

Chapter 4 PREVIOUS SOLUTIONS

The following chapters provide an exposé on the methods that have gone before this. The theory in this thesis is presented in two phases and each phase can be analyzed individually. The two phases are the determination of the rotation points of each joint and the generation of the whole skeleton. The first phase boils down to a single mathematical problem: that of finding the center of a sphere from noisy surface data. The second phase, drawing the skeleton, is described in Chapter 4.1.5.3.

4.1 Spherical Curve-fitting Approaches

There are three techniques in use today: iterative; least squares; and algebraic best fits. They each have their advantages and disadvantages. Iterative techniques are good for accuracy, least squares are faster than iterative but slower than algebraic, and algebraic techniques are good for speed.

4.1.1 Monte-Carlo Experiment

In order to compare the three, a Monte-Carlo experiment was run using 1000 trials, each of which had anywhere from 4 to 1,000,000 samples. The runs took over a week of computational effort to collect. Each trial had a fixed standard deviation of the samples from the sphere with values ranging from 10^{-13} to 10^{14} . Each trial was further varied by a limiting angle from some random point on a sphere. The limiting angle varied from 0 to 180° , where 180° means full sphere coverage. It is important to have a limiting angle in this experiment because all sphere-fit algorithms are error-prone when this angle gets smaller. Each trial also had a radius r_0 and center c_0 randomly chosen. The

radius was varied from 0.06 to 38.5 and the center moved as much as 3.6. The individual samples were multivariate normal random variates thus

$$(11) \quad x_i \sim N_3(\mu_i, \sigma^2)$$

where the covariance is a diagonal matrix with each diagonal equal to the square of the standard deviation. The expected value of each sample was a point that lies on a sphere thus

$$(12) \quad |\mu_i - c_0| = r_0.$$

Each sample was confined to cover only a partial part of the sphere by the following relationship (cf. Figure 6).

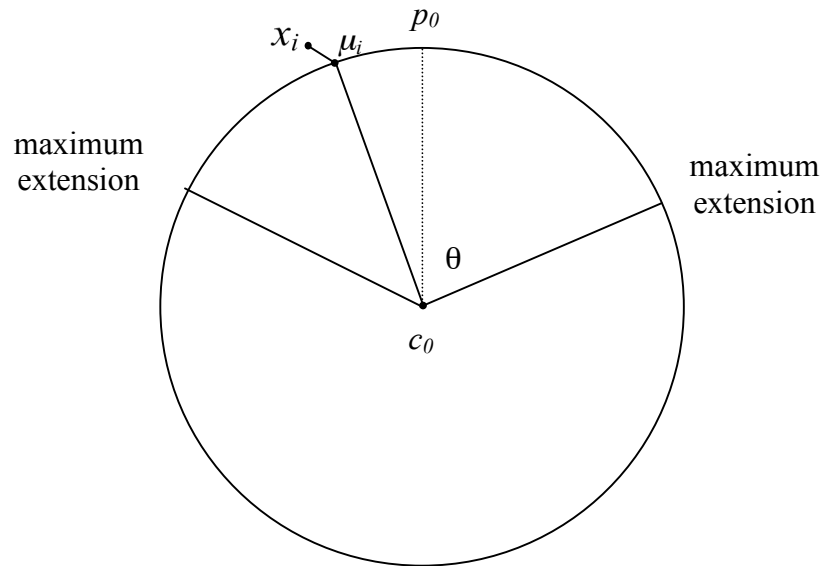


Figure 6 Constrained Measurements on Circle

The formula for checking if a point is acceptably within the limits becomes

$$(13) \quad (x_i - c_0)^T (p_0 - c_0) > |x_i - c_0| r_0 \cos(\theta).$$

Said in another way, the sample must be confined to be within an angle from a fixed point on the sphere. This experiment allows for the in-depth analysis of the error in the answer from four different estimators. The four techniques are Maximum-Likelihood Estimator (MLE), Linear Least-Squares (LLS), Generalized Delogne-Kása Estimator (GDKE), and the new Unbiased Generalized Delogne-Kása estimator (UGDK) explained in the Chapter 5.1. The first three are established formulae and has been used for two hundred years (MLE [100]) to as young as three years (GDKE [119]).

The following graph (Figure 7) shows the errors in the estimators compared to the standard deviation of the samples. The graph shows that the relative error versus relative standard deviation is a line for each of the four methods. The error is thus proportional to the standard deviation.

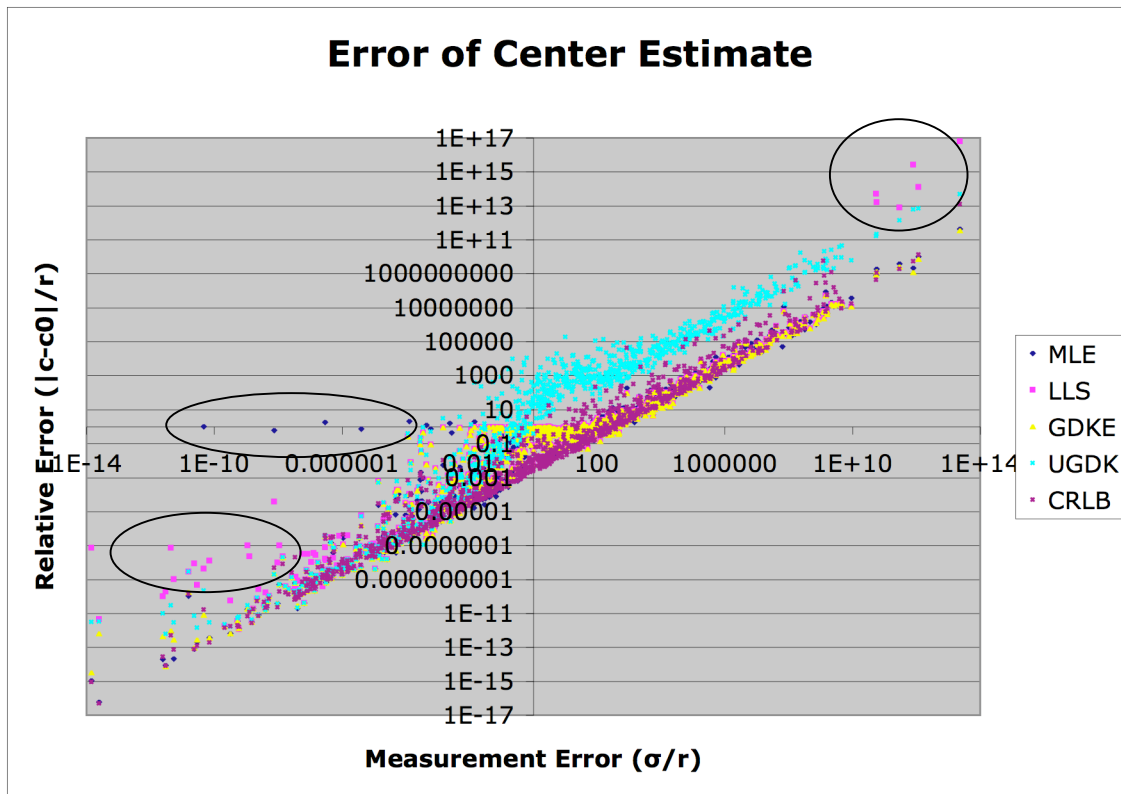


Figure 7 Relative Error Comparison

What is also clear from the graph is the comparison. The outliers on the graph have been circled. The obvious differences between the estimators show up in the graph by deviations from the straight line when error equals standard deviation. From the graph, it appears that there is a common limitation to the error in the estimator. This common limitation is named the Cramér-Rao Lower Bound to the covariance of an estimator. Section 4.1.2 discusses the limit to all estimators for this particular problem. The MLE has outliers when the error equals the radius. This is due to multiple solutions when the error equals the radius. The LLS has outliers when the standard deviation is below about 10^{-6} times the radius and greater than 10^{10} times the radius. These are due to numerical instability during the extremes of using finite representation of decimal numbers. The GDKE seems to be on par with the MLE except for the MLE outliers. The

UGDK has consistently larger error when the standard deviation is greater than about 0.1 times the radius. Each one of these shortcomings will be further discussed in sections 4.1.3, 4.1.4, 4.1.5, and 5.1.

4.1.2 Cramér-Rao Lower Bound

The Cramér-Rao Lower Bound (CRLB) is the proven lower bound for any estimator's covariance. It is equal to the inverse of the Fisher Information. It is an important measure when dealing with any estimator because it is the best error that an estimator can achieve. All estimators will have at best an error of the CRLB. The CRLB for the center and radius estimators are given for this particular problem [55] as

$$(14) \quad CRLB^{-1} = \sum_{i=1}^N \begin{pmatrix} (\mu_i - c_0)(\mu_i - c_0)^T & (\mu_i - c_0)r_0 \\ (\mu_i - c_0)^T r_0 & r_0^2 \end{pmatrix} \frac{1}{Tr((\mu_i - c_0)(\mu_i - c_0)^T \Sigma)}$$

where

μ_i is the expected value of the i^{th} measurement on the surface

c_0 is the true center

r_0 is the true radius

Σ is the measurement covariance.

