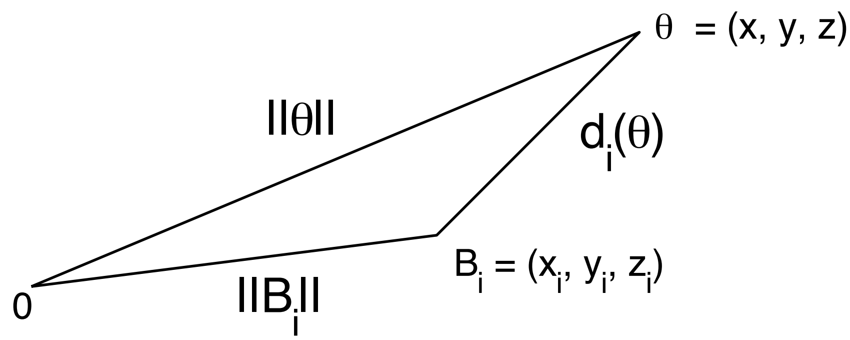
MODELING THE MEASUREMENT ERROR

Two potential sources of error:

1. Systematic error: misspecifying the coordinates of the beacons, calibration errors. Stay the same when measurements are repeated.
2. Random errors: Vary unpredictably with repeated measurements.

By modeling the error as random, statistical methods can be used to estimate the coordinates of the target point, and to calculate the likely size of the error.

We take n measurements. Perhaps one from each beacon, perhaps more.



$\boldsymbol{B_i} = (x_i, y_i, z_i)$ is the location of the beacon at which the ith measurement is taken. $\boldsymbol{B_i}$ is known.

Origin is at the center of mass of the beacons: $(\bar{x}, \bar{y}, \bar{z}) = (0, 0, 0)$.

$\boldsymbol{\theta} = (x, y, z)$ is the target point. $\boldsymbol{\theta}$ is unknown.

$d_i(\boldsymbol{\theta}) = ||\boldsymbol{B_i} - \boldsymbol{\theta}||$ is the true distance from the ith beacon to the target point.

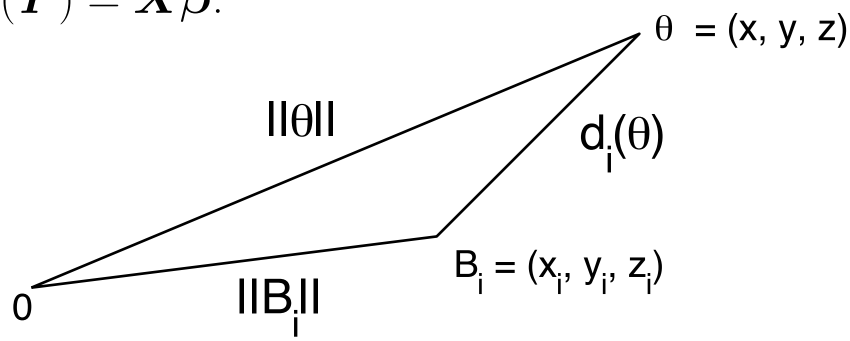
r_i is the measured distance from the ith beacon to the target point.

We assume $r_i = d_i(\boldsymbol{\theta}) + \varepsilon_i$, where $\mathrm{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, and the ε_i are independent.

Then $E(r_i) = d_i(\boldsymbol{\theta}) = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}$, which is not linear in the parameters x, y, z.

For a linear model we need:

- A vector $\boldsymbol{Y}$ of observable random variables.
- A matrix $\boldsymbol{X}$ of known constants.
- A vector β of unknown parameters containing x, y, and z, such that $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$.



Law of Cosines: $d_i(\boldsymbol{\theta})^2 = ||\boldsymbol{B_i}||^2 + ||\boldsymbol{\theta}||^2 - 2x_1x - 2y_iy - 2z_iz$

$E(r_i{}^2) = d_i(\boldsymbol{\theta})^2 + \sigma^2.$ Let $Y_i = ||\boldsymbol{B_i}||^2 - r_i{}^2.$

$\mathrm{E}(Y_i) = 2x_1x - 2y_iy - 2z_iz - ||\boldsymbol{\theta}||^2 - \sigma^2.$

Define

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2x_1 & 2y_1 & 2z_1 \\ 1 & 2x_2 & 2y_2 & 2z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2x_n & 2y_n & 2z_n \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} -||\boldsymbol{\theta}||^2 - \sigma^2 \\ x \\ y \\ z \end{bmatrix}$$

Let $\boldsymbol{Y}$ denote the vector whose ith component is Y_i. Then

$$\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{X\beta}.$$

The first component of $\boldsymbol{\beta}$ is a non-linear function of the other three, so it appears that this is not a linear regression. We ignore the functional dependence, and treat the first component as a free parameter.

