Week 10

More Issues in Non-linear Fitting
**Generic Fitting Algorithm**

1. Load $\{x_i, y_i, \sigma_i\}_N$
2. Select initial values for $\{a_j\}_n$
   - Set $\chi^2_{old} = \text{huge}$
3. Evaluate $\chi^2_{new}$
4. If $\Delta \chi^2 < 0.01$
   - Declare Victory! Compute $\sigma_{a_j}$'s
   - No
   - Compute $\delta a_j$'s
   - Update $a_j$'s using $\delta a_j$'s
5. Set $\chi^2_{old} = \chi^2_{new}$

Methods:
- Grid Search
- Gradient Search
- Expansion Method
- Marquardt’s Method

By hand or by Monte Carlo
Gradient Method pay-off:

**FIGURE 8.4**
Contour plot of $\chi^2$ as a function of two highly correlated variables. The zigzag line represents the search path approach to a local minimum by the grid-search method.
Gradient Search (Bevington 8.4; pp.153)

• Find the “steepest” descent down the valley of $\chi^2$ by computing the normalized gradient
  - Gradient always computable numerically
  - Sometimes can be computed algebraically

• Pro’s and Con’s:
  - Fast approach to the minimum from far away
  - Very slow near the minimum where the gradient is near zero
Summary from last time

- Linear vs. non-linear fitting functions
  - Linear functions can be fitted in one algebraic iteration
  - Non-linear function require iterated solutions when fitting to data
- Grid search is simple but tends to be slow
- Gradient search is fast when far away from minimum but slow near the minimum
- We need a method that is fast near the minimum...next time
Expansion Method (cf. Bevington 8.5; pp156)

Exploit the quadratic nature of $\chi^2$ near its minimum. Expand to 2nd order around whatever (nonideal) point in parameter space we are sitting, where $\chi^2 = \chi^2_0$:

$$
\chi^2 = \chi^2_0 + \sum_{j=1}^{m} \left\{ \frac{\partial \chi^2_0}{\partial a_j} \delta a_j \right\} + \frac{1}{2} \sum_{k=1}^{m} \sum_{j=1}^{m} \left\{ \frac{\partial^2 \chi^2_0}{\partial a_j \partial a_k} \delta a_j \delta a_k \right\} \quad \text{(Bev. Eq. 8.22)}
$$

Minimize $\chi^2$ with respect to the parameter increments $\delta a_j$, not the parameters themselves. Why? Because this linearizes the problem and it can be solved exactly.

$$
\frac{\partial \chi^2}{\partial (\delta a_k)} = \frac{\partial \chi^2_0}{\partial a_k} + \sum_{j=1}^{m} \left\{ \frac{\partial^2 \chi^2_0}{\partial a_k \partial a_j} \delta a_j \right\} = 0 \quad k = 1, m
$$

This is a set of $m$ linear equations in $m$ unknowns!

$$
\beta_k = -\frac{1}{2} \frac{\partial \chi^2_0}{\partial a_k} \quad \alpha_{jk} = \frac{1}{2} \frac{\partial^2 \chi^2_0}{\partial a_j \partial a_k}
$$

"Curvature Matrix" $\alpha$ measures curvature of $\chi^2$ hypersurface
Expansion Method

Matrix equation to solve:
\[ \vec{\beta} = \delta \vec{a} \, \vec{\alpha} \]

Invert the matrix \( \alpha \) to get the matrix \( \varepsilon \), the so-called "error matrix" \( \varepsilon = \alpha^{-1} \):

\[ \delta \vec{a} = \vec{\beta} \, \varepsilon \]

\[ \delta a_k = \sum_{j=1}^{m} \varepsilon_{kj} \beta_j \]

How to compute the matrix \( \alpha \) and the vector \( \beta \)?

\[ \beta_k = \sum \frac{1}{\sigma_i^2} \left[ y_i - y_{fit}(x_i) \right] \frac{\partial y_{fit}(x_i)}{\partial a_k} = -\frac{1}{2} \frac{\partial \chi_0^2}{\partial a_k} \]

\[ \alpha_{jk} = \sum \frac{1}{\sigma_i^2} \left\{ \frac{\partial y_{fit}(x_i)}{\partial a_j} \frac{\partial y_{fit}(x_i)}{\partial a_k} - \left[ y_i - y_{fit}(x_i) \right] \frac{\partial^2 y_{fit}(x_i)}{\partial a_j \partial a_k} \right\} \]

\[ = \frac{1}{2} \frac{\partial^2 \chi_0^2}{\partial a_j \partial a_k} \]

Small compared to first term; vanishes for linear functions; neglect this term

Product of first derivatives
The “error matrix”

\[ \alpha^{-1} = \mathbf{\varepsilon} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots \\ \sigma_{31}^2 & \cdots & \sigma_{33}^2 \\ \vdots & \vdots & \ddots \end{pmatrix} \]