This is a positive-definite matrix of size $(D+1) \times (D+1)$ where D is the number of dimensions of the hypersphere. The *CRLB* matrix is the smallest value of the covariance of an estimator for a particular measurement system. This can be reduced using the isotropic covariance assumption ($\Sigma = I\sigma^2$). The result of this substitution is

$$(15) \quad CRLB = \sigma^2 \left(\begin{array}{cc} \sum_{i=1}^N \frac{(\mu_i - c_0)(\mu_i - c_0)^T}{r_0^2} & \sum_{i=1}^N \frac{(\mu_i - c_0)r_0}{r_0^2} \\ \sum_{i=1}^N \frac{(\mu_i - c_0)^T r_0}{r_0^2} & \sum_{i=1}^N \frac{r_0^2}{r_0^2} \end{array} \right)^{-1}$$

because of the identity

$$(16) \quad (\mu_i - c_0)^T (\mu_i - c_0) = r_0^2$$

This summation can be expanded and simplified. The equivalent sum is

$$(17) \quad CRLB = \sigma^2 \left(\begin{array}{cc} \frac{NB}{r_0^2} & N \frac{(\bar{\mu} - c_0)}{r_0} \\ N \frac{(\bar{\mu} - c_0)^T}{r_0} & N \end{array} \right)^{-1}$$

where

$$(18) \quad \mathbf{B} = \frac{1}{N} \sum_{i=1}^N (\mu_i - c_0)(\mu_i - c_0)^T$$

or equivalently

$$(19) \quad CRLB = \frac{\sigma^2}{N} \left(\begin{array}{cc} \frac{\mathbf{B}}{r_0^2} & \frac{(\bar{\mu} - c_0)}{r_0} \\ \frac{(\bar{\mu} - c_0)^T}{r_0} & 1 \end{array} \right)^{-1}$$

This partitioned matrix is inverted to

$$(20) \quad CRLB = \frac{1}{N} \sigma^2 \left(\begin{array}{cc} r_0^2 \mathbf{B}^{-1} (\mathbf{I} + \alpha (\bar{\mu} - c_0)(\bar{\mu} - c_0)^T \mathbf{B}^{-1}) & -\alpha r_0 \mathbf{B}^{-1} (\bar{\mu} - c_0) \\ -\alpha r_0 (\bar{\mu} - c_0)^T \mathbf{B}^{-1} & \alpha \end{array} \right)$$

where

$$(21) \quad \alpha = \left(1 - \text{Tr}\left((\bar{\mu} - c_0)(\bar{\mu} - c_0)^T \mathbf{B}^{-1}\right)\right)^{-1}$$

The determinant of this matrix is

$$(22) \quad |\text{CRLB}| = \left(\frac{1}{N} \sigma^2 r_0^2\right)^{D+1} \frac{\alpha}{|\mathbf{B}|}$$

For large amounts of sampling, the Cramér-Rao Lower Bound has an even more refined formula for the constrained samples on the surface of the hypersphere. Like in Figure 6, the constrained sampling is confined to an angle θ from a particular point on the surface. The two-dimensional case (i.e. the circle) can be setup so that the calculus is made simpler. First, use a coordinate system that has its center where the center of the circle is. The y-axis is projected towards the particular point that lies on the surface and is the center of the confinement. A point on the circumference is then defined as

$$(23) \quad v_2 = \begin{pmatrix} r_0 \sin a \\ r_0 \cos a \end{pmatrix}$$

where a is the angle from the y-axis towards the x-axis. The center of the circle is then a zero vector in this coordinate frame.

$$(24) \quad c_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

For large amounts of sampling of the two-dimensional case, Equation (14) becomes an integral that averages the integrand over the region on the circumference of the circle thus

$$(25) \quad \lim_{N \rightarrow \infty} CRLB \equiv CRLB_2 = \frac{1}{N} \phi_2 r_0^2 \left(\begin{array}{cc} \int_{-\theta}^{\theta} \frac{(v_2 - c_0)(v_2 - c_0)^T}{(v_2 - c_0)^T \Sigma (v_2 - c_0)} da & \int_{-\theta}^{\theta} \frac{(v_2 - c_0)r_0}{(v_2 - c_0)^T \Sigma (v_2 - c_0)} da \\ \int_{-\theta}^{\theta} \frac{(v_2 - c_0)^T r_0}{(v_2 - c_0)^T \Sigma (v_2 - c_0)} da & \int_{-\theta}^{\theta} \frac{r_0^2}{(v_2 - c_0)^T \Sigma (v_2 - c_0)} da \end{array} \right)^{-1}$$

where

$$(26) \quad \phi_2 = \int_{-\theta}^{\theta} da = 2\theta$$

The $CRLB_2$ simplifies greatly with an isotropic covariance ($\Sigma = I\sigma^2$) as well as a zero center:

$$(27) \quad CRLB_2 = \frac{1}{N} \sigma^2 \phi_2 r_0^2 \left(\begin{array}{cc} \int_{-\theta}^{\theta} v_2 v_2^T da & r_0 \int_{-\theta}^{\theta} v_2 da \\ r_0 \int_{-\theta}^{\theta} v_2 da & r_0^2 \int_{-\theta}^{\theta} da \end{array} \right)^{-1}$$

Taking the integral produces the answer in this coordinate system.

$$(28) \quad CRLB_2 = \frac{1}{N} \sigma^2 2\theta \left(\begin{array}{ccc} \theta - \cos\theta \sin\theta & 0 & 0 \\ 0 & \theta + \cos\theta \sin\theta & 2\sin\theta \\ 0 & 2\sin\theta & 2\theta \end{array} \right)^{-1}$$

The inverse produces

$$(29) \quad CRLB_2 = \frac{1}{N} \sigma^2 \frac{2\theta}{\det_2} \left(\begin{array}{ccc} \frac{\det_2}{\theta - \cos\theta \sin\theta} & 0 & 0 \\ 0 & 2\theta & -2\sin\theta \\ 0 & -2\sin\theta & \theta + \cos\theta \sin\theta \end{array} \right)$$

where

$$(30) \quad \det_2 = 2\theta(\theta + \cos\theta \sin\theta) - 4\sin^2\theta$$

This has three distinct eigenvalues. The one for the x-direction (perpendicular to the line of symmetry) is

$$(31) \quad \lambda_{m2} = \frac{1}{N} \sigma^2 \frac{2\theta}{\theta - \cos \theta \sin \theta}$$

corresponding to the direction

$$(32) \quad v_{x2} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

The largest eigenvalue corresponds, for the most part, to the y-direction and is

$$(33) \quad \lambda_{M2} = \frac{1}{N} \sigma^2 \left(\frac{4\theta}{3\theta + \cos \theta \sin \theta - \sqrt{(\theta - \cos \theta \sin \theta)^2 + 16 \sin^2 \theta}} \right)$$

corresponding to the direction

$$(34) \quad v_{y2} = \begin{pmatrix} 0 \\ 1 \\ -\alpha \end{pmatrix}$$

where

$$(35) \quad \alpha = \frac{-\theta + \cos \theta \sin \theta + \sqrt{(\theta - \cos \theta \sin \theta)^2 + 16 \sin^2 \theta}}{4 \sin \theta}$$

The variable α is a function going from $\alpha=1$ at $\theta=0$ to $\alpha=0$ at $\theta=\pi$. It is graphed in the next figure:

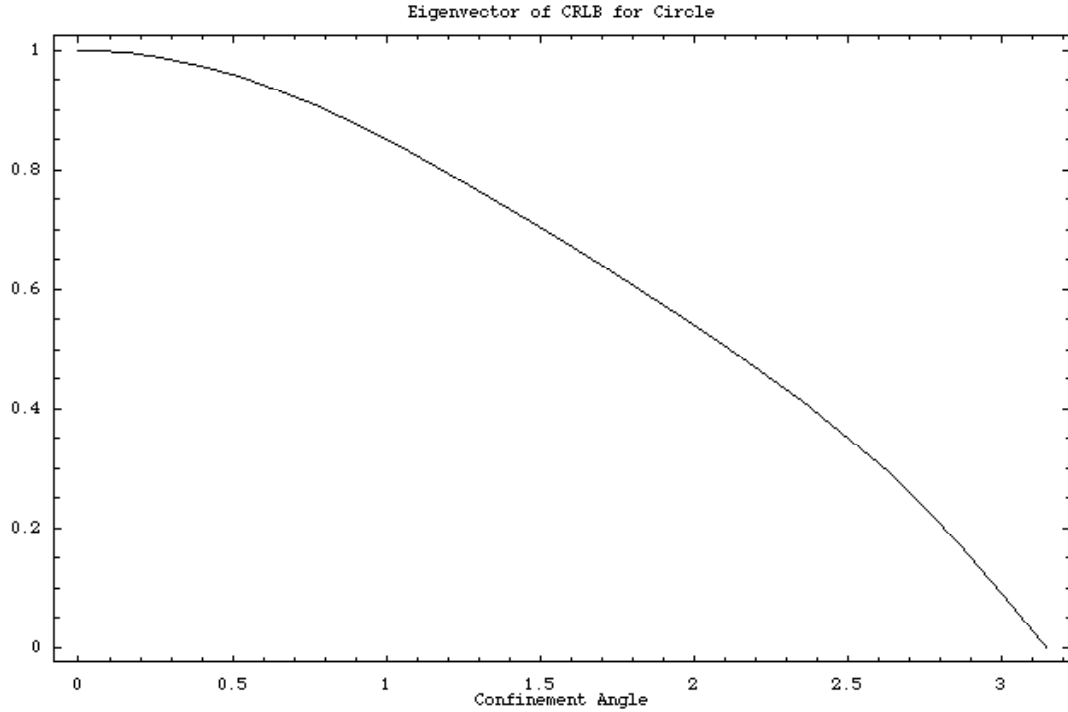


Figure 8 Eigenvectors of CRLB for Circle

The smallest eigenvalue corresponds, for the most part, to the r-direction (radius error) and is

$$(36) \quad \lambda_{r2} = \frac{1}{N} \sigma^2 \left(\frac{4\theta}{3\theta + \cos\theta \sin\theta + \sqrt{(\theta - \cos\theta \sin\theta)^2 + 16\sin^2\theta}} \right)$$

corresponding to the direction

$$(37) \quad v_{r2} = \begin{pmatrix} 0 \\ \alpha \\ 1 \end{pmatrix}$$

The eigenvalues follow the relationship

$$(38) \quad \lambda_{r2} \leq \lambda_{m2} \leq \lambda_{M2}$$

for all values of the confinement angle. They are graphed in the following figure.