The Y_i are independent, and

$\mathrm{Var}(Y_i) = \mathrm{Var}(r_i^2) = \mathrm{Var}[(d_i(\boldsymbol{\theta})+\varepsilon_i)^2] = 4d_i(\boldsymbol{\theta})^2\sigma^2+2d_i(\boldsymbol{\theta})\mu_3+\mu_4-\sigma^4$

where σ^2, μ_3, and μ_4 are the second, third, and fourth moments of the distribution of the ε_i.

$\mathrm{Var}(Y_i) \approx 4d_i(\boldsymbol{\theta})^2\sigma^2$ depends on the unknown $d_i(\boldsymbol{\theta})$.

Let $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$. $\boldsymbol{\Sigma}$ is diagonal with $\boldsymbol{\Sigma}_{ii} = \mathrm{Var}(Y_i)$.

If $\boldsymbol{\Sigma}$ were known, the best linear unbiased estimator of β would be the weighted least squares estimator:

$\hat{\boldsymbol{\beta}}_{WLS} = (\boldsymbol{X^T\Sigma^{-1}X})^{-1}\boldsymbol{X^T\Sigma^{-1}Y}.$

Since $\boldsymbol{\Sigma}$ is unknown, we consider two alternatives.

ORDINARY LEAST SQUARES

Recall

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2x_1 & 2y_1 & 2z_1 \\ 1 & 2x_2 & 2y_2 & 2z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2x_n & 2y_n & 2z_n \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} -||\boldsymbol{\theta}||^2 - \sigma^2 \\ x \\ y \\ z \end{bmatrix}$$

Let

$$\boldsymbol{X}_* = \begin{bmatrix} 2x_1 & 2y_1 & 2z_1 \\ 2x_2 & 2y_2 & 2z_2 \\ \vdots & \vdots & \vdots \\ 2x_n & 2y_n & 2z_n \end{bmatrix} \qquad \boldsymbol{\theta} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The ordinary least squares estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\boldsymbol{X}_*^{\boldsymbol{T}}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^{\boldsymbol{T}}\boldsymbol{Y}$.

$\hat{\boldsymbol{\theta}}$ is unbiased: $\mathrm{E}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}_*^{\boldsymbol{T}}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^{\boldsymbol{T}}\boldsymbol{X}_*\boldsymbol{\theta} = \boldsymbol{\theta}$.

Let s^2 be the mean squared residual (estimate of σ^2):

We estimate $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \approx s^2(\boldsymbol{X}_*^{\boldsymbol{T}}\boldsymbol{X}_*)^{-1}$.

ITERATIVELY REWEIGHTED LEAST SQUARES

A procedure in which one attempts iteratively to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

1. Choose an initial covariance matrix $\hat{\boldsymbol{\Sigma}}_{(0)}$ (e.g. $\hat{\boldsymbol{\Sigma}}_{(0)} = \boldsymbol{I}$)
2. Given $\hat{\boldsymbol{\Sigma}}_{(i)}$, compute the WLS estimate $\hat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{X}^T \hat{\boldsymbol{\Sigma}}_{(i)}^{-1} \boldsymbol{X})^{-1} X \hat{\boldsymbol{\Sigma}}_{(i)}^{-1} \boldsymbol{Y}$
3. $\hat{\boldsymbol{\theta}}_{(i)}$ is the last three components of $\hat{\boldsymbol{\beta}}_{(i)}$.
4. Use $\hat{\boldsymbol{\theta}}_{(i)}$ to update $\hat{\boldsymbol{\Sigma}}_{(i)}$.
5. Iterate 2 and 3 to convergence: obtaining $\hat{\boldsymbol{\beta}}_{\mathbf{irls}}$, $\hat{\boldsymbol{\Sigma}}$.