\( \sigma_{jj} = "\text{diagonal errors}" \equiv \sigma_j \)

\( \sigma_{jk} = "\text{correlated errors}" \); \( \sigma_{jk}^2 = "\text{covariances}" \)

Example...
Marquardt(-Levenberg) Method
(cf. Bevington 8.6; pp161)

\[ \beta = \delta a \alpha' \quad \text{with} \quad \alpha'_{jk} = \begin{cases} \alpha_{jk}(1 + \lambda) & \text{for } j = k \\ \alpha_{jk} & \text{for } j \neq k \end{cases} \]

- For small \( \lambda \) it behaves like the expansion method.
- For large \( \lambda \), fit behaves like the gradient method with small steps, since off-diagonals are negligible.
- Make \( \lambda \) just large enough such that \( \chi^2 \) decreases.

The "recipe" for fast fit convergence:

1. Compute \( \chi^2(a) \).
2. Start initially with \( \lambda = 0.001 \).
3. Compute \( \delta a \) and \( \chi^2(a + \delta a) \) with this choice of \( \lambda \).
4. If \( \chi^2(a + \delta a) > \chi^2(a) \), increase \( \lambda \) by a factor of 10 and repeat step 3.
5. If \( \chi^2(a + \delta a) < \chi^2(a) \), decrease \( \lambda \) by a factor of 10, consider \( a' = a + \delta a \) to be the new starting point, and return to step 3, substituting \( a' \) for \( a \).
Marquardt Method Fit

- Red = "true" parent
- Blue = initial guess

Iteration 0
# Numbers...

\[ N(t) = N_0 e^{-t/\tau} + B \]

<table>
<thead>
<tr>
<th>Trial</th>
<th>( \chi^2/\text{d.o.f.} )</th>
<th>( N_0 )</th>
<th>( \tau )</th>
<th>( B )</th>
<th>Value</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.364</td>
<td>150.0</td>
<td>20.0</td>
<td>0.0</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8.030</td>
<td>69.9</td>
<td>6.4</td>
<td>47.5</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.4</td>
<td>1.8</td>
<td>4.3</td>
<td>Uncertainty</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.841</td>
<td>100.6</td>
<td>8.1</td>
<td>29.1</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.0</td>
<td>1.0</td>
<td>1.2</td>
<td>Uncertainty</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.796</td>
<td>103.4</td>
<td>7.2</td>
<td>29.3</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.6</td>
<td>0.8</td>
<td>1.3</td>
<td>Uncertainty</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.795</td>
<td>102.9</td>
<td>7.4</td>
<td>29.1</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2</td>
<td>0.8</td>
<td>1.2</td>
<td>Uncertainty</td>
<td></td>
</tr>
</tbody>
</table>
Marquardt Method Fit

Iteration 1
red = “true” parent
blue = fit result
### Numbers...

\[
N(t) = N_0 e^{-t/\tau} + B
\]

<table>
<thead>
<tr>
<th>Trial</th>
<th>(\chi^2/d.o.f)</th>
<th>(N_0)</th>
<th>(\tau)</th>
<th>(B)</th>
<th>Value/Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.364</td>
<td>150.0</td>
<td>20.0</td>
<td>0.0</td>
<td>Value</td>
</tr>
<tr>
<td>1</td>
<td>8.030</td>
<td>69.9</td>
<td>6.4</td>
<td>47.5</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>1.8</td>
<td>4.3</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>2</td>
<td>0.841</td>
<td>100.6</td>
<td>8.1</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>1.0</td>
<td>1.2</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>3</td>
<td>0.796</td>
<td>103.4</td>
<td>7.2</td>
<td>29.3</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>0.8</td>
<td>1.3</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>4</td>
<td>0.795</td>
<td>102.9</td>
<td>7.4</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>8.2</td>
<td>0.8</td>
<td>1.2</td>
<td></td>
<td>Uncertainty</td>
</tr>
</tbody>
</table>

Carnegie Mellon University Physics
Marquardt Method Fit

Iteration 2

red = “true” parent
blue = fit result
### Numbers...

\[
N(t) = N_0 e^{-t/\tau} + B
\]