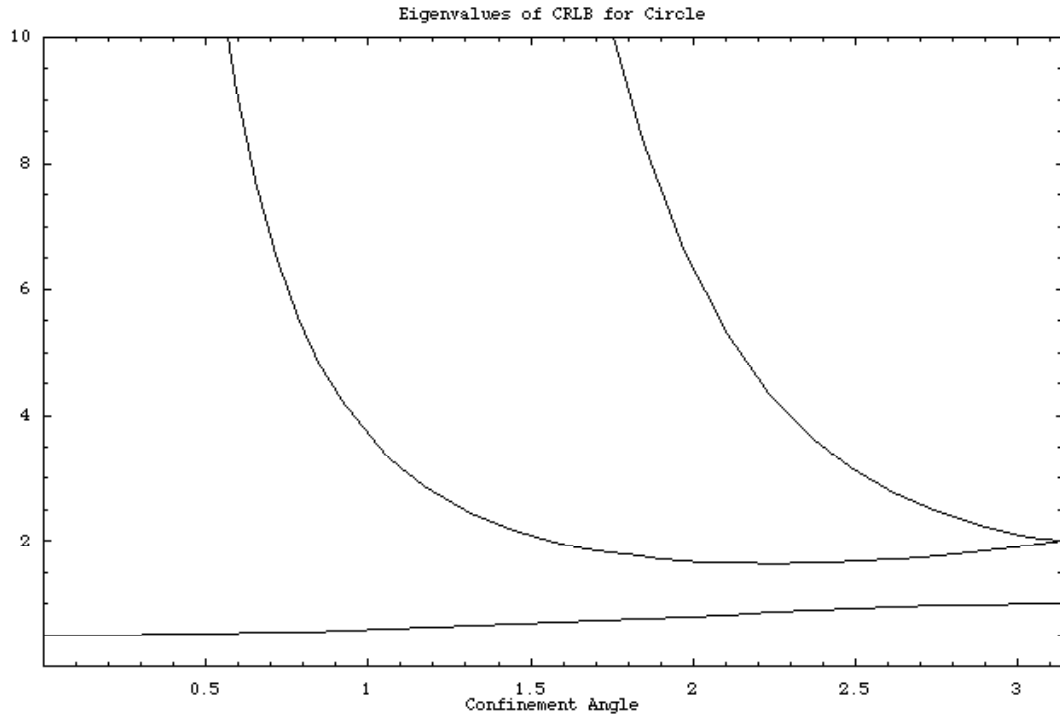


Figure 9 Eigenvalues of CRLB for Circle

Figure 9 shows that the variance for the center estimation increases to infinite as the confinement angle approaches zero. These eigenvalues determine the size of the error ellipse, which bulges towards the points on the surface of the circle.

The three-dimensional equivalent (i.e. sphere) is a bit different. The idea is still the same, which is to confine the measurements to be within an angle from a specific point on the surface. The coordinate system is set up so that the z-axis is pointing towards the confinement point on the surface. A point on the surface is then defined by two angles a and b thus

$$(39) \quad v_3 = \begin{pmatrix} r_0 \sin a \cos b \\ r_0 \sin a \sin b \\ r_0 \cos a \end{pmatrix}$$

The center of the sphere is at the center of the coordinate system thus

$$(40) \quad c_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The CRLB in three dimensions has the limit for large sampling as

$$(41) \quad \lim_{N \rightarrow \infty} CRLB \equiv CRLB_3 = \frac{1}{N} \sigma^2 r_0^2 \phi_3 \begin{pmatrix} \int_0^\theta \int_0^{2\pi} v_3 v_3^T \sin a \, db \, da & r_0 \int_0^\theta \int_0^{2\pi} v_3 \sin a \, db \, da \\ 0 & 0 \\ r_0 \int_0^\theta \int_0^{2\pi} v_3^T \sin a \, db \, da & r_0^2 \int_0^\theta \int_0^{2\pi} \sin a \, db \, da \\ 0 & 0 \end{pmatrix}^{-1}$$

where

$$(42) \quad \phi_3 = \int_0^\theta \int_0^{2\pi} \sin a \, db \, da = 4\pi \sin^2\left(\frac{\theta}{2}\right)$$

Taking the double integral produces

$$(43) \quad CRLB_3 = \frac{1}{N} \sigma^2 \begin{pmatrix} \frac{(2 + \cos\theta) \sin^2\left(\frac{\theta}{2}\right)}{3} & 0 & 0 & 0 \\ 0 & \frac{(2 + \cos\theta) \sin^2\left(\frac{\theta}{2}\right)}{3} & 0 & 0 \\ 0 & 0 & \frac{3 \cos^4\left(\frac{\theta}{2}\right) + \sin^4\left(\frac{\theta}{2}\right)}{3} & \cos^2\left(\frac{\theta}{2}\right) \\ 0 & 0 & \cos^2\left(\frac{\theta}{2}\right) & 1 \end{pmatrix}^{-1}$$

$$(44) \quad CRLB_3 = \frac{1}{N} \sigma^2 \frac{3}{\sin^4\left(\frac{\theta}{2}\right)} \begin{pmatrix} \frac{\sin^2\left(\frac{\theta}{2}\right)}{(2 + \cos\theta)} & 0 & 0 & 0 \\ 0 & \frac{\sin^2\left(\frac{\theta}{2}\right)}{(2 + \cos\theta)} & 0 & 0 \\ 0 & 0 & 1 & -\cos^2\left(\frac{\theta}{2}\right) \\ 0 & 0 & -\cos^2\left(\frac{\theta}{2}\right) & \frac{3 \cos^4\left(\frac{\theta}{2}\right) + \sin^4\left(\frac{\theta}{2}\right)}{3} \end{pmatrix}$$

The middle sized eigenvalue is

$$(45) \quad \lambda_{m3} = \frac{1}{N} \sigma^2 \frac{3}{(2 + \cos\theta) \sin^2\left(\frac{\theta}{2}\right)}$$

corresponding to the directions

$$(46) \quad v_{x3} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad v_{y3} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

The largest eigenvalue is

(47)

$$\lambda_{M3} = \frac{1}{N} \sigma^2 \left(\frac{24}{18 + 4 \cos \theta + 2 \cos(2\theta) - \sqrt{2} \sqrt{131 + 124 \cos \theta + 28 \cos(2\theta) + 4 \cos(3\theta) + \cos(4\theta)}} \right)$$

corresponding to the directions

$$(48) \quad v_{z3} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \alpha \end{pmatrix}$$

where

$$(49) \quad \alpha = -\frac{1}{3} (2 + \cos \theta) \tan^2\left(\frac{\theta}{2}\right) + \frac{\sqrt{2} \sqrt{131 + 124 \cos \theta + 28 \cos(2\theta) + 4 \cos(3\theta) + \cos(4\theta)}}{12(1 - \cos \theta) \cos^2\left(\frac{\theta}{2}\right)}$$

The third and final eigenvalue corresponds to the r -direction and is

$$(50) \quad \lambda_{r3} = \frac{1}{N} \sigma^2 \left(2 + \frac{3}{\tan^2\left(\frac{\theta}{2}\right) \sin^2\left(\frac{\theta}{2}\right)} - \frac{\sqrt{131 + 124 \cos \theta + 28 \cos(2\theta) + 4 \cos(3\theta) + \cos(4\theta)}}{4\sqrt{2} \sin^4\left(\frac{\theta}{2}\right)} \right)$$

corresponding to the directions

$$(51) \quad v_{r3} = \begin{pmatrix} 0 \\ 0 \\ -\alpha \\ 1 \end{pmatrix}$$

The last two eigenvectors are determined by the one function α of the confinement angle.

This function is graphed in Figure 10.

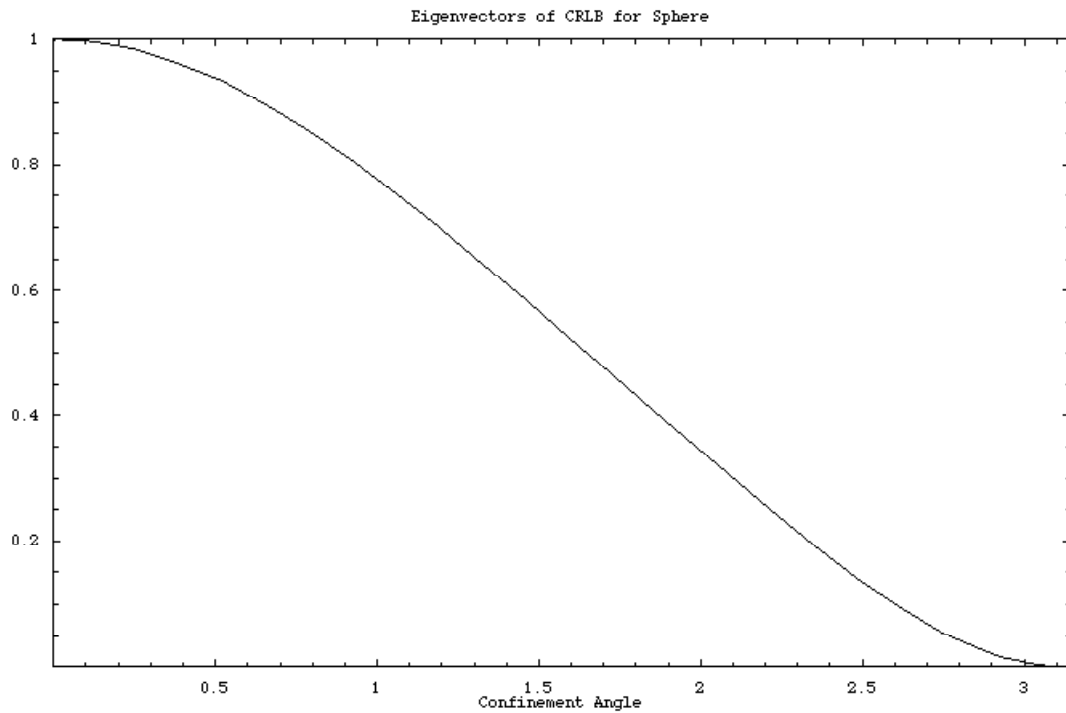


Figure 10 Eigenvectors of CRLB for Sphere

All three eigenvalues for the confined points on a sphere are graphed in Figure 11.

