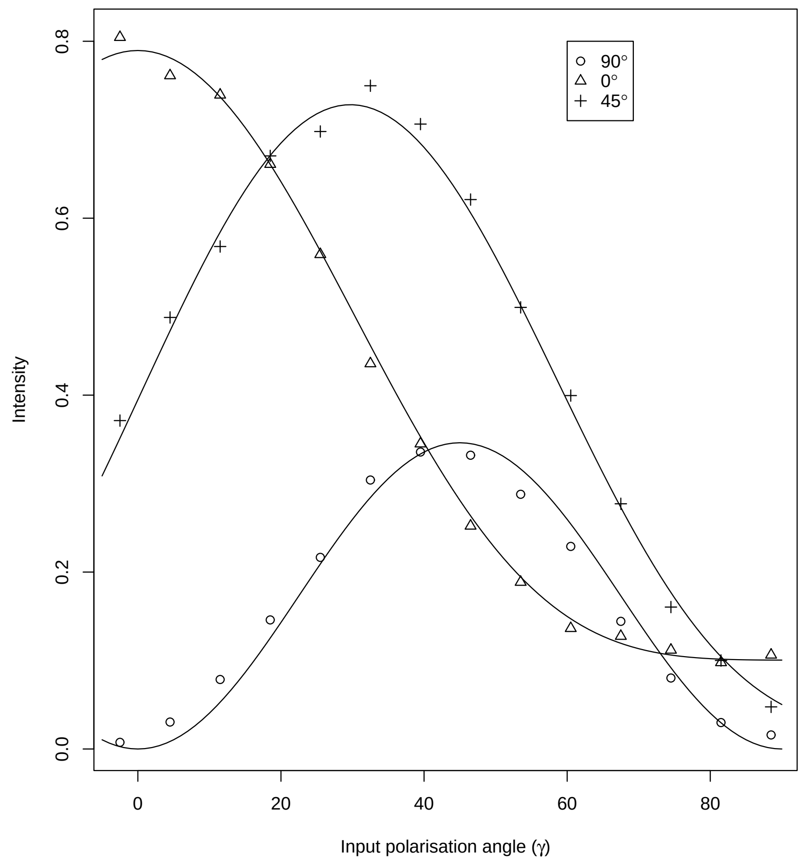
Parameter estimates for the magnitudes and angles for this data are

$$(\mathbf{r_a}, \mathbf{r_b}, \mathbf{r_c}, \boldsymbol{\phi}_\mathbf{b}, \boldsymbol{\phi}_\mathbf{c}) = (0.889, 0.317, 0.588, 1.142, 0.301)$$

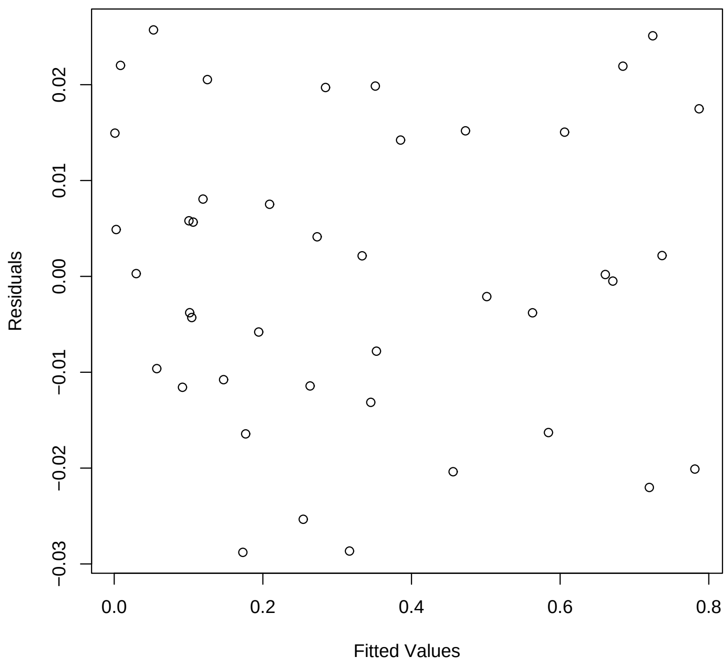
The magnitude of A is the largest so we will consider ratios B/A and C/A. Confidence intervals for the ratio of magnitudes in these two cases based on the different approaches are

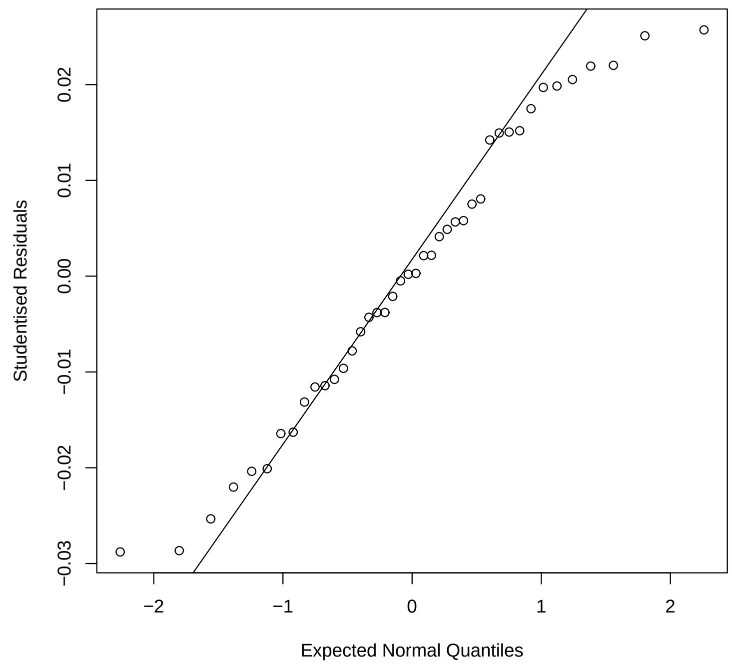
Ratio	Delta Method	Profile	Bootstrap	Fieller
B/A	$(0.326, 0.388)$	$(0.325, 0.389)$	$(0.327, 0.383)$	$(0.326, 0.388)$
C/A	$(0.648, 0.676)$	$(0.649, 0.676)$	$(0.651, 0.675)$	$(0.649, 0.675)$

Example - Phenylalanine II



Example - Phenylalanine III





Example - Toluene

There are multiple sets of data from experiments run under the same conditions. Ratios of coefficients should be similar for models fitted to these experiments. Confidence intervals for the ratios of magnitudes for four experiments using Toluene are

Experiment Number	B/A	C/A
1	$(0.144, 0.387)$	$(0.764, 0.849)$
2	$(0.149, 0.263)$	$(0.543, 0.574)$
3	$(0.165, 0.342)$	$(0.835, 0.890)$
4	$(0.236, 0.323)$	$(0.577, 0.647)$

There is no consistency for the angles of B and C between the four experiments. As a result it is not possible to fit a common model to the data sets and use these common parameter estimates for further calculations.

Future Research

Design of the Experiment

- Number of observations and choice of polarisation angles.
- Effect of parameter estimates on the *locally optimal* design.
- Location of support points for design under different models.

Analysis of the Experiment

- Estimate the parameters of the second-order susceptibility tensor directly.
- Other models including temperature, concentration, etc.