<table>
<thead>
<tr>
<th>Trial</th>
<th>$\chi^2$/d.o.f.</th>
<th>$N_0$</th>
<th>$\tau$</th>
<th>$B$</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.364</td>
<td>150.0</td>
<td>20.0</td>
<td>0.0</td>
<td>Value</td>
</tr>
<tr>
<td>1</td>
<td>8.030</td>
<td>69.9</td>
<td>6.4</td>
<td>47.5</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.4</td>
<td>1.8</td>
<td>4.3</td>
<td>Uncertainty</td>
</tr>
<tr>
<td>2</td>
<td>0.841</td>
<td>100.6</td>
<td>8.1</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.0</td>
<td>1.0</td>
<td>1.2</td>
<td>Uncertainty</td>
</tr>
<tr>
<td>3</td>
<td>0.796</td>
<td>103.4</td>
<td>7.2</td>
<td>29.3</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.6</td>
<td>0.8</td>
<td>1.3</td>
<td>Uncertainty</td>
</tr>
<tr>
<td>4</td>
<td>0.795</td>
<td>102.9</td>
<td>7.4</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2</td>
<td>0.8</td>
<td>1.2</td>
<td>Uncertainty</td>
</tr>
</tbody>
</table>
Marquardt Method Fit

Iteration 3
red = "true" parent
blue = fit result

Counts
100
90
80
70
60
50
40
30
20
10
0
Time (arbitrary units)
0
10
20
30
40
50
### Numbers...

\[ N(t) = N_0 e^{-t/\tau} + B \]

<table>
<thead>
<tr>
<th>Trial</th>
<th>$\chi^2$/d.o.f.</th>
<th>$N_0$</th>
<th>$\tau$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.364</td>
<td>150.0</td>
<td>20.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>8.030</td>
<td>69.9</td>
<td>6.4</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.4</td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>0.841</td>
<td>100.6</td>
<td>8.1</td>
<td>29.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>0.796</td>
<td>103.4</td>
<td>7.2</td>
<td>29.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7.6</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>0.795</td>
<td>102.9</td>
<td>7.4</td>
<td>29.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>


Marquardt Method Fit

Iteration 4

red = “true” parent
blue = fit result
### Numbers...

\[ N(t) = N_0 e^{-t/\tau} + B \]

<table>
<thead>
<tr>
<th>Trial</th>
<th>( \chi^2/d.o.f. )</th>
<th>( N_0 )</th>
<th>( \tau )</th>
<th>( B )</th>
<th>Value/Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.364</td>
<td>150.0</td>
<td>20.0</td>
<td>0.0</td>
<td>Value</td>
</tr>
<tr>
<td>1</td>
<td>8.030</td>
<td>69.9</td>
<td>6.4</td>
<td>47.5</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>1.8</td>
<td>4.3</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>2</td>
<td>0.841</td>
<td>100.6</td>
<td>8.1</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>1.0</td>
<td>1.2</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>3</td>
<td>0.796</td>
<td>103.4</td>
<td>7.2</td>
<td>29.3</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
<td>0.8</td>
<td>1.3</td>
<td></td>
<td>Uncertainty</td>
</tr>
<tr>
<td>4</td>
<td>0.795</td>
<td>102.9</td>
<td>7.4</td>
<td>29.1</td>
<td>Value</td>
</tr>
<tr>
<td></td>
<td>8.2</td>
<td>0.8</td>
<td>1.2</td>
<td></td>
<td>Uncertainty</td>
</tr>
</tbody>
</table>

\[ \rightarrow \text{The “real” parent numbers} \]
### Covariance Among Parameters

#### Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>$N_0$</th>
<th>$\tau$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>67.64</td>
<td>-4.35</td>
<td>1.84</td>
</tr>
<tr>
<td>$a_3$</td>
<td>-4.35</td>
<td>0.58</td>
<td>-0.59</td>
</tr>
<tr>
<td>$B$</td>
<td>1.84</td>
<td>-0.59</td>
<td>1.55</td>
</tr>
</tbody>
</table>

\[
\alpha^{-1} = \varepsilon = \begin{pmatrix}
\sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \cdots \\
\sigma_{21}^2 & \sigma_{22}^2 & \cdots \\
\sigma_{31}^2 & \cdots & \sigma_{33}^2 \\
\vdots & \ddots & \vdots \\
\end{pmatrix}
\]

\[
\sigma_{N_0} = \sqrt{67.64} = 8.2, \text{ etc}...
\]

The off-diagonal terms specify the covariances or "correlations" among fit parameters.
Covariance Among Parameters

"Reduced" Covariance Matrix

\[
\begin{pmatrix}
N_0 & \tau & B \\
\text{a}_1 & \text{a}_2 & \text{a}_3 \\
1.000 & -0.112 & 1.000 \\
0.018 & -0.664 & 1.000 \\
\end{pmatrix}
\]

\[\epsilon'_{ij} = \frac{\epsilon_{ij}}{\epsilon_{ii} \epsilon_{jj}}, i \neq j\]

We see that the lifetime and background rate are strongly correlated.
Summary

- Expansion method uses quadratic approx. of the $\chi^2$ function near the minimum to zero in on the minimum.
- Marquardt method combines the best features of the gradient and expansion methods to fit a function to data quickly.
- Covariance matrix of parameter errors is one outcome of the expansion method algebra.
Backup Slides...