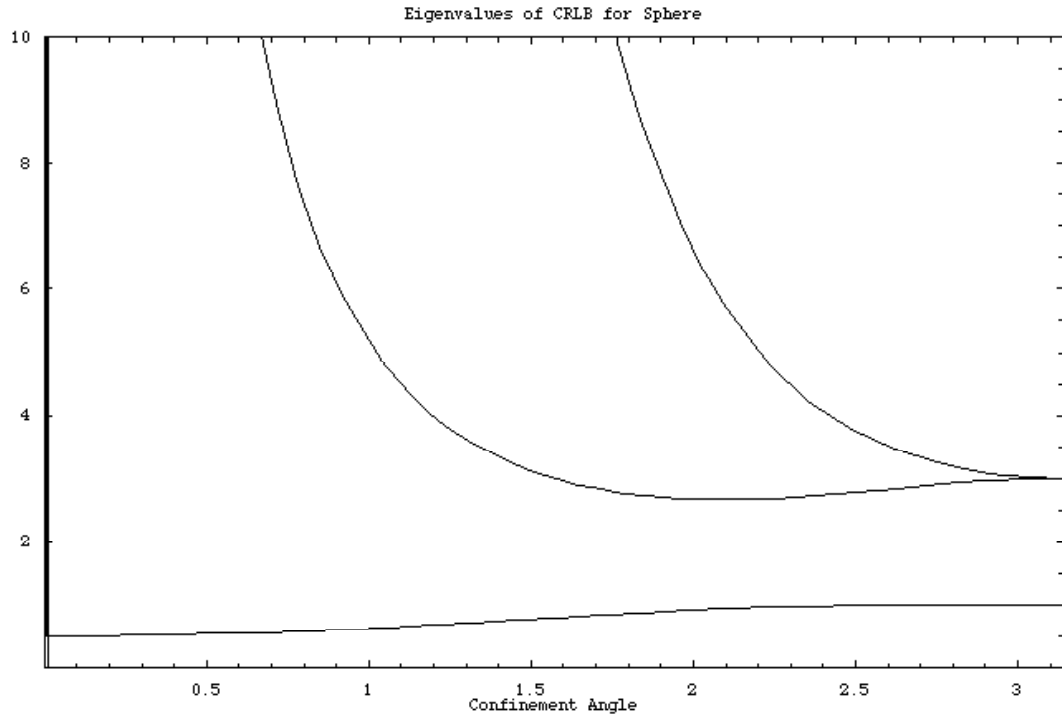


Figure 11 Eigenvalues of CRLB for Sphere

So what does all this gives me? The formulae for the eigenvalues provide the best-case error of problem at hand, i.e. all estimators aspire to these limits. There is a condition where the error in the estimate exceeds the error in the measurement. This condition occurs, for the circle, as

$$(52) \quad CRLB_2 > \sigma^2$$

The comparison here is element by element in the matrix. The worst case for the center estimator is

$$(53) \quad N < \frac{2\theta^2}{\theta(\theta + \cos\theta \sin\theta) - 2\sin^2\theta}$$

The sphere equivalent limit is

$$(54) \quad N < \frac{3}{\sin^4\left(\frac{\theta}{2}\right)}$$

These two lower limits for number of samples is a good start when the circumstances arise for picking the number of samples.

4.1.3 Non-linear Maximum-Likelihood Estimator

According to the National Institute of Standards and Technology (NIST) [100], the best way to find the center of a sphere is through non-linear minimization of the variance of the radius. This is also called the Maximum-Likelihood Estimator (MLE) for the center \hat{c} and radius \hat{r} . The variance is written as

$$(55) \quad s_{MLE}^2 = \frac{1}{N-1} \sum_{i=1}^N (r_i - \hat{r})^2$$

where N is the number of samples whose positions are x_i on or near the surface of the sphere. The sample radius r_i is expressed as the vector norm (distance) of the sample from the center of the sphere thus

$$(56) \quad r_i \equiv |x_i - \hat{c}|.$$

The two estimators, i.e. the radius \hat{r} and center \hat{c} , are unknown in the equation and can be solved by minimizing the variance. The radius is solved easily through taking the derivative of the variance and setting it equal to zero. The derivative is

$$(57) \quad \frac{\partial s_{MLE}^2}{\partial \hat{r}} = \frac{1}{N-1} \sum_{i=1}^N 2(\hat{r} - r_i) = 0$$

This summation is simplified to

$$(58) \quad \sum_{i=1}^N \hat{r} = \sum_{i=1}^N r_i$$

which ultimately simplifies to

$$(59) \quad \hat{r} = \frac{1}{N} \sum_{i=1}^N r_i$$

This estimator of the radius is still dependent on the center estimator \hat{c} because of the definition of r_i . The minimization is usually carried out by iterative methods like the Levenberg-Marquardt Method [100] and cannot be solved directly. One disadvantage of the iterative technique is the need to produce an initial guess. The following graphs comes from the Monte-Carlo experiment previously mentioned and is another way of showing the error, this time dividing by the square root of the CRLB instead of the radius. The first graph compares the error to the standard deviation.

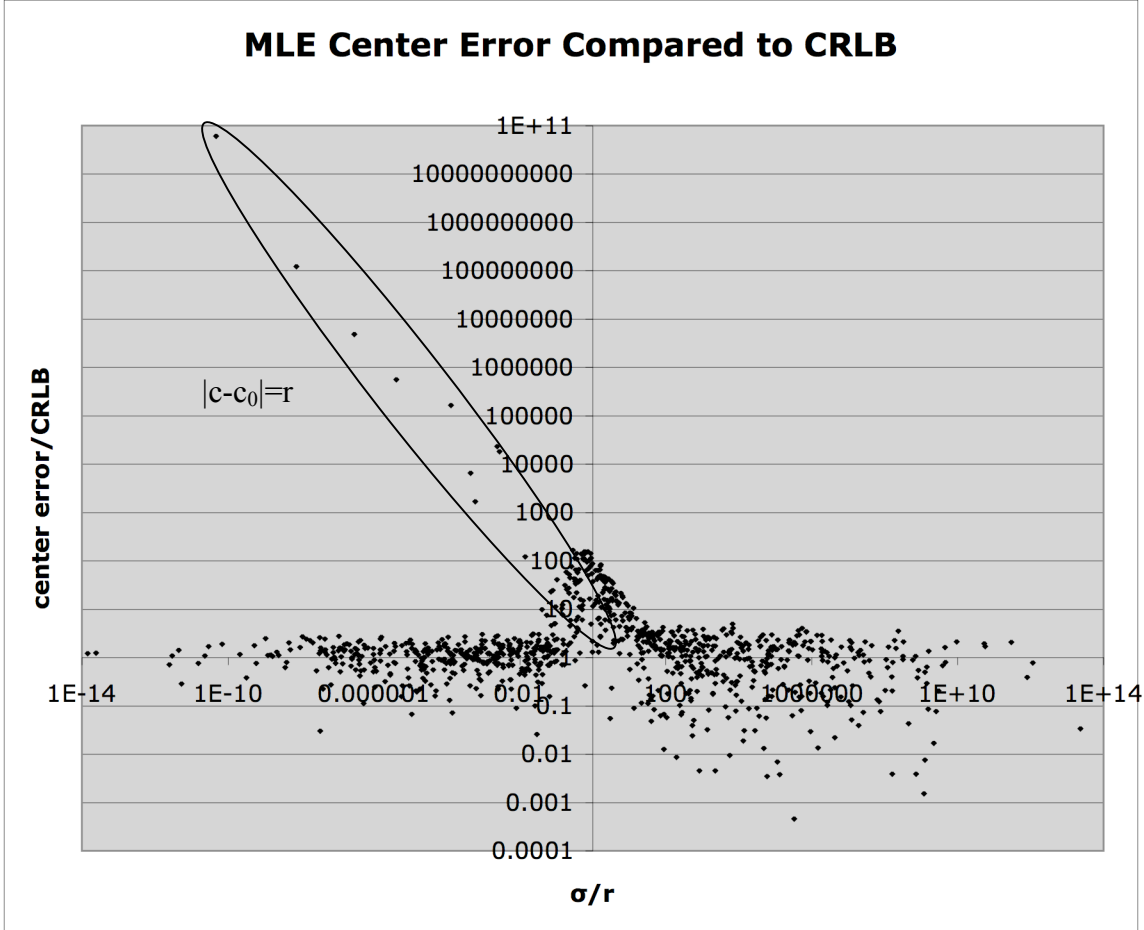


Figure 12 MLE Compared to CRLB

The outliers are clearly visible and correspond to when the error equals the radius. This is the case when the coverage angle is small and the MLE converges to the middle of the points instead of nearer to the center. The next graph explains when the outliers occur in comparison to the coverage angle. An angle of 180° means full coverage of the sphere, whereas an angle of 0° means all points confined to a single spot on the sphere.

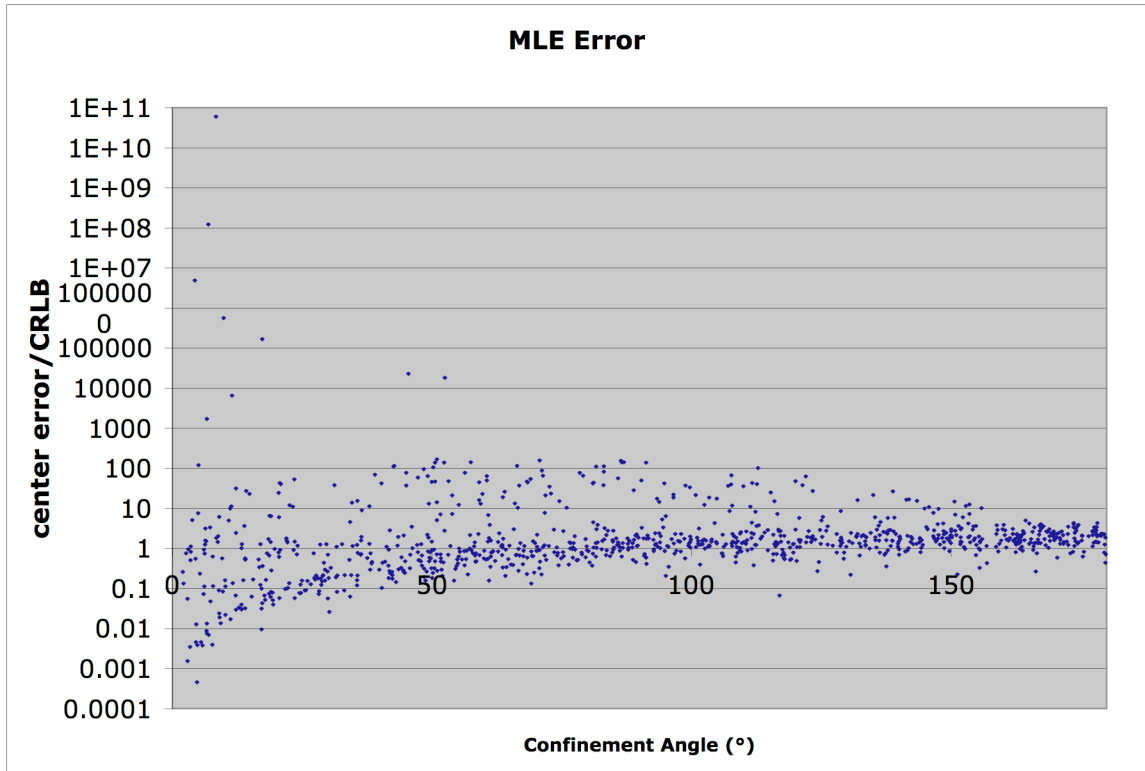


Figure 13 MLE Error Versus Sphere Coverage

This graph shows the outliers occur at low angles and indicate that the Levenberg-Marquardt technique converged on an answer that was embedded in the set of points that reside on the sphere as in the equation

$$(60) \quad |\hat{c} - c_0| \approx r_0.$$

These outliers occur when the samples during a run are confined to a small area on the sphere. For the most part, the answer error is proportional to the square root of the CRLB as evident in Figure 12 and Figure 13.

