Derivation of weighted ordinary least squares

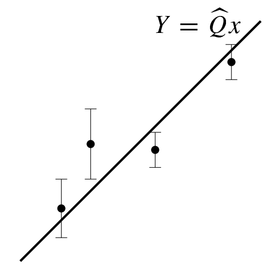
- Both ordinary least squares (OLS) and total least squares (TLS), as applied to estimate total capacity, seek to find $\widehat{Q}$ such that $y \approx \widehat{Q}x$ using N-vectors of measured data $\mathbf{x}$ and $\mathbf{y}$ $(N \geq 1)$
- ith element (x_i, y_i) corresponds to data collected from a cell over time interval i
  - x_i is estimated change in state-of-charge over that interval
  - y_i is net accumulated ampere hours passing through cell during that period
- Specifically,

$$x_i = z[k_{2,i}] - z[k_{1,i}] \qquad \text{for time interval } i$$

$$y_i = -\Delta t \sum_{k=k_{1,i}}^{k_{2,i}-1} \eta[k] i[k]$$

Major assumption of OLS

- OLS approach assumes that there is no error on x_i and models data as $\mathbf{y} = Q\mathbf{x} + \Delta\mathbf{y}$, where $\Delta\mathbf{y}$ is a vector of measurement errors
- Error bars on data points meant to illustrate uncertainties, which are proportional to σ_{y_i}
- We assume that $\Delta\mathbf{y}$ comprises zero-mean Gaussian random variables, with known variances $\sigma_{y_i}^2$ (which are not necessarily equal to each other)
- OLS attempts to find an estimate $\widehat{Q}$ of the true cell total capacity Q that minimizes sum of squared errors Δy_i



Cost function

- We generalize approach here slightly to allow for finding a $\widehat{Q}$ that minimizes sum of *weighted* squared errors
  - Weighting takes into account uncertainty of specific measurement
- That is, we seek $\widehat{Q}$ that minimizes weighted-least-squares (WLS) cost function

$$\chi^2_{\mathrm{WLS}} = \sum_{i=1}^{N} \frac{(y_i - Y_i)^2}{\sigma_{y_i}^2} = \sum_{i=1}^{N} \frac{(y_i - \widehat{Q}x_i)^2}{\sigma_{y_i}^2}$$

where Y_i is a point on the line $Y_i = \widehat{Q}x_i$ corresponding to measured data pair (x_i, y_i), where y_i is assumed to have noise but x_i has no noise

Solution

- A number of solution approaches may be taken
- One that will serve us well is to differentiate cost function with respect to $\widehat{Q}$ and solve for $\widehat{Q}$ by setting partial derivative to zero

$$\frac{\partial \chi^2_{\mathrm{WLS}}}{\partial \widehat{Q}} = -2 \sum_{i=1}^{N} \frac{x_i(y_i - \widehat{Q}x_i)}{\sigma_{y_i^2}} = 0$$

$$\widehat{Q} \sum_{i=1}^{N} \frac{x_i^2}{\sigma_{y_i}^2} = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_{y_i}^2}$$

$$\widehat{Q} = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_{y_i}^2} \Bigg/ \sum_{i=1}^{N} \frac{x_i^2}{\sigma_{y_i}^2}$$

Summary

- Have now defined and solved weighted ordinary-least-squares problem for estimating cell total capacity
- Desired to find $\widehat{Q}$ to minimize $\chi^2_{\mathrm{WLS}} = \sum_{i=1}^{N}(y_i - Y_i)^2/\sigma_{y_i}^2$
- Solution: Each time new data pair (x_i, y_i) available, compute

$$c_1 = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_{y_i}^2} \quad \text{and} \quad c_2 = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_{y_i}^2}$$

- Then, $\widehat{Q} = c_2/c_1$