How to do 3:

$$\boldsymbol{\Sigma}_{ii} = 4d_i(\boldsymbol{\theta})^2 \sigma^2 + 2d_i(\boldsymbol{\theta})\mu_3 + \mu_4 - \sigma^4 \approx 4d_i(\boldsymbol{\theta})^2.$$

kth diagonal element of $\hat{\boldsymbol{\Sigma}}_{(i+1)}$ is $4d_i(\hat{\boldsymbol{\theta}}_i)^2$.

Estimate $\text{Cov}(\hat{\boldsymbol{\beta}}_{\mathbf{irls}})$ with $(\boldsymbol{X}^{\boldsymbol{T}} \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{X})^{-1}$.

Estimate $\text{Cov}(\boldsymbol{\theta})$ with appropriate 3×3 submatrix.

NONLINEAR LEAST SQUARES (QUASI-LIKELIHOOD)

Let $\boldsymbol{\theta} = (x,\ y,\ z)^T$ be the target point.

Let $d_i(\boldsymbol{\theta}) = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2}$ be the true distance to the ith beacon.

Let r_i be the measured distance. $r_i = d_i(\boldsymbol{\theta}) + \varepsilon_i$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$.

Estimate $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$, the minimizer of $F(\boldsymbol{\theta}) = \Sigma_{i=1}^{n}(r_i - d_i(\boldsymbol{\theta}))^2$.

It can be shown that $\hat{\boldsymbol{\theta}}$ is consistent, i.e. $\lim_{n\to\infty} \mathrm{P}(||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|| > \epsilon) = 0$

This procedure yields the maximum likelihood estimator when the errors are normally distributed.

Define $J(\boldsymbol{\theta})$ to be the Jacobian of $d(\boldsymbol{\theta})$: the $n \times 3$ matrix whose ith row is $(\partial d_i(\boldsymbol{\theta})/\partial x \quad \partial d_i(\boldsymbol{\theta})/\partial y \quad \partial d_i(\boldsymbol{\theta})/\partial z)$.

$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \approx \sigma^2(\boldsymbol{J(\theta)^T J(\theta)})^{-1}$

Let $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{\theta})$

$$\mathbf{0} = \boldsymbol{\nabla F}(\hat{\boldsymbol{\theta}}) = \boldsymbol{J}^{\boldsymbol{T}}(\hat{\boldsymbol{\theta}})[\boldsymbol{r} - \boldsymbol{d}(\hat{\boldsymbol{\theta}})]$$

$$\approx \boldsymbol{J}^{\boldsymbol{T}}[\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})] - \boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

$$\boldsymbol{J}^{\boldsymbol{T}}[\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})][\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})]^{\boldsymbol{T}}\boldsymbol{J} \approx \boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\boldsymbol{T}}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J}$$

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\boldsymbol{T}} \approx (\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J})^{-1}\boldsymbol{J}^{\boldsymbol{T}}[\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})][\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})]^{\boldsymbol{T}}\boldsymbol{J}(\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J})^{-1}$$

Now $\mathbf{Cov}(\hat{\boldsymbol{\theta}}) = \boldsymbol{E}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\boldsymbol{T}})$.

and $\boldsymbol{E}([\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})][\boldsymbol{r} - \boldsymbol{d}(\boldsymbol{\theta})]^{\boldsymbol{T}}) = \boldsymbol{\sigma}^2\boldsymbol{I}$.