$$(61) \quad |\hat{c} - c_0|^2 \approx CRLB$$

4.1.4 Linear Least-Squares Solution

The next best thing to the very slow MLE method is through linear least-squares solution. Most authors use a similar solution to the one presented here. The method starts out the equation for a hypersphere

$$(62) \quad \hat{r}^2 = (x_i - \hat{c})^T (x_i - \hat{c}).$$

Expanding the vector product produces N linear equations for the sphere thus

$$(63) \quad x_i^T x_i = 2x_i^T \hat{c} + (\hat{r}^2 - \hat{c}^T \hat{c}).$$

This is usually an over-constrained problem since there are N equations and four (i.e. $D+1$) unknowns. The N equations can be put into a single matrix equation to solve with standard linear algebra techniques. The linear equation is

$$(64) \quad Y = X\hat{A}$$

where

$$(65) \quad Y = \begin{pmatrix} x_1^T x_1 \\ x_2^T x_2 \\ \vdots \end{pmatrix}$$

$$(66) \quad X = \begin{pmatrix} 2x_{1x} & 2x_{1y} & 2x_{1z} & 1 \\ 2x_{2x} & 2x_{2y} & 2x_{2z} & 1 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

$$(67) \quad \hat{A} = \begin{pmatrix} \hat{c}_x \\ \hat{c}_y \\ \hat{c}_z \\ \hat{r}^2 - \hat{c}^T \hat{c} \end{pmatrix}$$

The Singular Value Decomposition (SVD) method is used to solved for \hat{A} thereby retrieving \hat{c} and \hat{r} . SVD has the advantage of being able to solve equations that contain near singular matrices. It has the disadvantage of being a slow algorithm.

Figure 14 shows the error of the center estimator. Once again, the outlier case shows when the answer erroneously lies on the sphere. This example, the LLS method shows possible bad results at the extreme ends of the standard deviation spectrum.

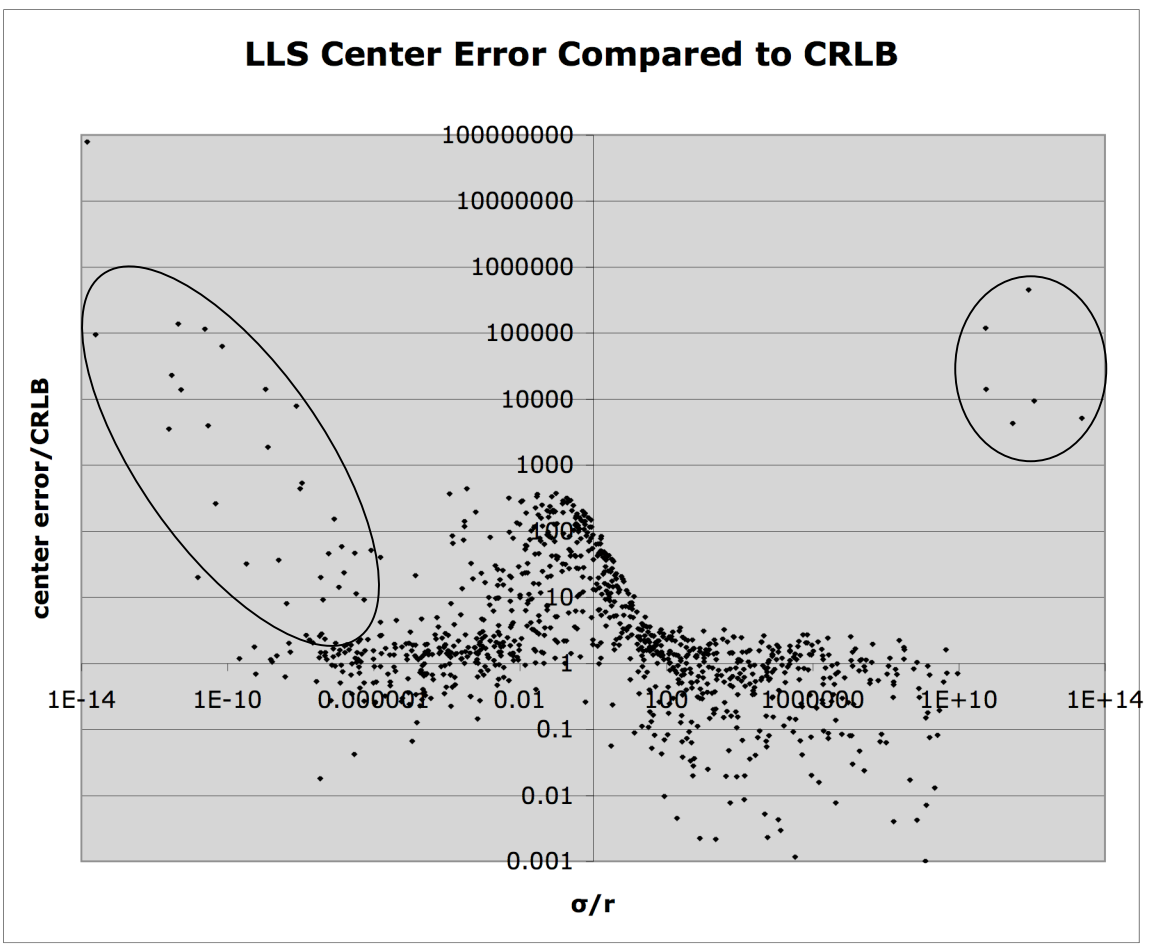


Figure 14 LLS Compared to CRLB

For the most part, once again, the answer error is proportional to the square root of the CRLB.

$$(68) \quad |\hat{c} - c_0|^2 \approx CRLB.$$

The speed of the algorithm can be measured in flop count. Flops are floating point operations and the operations included in the count are not very well defined. Many authors allow only multiplication, division, addition and subtraction of floating point numbers to be included. Others also include the square-root function. Some exclude addition and subtraction. We adopt the one with the most (i.e. * / + - $\sqrt{\quad}$). The LLS method here involves the Singular Value Decomposition. The Numerical Recipes [94] implementation is the most popular. Counting the individual + - * / $\sqrt{\quad}$ operations proves cumbersome and involves an assumption about an iteration cycle in the implementation. The setup and breakdown of Equation (64) contributes

$$(69) \quad FLOPs = N(3D - 1) + (2D + 1)$$

Each iteration for the algorithm increases the count by

$$(70) \quad FLOPs = N(6D^2 + 12D + 6) + \frac{1}{2}(53D^2 + 117D + 70)$$

If the iterations are assumed to be one pass then the total count is

$$(71) \quad FLOPs = N(14D^2 + 32D + 14) + \frac{1}{3}(13D^3 + 102D^2 + 245D + 123)$$

where D is the dimension of the sphere (e.g. 2 for circle, 3 for sphere). So for a sphere, the Linear Least-Squares method has

$$(72) \quad FLOPs = N236 + 709$$

4.1.5 Generalized Delogne-Kása Estimator

The GDKE was first introduced in 2004 by Zelniker [119] as an extension of the estimator of the circle center. The circle estimator had been published since 1972 starting with Paul Delogne [41]. István Kása [57] further expanded the 2D theory in 1976. In the process of my investigating this problem, the research for this dissertation took the same path as Zelniker and independently discovered the equation in a more efficient form in 2005. The new but algebraically equivalent equation was presented at the Winter School of Computer Graphics Conference at Plzeň, Česká Republika in 2007 [63]. The estimator has an algebraic solution and is considered fast in relation to the previously mentioned methods.

4.1.5.1 Derivation

The problem starts out similar to the MLE, where a minimum to a variance is needed. The variance takes on the form

$$(73) \quad s_{GDK}^2 = \frac{1}{N-1} \sum_{i=1}^N (r_i^2 - \hat{r}^2)^2$$

where

$$(74) \quad r_i^2 = (x_i - \hat{c})^T (x_i - \hat{c})$$

The solution for the radius estimator falls out similar to the MLE except this time it is the square root of the sum of the squares (RMS) of the sample radii.

$$(75) \quad \hat{r} = \sqrt{\frac{1}{N} \sum_{i=1}^N r_i^2}$$

If this estimator is then plugged back into the variance, it becomes evident that this estimator is the absolute minimum of the variance for any center estimator. Choose an estimator r' other than \hat{r} and the difference of the variances becomes

$$(76) \quad s^2(r') - s^2(\hat{r}) = (r'^2 - \hat{r}^2)^2 \geq 0$$

Don't forget that the center estimator \hat{c} is in Equation (56) of the sample radii r_i for the radius estimator. To solve for the center estimator, the gradient of the variance with respect to the changing center estimate must be taken and then set equal to zero

$$(77) \quad \nabla s_{GDK}^2 = 0$$

where the gradient is defined as the vector operator with respect to the center estimate

$$(78) \quad \nabla = \left(\frac{\partial}{\partial \hat{c}_x} \quad \frac{\partial}{\partial \hat{c}_y} \quad \frac{\partial}{\partial \hat{c}_z} \right)^T$$

Applying the derivative produces

$$(79) \quad \nabla s_{GDK}^2 = \frac{1}{N-1} \left(\sum_{i=1}^N \nabla r_i^4 - 2N\hat{r}^2 \nabla \hat{r}^2 \right)$$

The further derivatives can be shown to be

$$(80) \quad \nabla \hat{r}^2 = 2(\hat{c} - \bar{x})$$

$$(81) \quad \nabla r_i^4 = 4r_i^2(\hat{c} - x_i)$$

These relationships and further manipulation can produce

$$(82) \quad \sum_{i=1}^N \nabla r_i^4 = 4(N-1)(2\mathbf{C}(\hat{c} - \bar{x}) - \mathbf{S}) + 4N\hat{r}^2(\hat{c} - \bar{x})$$

where the special values \mathbf{C} and \mathbf{S} are defined as

$$(83) \quad \mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$$

$$(84) \quad \mathbf{S} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T (x_i - \bar{x})$$

\mathbf{C} is the standard definition of the sample covariance matrix. \mathbf{S} is the multidimensional equivalent of the third central moment and as such can be called the sample third central vector moment. Combining all of these makes the gradient of the variance end up being

$$(85) \quad \nabla s_{GDK}^2 = 4(2\mathbf{C}(\hat{c} - \bar{x}) - \mathbf{S})$$

Setting this to zero, the solution for the center estimator becomes

$$(86) \quad \hat{c} = \bar{x} + \frac{1}{2}\mathbf{C}^{-1}\mathbf{S}$$

If this estimator is then plugged back into the variance, it becomes evident that this estimator is the absolute minimum of the variance. Choose an estimator c' other than \hat{c} and the difference of the variances becomes

$$(87) \quad s^2(c') - s^2(\hat{c}) = (c' - \hat{c})^T \mathbf{C}(c' - \hat{c}) \geq 0$$

Applying this formula produced statistically similar results to the MLE iterations in the Monte-Carlo experiment. The following graph shows that, similar to the MLE, it has error

$$(88) \quad |\hat{c} - c_0| = O(\sigma)$$

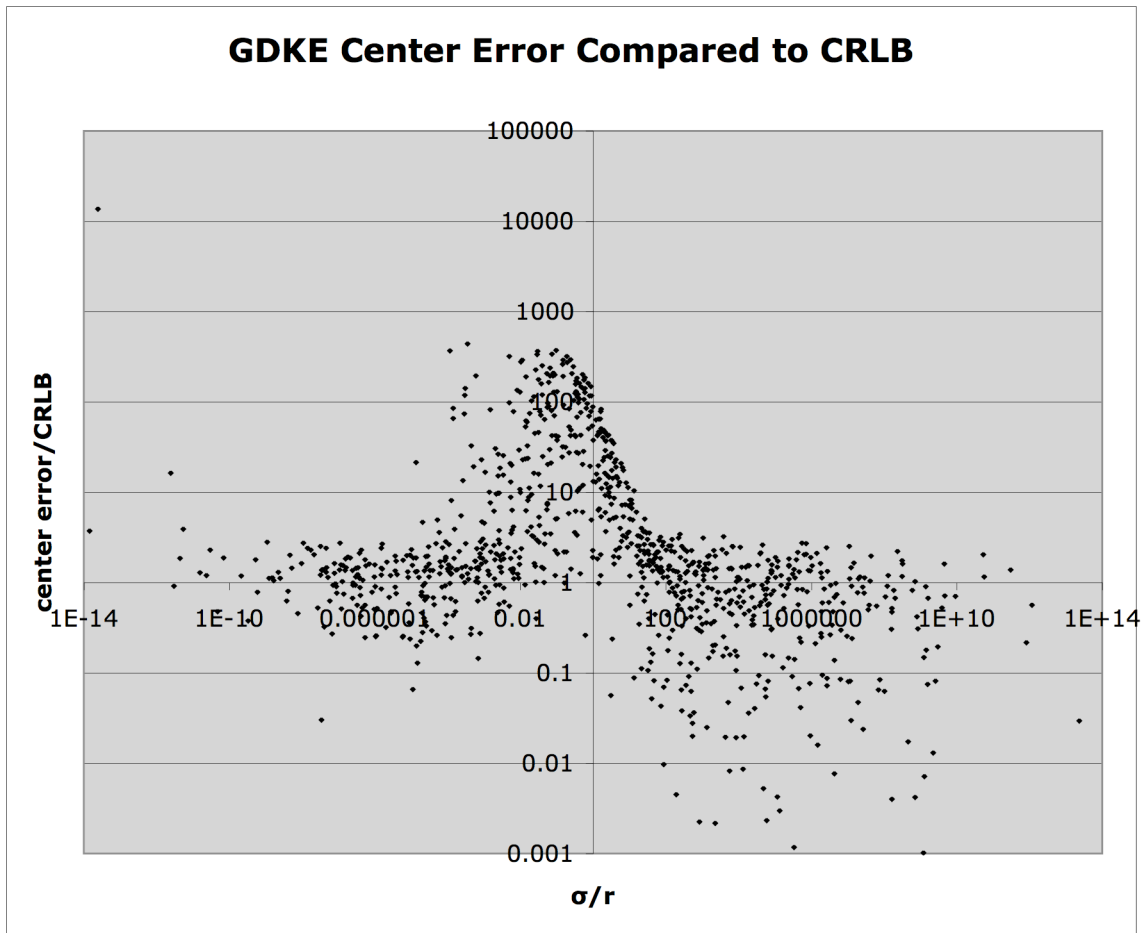


Figure 15 GDKE Compared to CRLB

The floating-point operations count for the GDKE of the center of a hypersphere of dimension D is

$$(89) \quad FLOPs = N(D^2 + 6D - 1) + \left(\frac{1}{3}D^3 + 3D^2 + \frac{8}{3}D + 1\right)$$

where N is the number of samples. The spherical solution ($D=3$) becomes

$$(90) \quad FLOPs = N26 + 45$$

You might ask at this point - it's faster, produces similar results, what's wrong with it? This fast method has not been fully accepted because it is a biased estimator. An estimator of a parameter is considered biased if it is expected to be a little off of the real

answer. Zelniker [119] has shown that the bias of the GDKE is on the order of the measurement standard deviation. The following statistical analysis will show exactly what the bias is and that there are cases that the bias is quite significant and does not disappear when more samples are taken.

4.1.5.2 Statistical Behavior

To understand completely the statistical behavior of the estimator, one must multiply multivariate Normal dependent variables six times to find the standard deviation of the center estimator. Another obstacle to finding the behavior is the inverse of a matrix. All these make it difficult to find an exact answer to the mean and standard deviation. A limiting approximation can be achieved if the individual components are analyzed.

Let's start off with our samples. Let us assume that each measurement x_i is a D -dimensional multivariate random Normal with constant covariance and differing means that can be expressed as

$$(91) \quad x_i \sim N_D(\mu_i, \Sigma)$$

where all of the means lie on the D -dimensional hypersphere.

$$(92) \quad |\mu_i - c_0| = r_0$$

If the average of the samples is taken, then the new number is still a multivariate Normal

$$(93) \quad \bar{x} \sim N_D(\bar{\mu}, \frac{1}{N}\Sigma)$$

Section 4.1.5.3 proves the following property of the sample covariance matrix:

$$(94) \quad E(\mathbf{C}) = \mathbf{C}_0 + \Sigma$$

$$(95) \quad \text{Cov}(\mathbf{C}, \mathbf{C}^T) = \frac{1}{(N-1)} \left(\Sigma(\mathbf{C}_0 + \text{Tr}(\mathbf{C}_0)) + (\mathbf{C}_0 + \Sigma)(\Sigma + \text{Tr}(\Sigma)) \right)$$

where

$$(96) \quad \mathbf{C}_0 = \frac{1}{N-1} \sum_{i=1}^N (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T.$$

The expectation for \mathbf{C} is here shown to be a biased estimator of the true covariance \mathbf{C}_0 . Most studies around this matrix are trying to estimate the measurement covariance Σ and say that the sample covariance matrix is unbiased, but that is only in the case when all samples have the same mean, thereby setting $\mathbf{C}_0=0$.

Similar analysis (cf. Section 4.1.5.4) will show that the expectation of the sample third central moment is

$$(97) \quad E(\mathbf{S}) = \mathbf{S}_0$$

where

$$(98) \quad \mathbf{S}_0 = \frac{1}{N-1} \sum_{i=1}^N (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T (\mu_i - \bar{\mu})$$

Now with the properties of the sample mean, sample covariance \mathbf{C} and sample third central moment \mathbf{S} , the following property holds for the GDKKE center estimator. The inverse of the sample covariance can be expanded by Leontief inverse as

$$(99) \quad \mathbf{C}^{-1} = (\mathbf{C}_0 + \Sigma)^{-1} + \sum_{k=1}^{\infty} \left(\mathbf{I} - (\mathbf{C}_0 + \Sigma)^{-1} \mathbf{C} \right)^k (\mathbf{C}_0 + \Sigma)^{-1}$$

Substituting this into the equation of the center produces

$$(100) \quad \hat{c} = \bar{x} + \frac{1}{2} (\mathbf{C}_0 + \Sigma)^{-1} \mathbf{S} + \sum_{k=1}^{\infty} \left(\mathbf{I} - (\mathbf{C}_0 + \Sigma)^{-1} \mathbf{C} \right)^k (\mathbf{C}_0 + \Sigma)^{-1} \mathbf{S}$$

This approximation results in the expectation of the center estimation as

$$(101) \quad E(\hat{c}) = c_0 + \frac{1}{2} \left((\mathbf{C}_0 + \Sigma)^{-1} - \mathbf{C}_0^{-1} \right) \mathbf{S}_0 + \dots$$