Taking expectations,

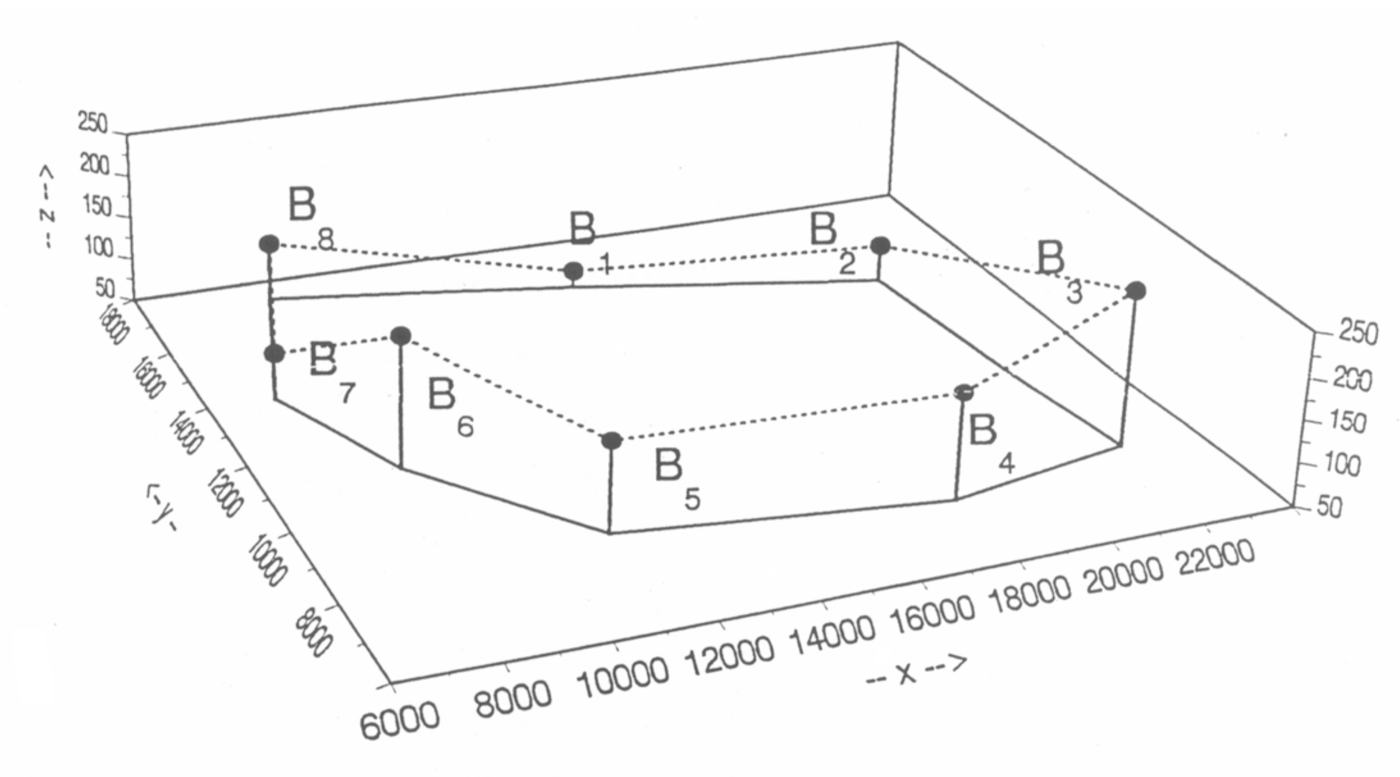
$$\mathbf{Cov}(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{\sigma}^2(\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J})^{-1}$$

Simulation:

Eight beacons were used.

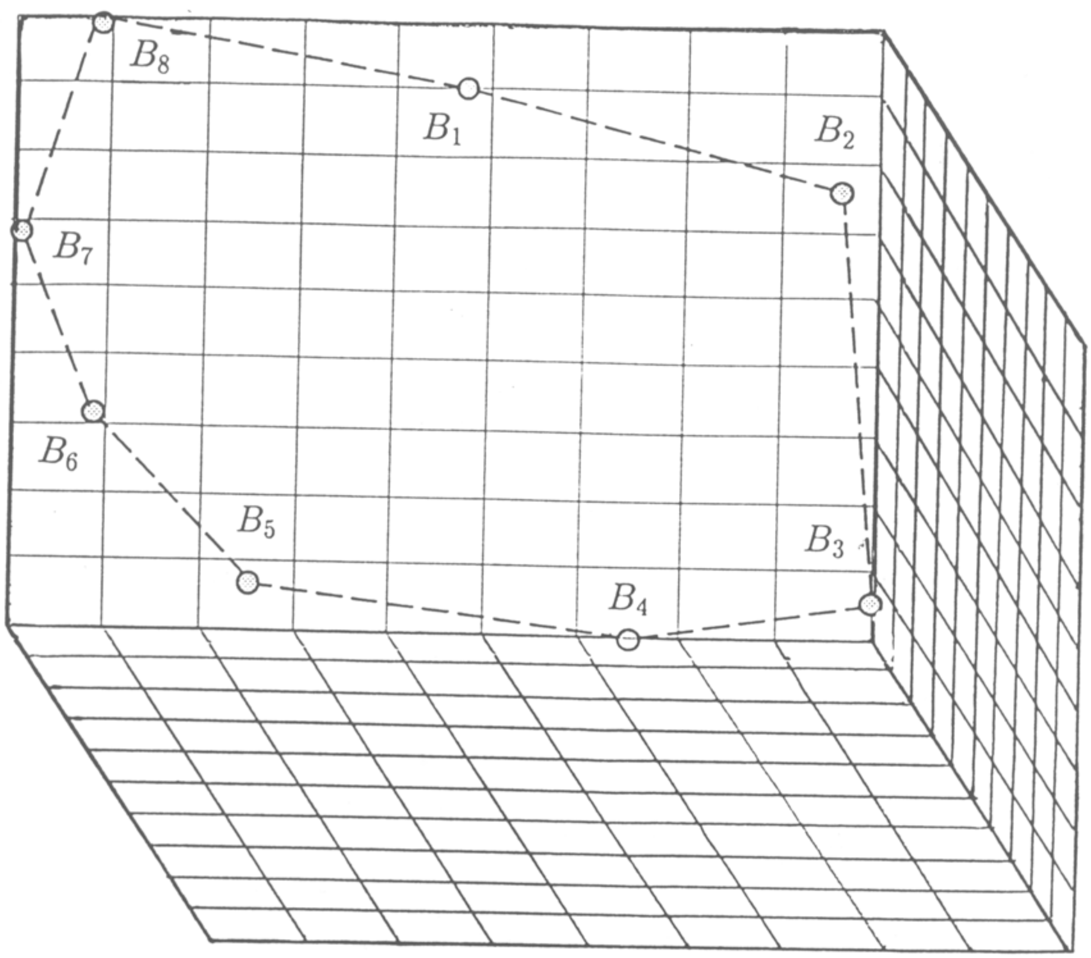
Beacon Coordinates

X	Y	Z
920	3977.5	−77.125
7360	2577.5	−53.125
8090	−3892.5	83.875
3910	−4512.5	27.875
−2710	−3742.5	4.875
−5420	−1082.5	55.875
−6740	1657.5	−42.125
−5410	5017.5	−0.125



One thousand target points, forming a rectangular lattice, were considered.

Top of box is 5 feet below lowest beacon. Dimensions are
$14830 \times 9530 \times 600$



For each target point, one thousand data sets were generated.

Each data set consisted of one measurement per beacon.

Each measurement was obtained by adding to the true distance a random error distributed uniformly on $(-0.5,\ 0.5)$

For each of the 1000 target points, we calculated the following:

- Monte Carlo estimate of Mean Square Error: $\sqrt{\dfrac{\Sigma_{i=1}^{1000} ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2}{1000}}$

Take this to be the true standard error.

- RMS Nominal MSE: $\sqrt{\dfrac{\Sigma_{i=1}^{1000} \mathrm{trace}(\mathrm{Cov}(\hat{\boldsymbol{\theta}}))}{1000}}$

This is the average estimated standard error.

- Coverage Probability for the Nominal 95% Confidence Ellipsoid:

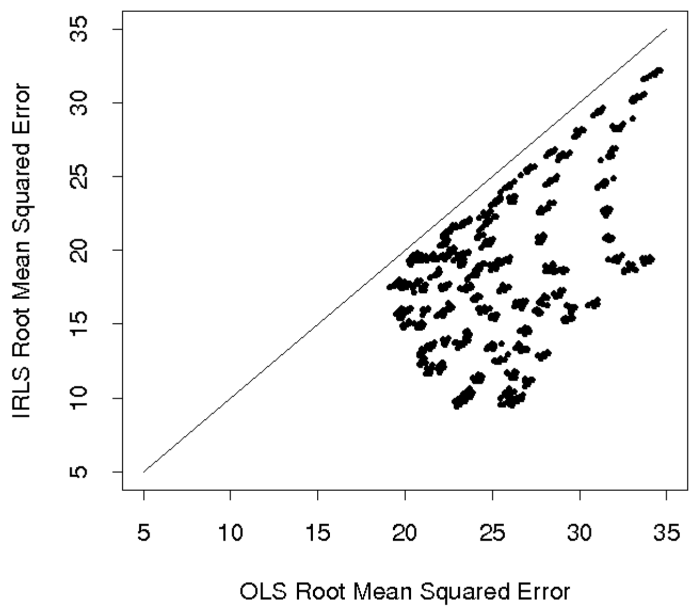
The proportion of the 1000 data sets for which