The covariance of the center estimator can be expanded to be

$$(102) \quad \begin{aligned} \text{Cov}(\hat{c}, \hat{c}^T) &= \frac{1}{N} \Sigma + \frac{1}{2} (\mathbf{C}_0 + \Sigma)^{-1} \text{Cov}(\mathbf{S}, \bar{x}^T) + \frac{1}{2} \text{Cov}(\bar{x}, \mathbf{S}^T) (\mathbf{C}_0 + \Sigma)^{-1} \\ &+ \frac{1}{4} (\mathbf{C}_0 + \Sigma)^{-1} \text{Cov}(\mathbf{S}, \mathbf{S}^T) (\mathbf{C}_0 + \Sigma)^{-1} + \dots \end{aligned}$$

The intermediate covariances are found to be independent (cf. Chapter 4.1.5.4)

$$(103) \quad \text{Cov}(\mathbf{S}, \bar{x}^T) = \text{Cov}(\bar{x}, \mathbf{S}^T) = 0$$

These reduce the covariance to

$$(104) \quad \text{Cov}(\hat{c}, \hat{c}^T) = \frac{1}{N} \Sigma + \frac{1}{4} (\mathbf{C}_0 + \Sigma)^{-1} \text{Cov}(\mathbf{S}, \mathbf{S}^T) (\mathbf{C}_0 + \Sigma)^{-1} + \dots$$

The covariance of the sample third central moment is

$$(105) \quad \text{Cov}(\mathbf{S}, \mathbf{S}^T) = \frac{1}{N} \sigma^2 \left(8\mathbf{F}_0 + \text{Tr}(\mathbf{F}_0) - 4\mathbf{C}_0^2 - 4\text{Tr}(\mathbf{C}_0)\mathbf{C}_0 - \text{Tr}(\mathbf{C}_0)^2 \right) + O\left(\frac{1}{N}\sigma^4\right)$$

which makes the covariance of the center estimator

$$(106) \quad \begin{aligned} \text{Cov}(\hat{c}, \hat{c}^T) &= \frac{1}{N} \sigma^2 \left(\mathbf{I} - (\mathbf{C}_0 + \sigma^2)^{-1} \mathbf{C}_0^2 (\mathbf{C}_0 + \sigma^2)^{-1} \right) \\ &+ \frac{1}{4N} \sigma^2 (\mathbf{C}_0 + \sigma^2)^{-1} \left(8\mathbf{F}_0 + \text{Tr}(\mathbf{F}_0) - 4\text{Tr}(\mathbf{C}_0)\mathbf{C}_0 - \text{Tr}(\mathbf{C}_0)^2 \right) (\mathbf{C}_0 + \sigma^2)^{-1} \\ &+ O\left(\frac{1}{N}\sigma^4\right) \end{aligned}$$

The radius estimator is similarly biased. Starting from the square of the radius estimator

$$(107) \quad \hat{r}^2 = \frac{N-1}{N} \text{Tr}(\mathbf{C}) + (\hat{c} - \bar{x})^T (\hat{c} - \bar{x})$$

The expectation of the square is then

$$(108) \quad E(\hat{r}^2) = \frac{N-1}{N} Tr(\mathbf{C}_0 + \Sigma) + Tr(Cov(\hat{c}, \hat{c}^T) + E(\hat{c})E(\hat{c}^T)) - 2E(\bar{x}^T \hat{c}) + \bar{\mu}^T \bar{\mu} + \frac{1}{N} Tr(\Sigma)$$

The intermediate expectation, which is related to the covariance between the mean and the center estimator, is

$$(109) \quad E(\hat{c}\bar{x}^T) = E(\bar{x}\bar{x}^T) + \frac{1}{2} E(\mathbf{C}^{-1}\mathbf{S}\bar{x})$$

$$(110) \quad E(\hat{c}\bar{x}^T) = c_0 \bar{\mu}^T + \frac{1}{N} \Sigma + \frac{1}{2} ((\mathbf{C}_0 + \Sigma)^{-1} - \mathbf{C}_0^{-1}) \mathbf{S}_0 \bar{\mu} + \dots$$

Inserting this into the expectation of the square, reduces it to

$$(111) \quad E(\hat{r}^2) = r_0^2 + \frac{N-1}{N} Tr(\Sigma) + \frac{1}{4} \mathbf{S}_0^T ((\mathbf{C}_0 + \Sigma)^{-1} + \mathbf{C}_0^{-1}) ((\mathbf{C}_0 + \Sigma)^{-1} - \mathbf{C}_0^{-1}) \mathbf{S}_0 + \dots$$

This expectation shows the bias in the radius estimator that remains non-zero as long as there is error in the measurement system. The variance of the square of the radius estimator comes out to be

$$(112) \quad Var(\hat{r}^2) = O\left(\frac{1}{N} \Sigma\right)$$

A typical example of the (mis-)use of these estimators can be displayed using Mathematica as in Figure 16.

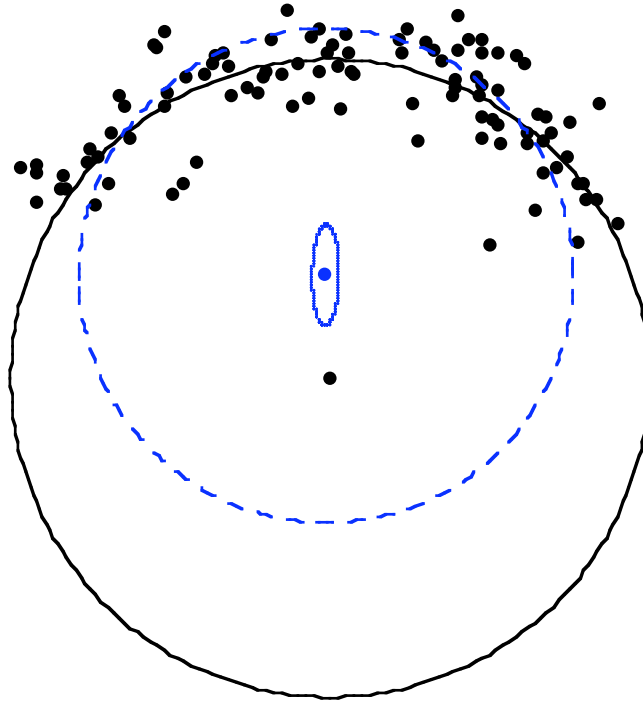


Figure 16 GDKE Error Ellipse

The dashed bars in Figure 16 represent the error involved in the center estimate. As can be seen, the true center is not anywhere within the error ellipsoid of the estimate. This is due solely to the bias in the estimator. Figure 16 is a good example of why the GDKE has not been adopted as much as the others. There is a simple way to remove this non-diminishing bias and it is the method of choice for speed and accuracy as long as a bit of a-priori knowledge is available (cf. Section 5.1).

4.1.5.3 Expectation of Sample Covariance

Theorem:

The sample covariance matrix has biased expectation when the samples are independently measured from the multivariate normal distribution with the same measurement covariance but differing expectations.

Proof:

The sample covariance is defined as

$$(113) \quad \mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$$

where the samples are independent of each other and come from the multivariate normal distribution

$$(114) \quad x_i \sim N_D(\mu_i, \Sigma)$$

The sample covariance can be expanded to

$$(115) \quad \mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (x_i x_i^T - \bar{x} x_i^T - x_i \bar{x}^T + \bar{x} \bar{x}^T)$$

When the summation is expanded, the covariance becomes

$$(116) \quad \mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N x_i x_i^T - \frac{N}{N-1} \bar{x} \bar{x}^T$$

and ultimately to

$$(117) \quad \mathbf{C} = \frac{1}{N} \sum_{i=1}^N x_i x_i^T - \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} x_i x_j^T.$$

The expectation of the sample covariance is

$$(118) \quad E(\mathbf{C}) = \frac{1}{N} \sum_{i=1}^N E(x_i x_i^T) - \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N E(x_i x_j^T)$$

Since these samples are independent, the expectation breaks down to

$$(119) \quad E(\mathbf{C}) = \frac{1}{N} \sum_{i=1}^N E(x_i x_i^T) - \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i}^N E(x_i) E(x_j)^T$$

The expectation of the same two multivariate normal vectors multiplied together is

$$(120) \quad E(x_i x_i^T) = \mu_i \mu_i^T + \Sigma$$

The expectation of the sample covariance then becomes

$$(121) \quad E(\mathbf{C}) = \mathbf{C}_0 + \Sigma$$

where

$$(122) \quad \mathbf{C}_0 = \frac{1}{N-1} \sum_{i=1}^N (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T.$$

The sample covariance is thus biased whether it is an estimate of Σ or \mathbf{C}_0 . The GDKKE is using the matrix as an estimate of \mathbf{C}_0 with Σ interfering. The variance of \mathbf{C} is more complicated involving six multivariate normal vectors multiplied together in a non-independent manner. A solution was achieved by analyzing the individual elements of the matrix. The covariance of each element in the matrix is defined as

$$(123) \quad \text{Cov}(\mathbf{C}_{ij}, \mathbf{C}_{mn}) = E(\mathbf{C}_{ij} \mathbf{C}_{mn}) - E(\mathbf{C}_{ij}) E(\mathbf{C}_{mn})$$

The square of the expectation of each element of the matrix is then

$$(124) \quad E(\mathbf{C}_{ij}) E(\mathbf{C}_{mn}) = \mathbf{C}_{0ij} \mathbf{C}_{0mn} + \mathbf{C}_{0ij} \Sigma_{mn} + \mathbf{C}_{0mn} \Sigma_{ij} + \Sigma_{ij} \Sigma_{mn}$$

where

$$(125) \quad \mathbf{C}_{0ij} = \frac{1}{N-1} \sum_{k=1}^N (\mu_{ki} - \bar{\mu}_i)(\mu_{kj} - \bar{\mu}_j) = \frac{1}{N-1} \sum_{k=1}^N \mu_{ki} \mu_{kj} - \frac{N}{N-1} \bar{\mu}_i \bar{\mu}_j$$