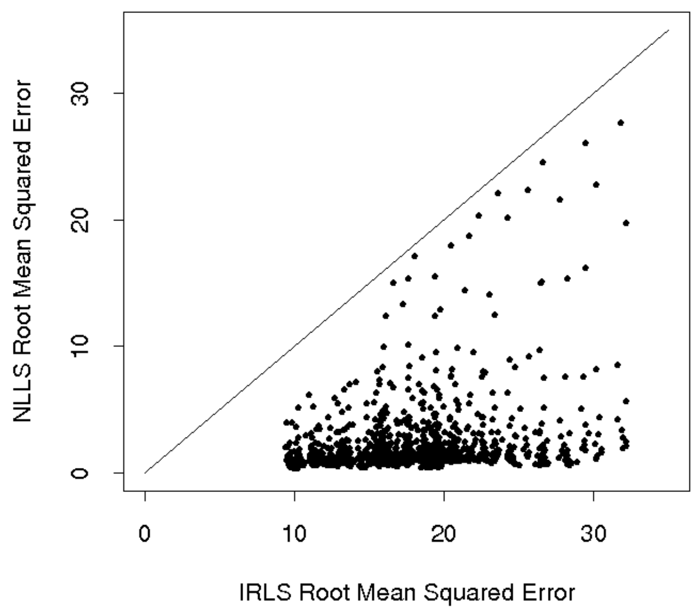
$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\mathrm{Cov}(\hat{\boldsymbol{\theta}}))^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) < \mathrm{F}_{3,\ \mathrm{n}-4,\ .95}$$

Table 2: Comparison of methods of estimation.

Estimator	RMSE	Nominal RMSE	Coverage Probability
OLS	25.34	24.12	0.9409
IRLS	18.68	172.92	0.9613
NLLS	3.96	3.78	0.9434

Units are feet.





Comparison of Nonlinear and Optimal Linear Estimators

The asymptotic covariance of the nonlinear estimator is $(\boldsymbol{J(\theta)^T J(\theta)})^{-1} =$

$$\begin{pmatrix} \sum_{i=1}^{n} \frac{(x_i - x)^2}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(x_i - x)(y_i - y)}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(x_i - x)(z_i - z)}{d_i(\theta)^2} \\ \sum_{i=1}^{n} \frac{(x_i - x)(y_i - y)}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(y_i - y)^2}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(y_i - y)(z_i - z)}{d_i(\theta)^2} \\ \sum_{i=1}^{n} \frac{(x_i - x)(z_i - z)}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(y_i - y)(z_i - z)}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{(z_i - z)^2}{d_i(\theta)^2} \end{pmatrix}^{-1}$$

Using the approximation $\Sigma_{ii} = 4d_i(\boldsymbol{\theta})^2\sigma^2$, the theoretically optimal linear estimator has covariance $(\boldsymbol{X^T \Sigma^{-1} X})^{-1} =$

$$\begin{pmatrix} \sum_{i=1}^{n} \frac{x_i^2}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{x_i y_i}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{x_i z_i}{d_i(\theta)^2} \\ \sum_{i=1}^{n} \frac{x_i y_i}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{y_i^2}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{y_i z_i}{d_i(\theta)^2} \\ \sum_{i=1}^{n} \frac{x_i z_i}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{y_i z_i}{d_i(\theta)^2} & \sum_{i=1}^{n} \frac{z_i^2}{d_i(\theta)^2} \end{pmatrix}^{-1}$$

The difference between the covariance matrices will be positive definite unless the point is much closer to the beacons in one dimension than in the others.

There were four points out of 1000 for which the theoretically optimal linear estimator outperformed the nonlinear estimator.

Issues for Further Study:

- Behavior of estimation procedures when distribution of measurement errors is skewed.

- Behavior of estimation procedures when measurement errors are proportional to the true distance.

- Bootstrap method of estimating IRLS error.

Conclusions:

When the error distribution is symmetric:

- Ordinary least squares is reliable even with small samples, in that conventional measures of error are fairly precise.

- Iteratively reweighted least squares is more precise than ordinary least squares, but conventional measures of error can greatly underestimate the precision. A bootstrap estimate of error might be much better.

- Non-linear least squares has the best performance asymptotically. It appears to hold up well for small samples.