The expectation of the square of each element is much more complicated involving combinations of four random variates multiplied to each other. The square starts as

$$(126) \quad \begin{aligned} \mathbf{C}_{ij} \mathbf{C}_{mn} &= \frac{1}{(N-1)^2} \sum_{k=1}^N \sum_{l=1}^N x_{ki} x_{kj} x_{lm} x_{nl} \\ &\quad - \frac{N}{(N-1)^2} \sum_{k=1}^N x_{ki} x_{kj} \bar{x}_m \bar{x}_n + x_{km} x_{kn} \bar{x}_i \bar{x}_j \\ &\quad + \frac{N^2}{(N-1)^2} \bar{x}_i \bar{x}_j \bar{x}_m \bar{x}_n \end{aligned}$$

Since the points are independent of each other, the expectation of this is

$$(127) \quad \begin{aligned} E(\mathbf{C}_{ij} \mathbf{C}_{mn}) &= \frac{1}{(N-1)^2} \sum_{k=1}^N \sum_{l=1}^N E(x_{ki} x_{kj}) E(x_{lm} x_{nl}) \\ &\quad + \frac{1}{(N-1)^2} \sum_{k=1}^N E(x_{ki} x_{kj} x_{km} x_{nk}) - E(x_{ki} x_{kj}) E(x_{km} x_{nk}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l=1}^N \sum_{p=1}^N E(x_{ki} x_{kj}) E(x_{lm}) E(x_{pn}) + E(x_{km} x_{kn}) E(x_{li}) E(x_{pj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj} x_{km}) E(x_{ln}) + E(x_{km} x_{kn} x_{ki}) E(x_{lj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj}) E(x_{km}) E(x_{ln}) - E(x_{km} x_{kn}) E(x_{ki}) E(x_{lj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj} x_{kn}) E(x_{lm}) + E(x_{km} x_{kn} x_{kj}) E(x_{li}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj}) E(x_{lm}) E(x_{kn}) - E(x_{km} x_{kn}) E(x_{li}) E(x_{kj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj}) E(x_{lm} x_{ln}) + E(x_{km} x_{kn}) E(x_{li} x_{lj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N E(x_{ki} x_{kj}) E(x_{lm}) E(x_{ln}) - E(x_{km} x_{kn}) E(x_{li}) E(x_{lj}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N 2E(x_{ki} x_{kj} x_{km} x_{kn}) - E(x_{ki} x_{kj}) E(x_{km}) E(x_{kn}) \\ &\quad - \frac{1}{(N-1)^2 N} \sum_{k=1}^N E(x_{km} x_{kn}) E(x_{ki}) E(x_{kj}) \\ &\quad + \frac{N^2}{(N-1)^2} E(\bar{x}_i \bar{x}_j \bar{x}_m \bar{x}_n) \end{aligned}$$

$$\begin{aligned}
E(\mathbf{C}_{ij}\mathbf{C}_{mn}) &= \frac{1}{(N-1)^2} \sum_{k=1}^N \sum_{l=1}^N (\mu_{ki}\mu_{kj} + \Sigma_{ij})(\mu_{lm}\mu_{nl} + \Sigma_{mn}) \\
&+ \frac{1}{(N-1)^2} \sum_{k=1}^N \left(\begin{aligned} &\mu_{ki}\mu_{kj}\mu_{km}\mu_{kn} + \Sigma_{ij}\Sigma_{mn} + \Sigma_{im}\Sigma_{nj} + \Sigma_{in}\Sigma_{jm} \\ &+ \Sigma_{ij}\mu_{km}\mu_{kn} + \Sigma_{im}\mu_{kj}\mu_{kn} + \Sigma_{in}\mu_{kj}\mu_{km} + \Sigma_{jm}\mu_{ki}\mu_{kn} \\ &+ \Sigma_{jn}\mu_{ki}\mu_{km} + \Sigma_{mn}\mu_{ki}\mu_{kj} - (\mu_{ki}\mu_{kj} + \Sigma_{ij})(\mu_{km}\mu_{kn} + \Sigma_{mn}) \end{aligned} \right) \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l=1}^N \sum_{p=1}^N (\mu_{ki}\mu_{kj} + \Sigma_{ij})\mu_{lm}\mu_{pn} + (\mu_{km}\mu_{kn} + \Sigma_{mn})\mu_{li}\mu_{pj} \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N \left(\begin{aligned} &(\mu_{ki}\mu_{kj}\mu_{km} + \Sigma_{ij}\mu_{km} + \Sigma_{im}\mu_{kj} + \Sigma_{jm}\mu_{ki})\mu_{nl} \\ &+ (\mu_{km}\mu_{kn}\mu_{ki} + \Sigma_{mn}\mu_{ki} + \Sigma_{mi}\mu_{kn} + \Sigma_{ni}\mu_{km})\mu_{lj} \\ &- (\mu_{ki}\mu_{kj} + \Sigma_{ij})\mu_{km}\mu_{ln} - (\mu_{km}\mu_{kn} + \Sigma_{mn})\mu_{ki}\mu_{lj} \end{aligned} \right) \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N \left(\begin{aligned} &(\mu_{ki}\mu_{kj}\mu_{kn} + \Sigma_{ij}\mu_{kn} + \Sigma_{in}\mu_{kj} + \Sigma_{jn}\mu_{ki})\mu_{lm} \\ &+ (\mu_{km}\mu_{kn}\mu_{kj} + \Sigma_{mn}\mu_{kj} + \Sigma_{mj}\mu_{kn} + \Sigma_{nj}\mu_{km})\mu_{li} \\ &- (\mu_{ki}\mu_{kj} + \Sigma_{ij})\mu_{lm}\mu_{kn} - (\mu_{km}\mu_{kn} + \Sigma_{mn})\mu_{li}\mu_{kj} \end{aligned} \right) \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l \neq k}^N \left(\begin{aligned} &(\mu_{ki}\mu_{kj} + \Sigma_{ij})(\mu_{lm}\mu_{ln} + \Sigma_{mn}) \\ &+ (\mu_{km}\mu_{kn} + \Sigma_{mn})(\mu_{li}\mu_{lj} + \Sigma_{ij}) \\ &- (\mu_{ki}\mu_{kj} + \Sigma_{ij})\mu_{lm}\mu_{ln} - (\mu_{km}\mu_{kn} + \Sigma_{mn})\mu_{li}\mu_{lj} \end{aligned} \right) \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \left(\begin{aligned} &2 \left(\begin{aligned} &\mu_{ki}\mu_{kj}\mu_{km}\mu_{kn} + \Sigma_{ij}\Sigma_{mn} + \Sigma_{im}\Sigma_{nj} + \Sigma_{in}\Sigma_{jm} \\ &+ \Sigma_{ij}\mu_{km}\mu_{kn} + \Sigma_{im}\mu_{kj}\mu_{kn} + \Sigma_{in}\mu_{kj}\mu_{km} + \Sigma_{jm}\mu_{ki}\mu_{kn} \\ &+ \Sigma_{jn}\mu_{ki}\mu_{km} + \Sigma_{mn}\mu_{ki}\mu_{kj} \end{aligned} \right) \\ &- (\mu_{ki}\mu_{kj} + \Sigma_{ij})\mu_{km}\mu_{kn} - (\mu_{km}\mu_{kn} + \Sigma_{mn})\mu_{ki}\mu_{kj} \end{aligned} \right) \\
&+ \frac{N^2}{(N-1)^2} \left(\begin{aligned} &\bar{\mu}_i\bar{\mu}_j\bar{\mu}_m\bar{\mu}_n + \frac{1}{N^2}\Sigma_{ij}\Sigma_{mn} + \frac{1}{N^2}\Sigma_{im}\Sigma_{nj} + \frac{1}{N^2}\Sigma_{in}\Sigma_{jm} \\ &+ \frac{1}{N}\Sigma_{ij}\bar{\mu}_m\bar{\mu}_n + \frac{1}{N}\Sigma_{im}\bar{\mu}_j\bar{\mu}_n + \frac{1}{N}\Sigma_{in}\bar{\mu}_j\bar{\mu}_m + \frac{1}{N}\Sigma_{jm}\bar{\mu}_i\bar{\mu}_n \\ &+ \frac{1}{N}\Sigma_{jn}\bar{\mu}_i\bar{\mu}_m + \frac{1}{N}\Sigma_{mn}\bar{\mu}_i\bar{\mu}_j \end{aligned} \right)
\end{aligned}
\tag{128}$$

Expanding the sums and canceling terms reduces this gigantic equation to

$$\begin{aligned}
(129) \quad E(\mathbf{C}_{ij} \mathbf{C}_{mn}) &= \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l=1}^N \left(N \mu_{ki} \mu_{kj} \mu_{lm} \mu_{nl} + (N-1) \Sigma_{ij} \mu_{km} \mu_{nk} + (N-1) \Sigma_{mn} \mu_{ki} \mu_{kj} \right) \\
&+ \frac{1}{(N-1)^2 N} \sum_{k=1}^N \left((N-2) \Sigma_{im} \Sigma_{nj} + (N-2) \Sigma_{in} \Sigma_{jm} + N(N-2) \Sigma_{ij} \Sigma_{mn} \right) \\
&\quad + N \Sigma_{im} \mu_{kj} \mu_{kn} + N \Sigma_{in} \mu_{kj} \mu_{km} + N \Sigma_{jm} \mu_{ki} \mu_{kn} + N \Sigma_{jn} \mu_{ki} \mu_{km} \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l=1}^N \sum_{p=1}^N \left(\mu_{ki} \mu_{kj} \mu_{lm} \mu_{pn} + \mu_{km} \mu_{kn} \mu_{li} \mu_{pj} \right) \\
&- \frac{1}{(N-1)^2 N} \sum_{k=1}^N \sum_{l=1}^N \left(2 \Sigma_{jm} \mu_{ki} \mu_{nl} + 2 \Sigma_{mi} \mu_{kn} \mu_{lj} + 2 \Sigma_{ni} \mu_{lm} \mu_{kj} + 2 \Sigma_{jn} \mu_{ki} \mu_{lm} \right) \\
&\quad + N \Sigma_{ij} \mu_{lm} \mu_{kn} + N \Sigma_{mn} \mu_{li} \mu_{kj} \\
&+ \frac{N^2}{(N-1)^2} \left(\begin{aligned} &\bar{\mu}_i \bar{\mu}_j \bar{\mu}_m \bar{\mu}_n + \frac{1}{N^2} \Sigma_{ij} \Sigma_{mn} + \frac{1}{N^2} \Sigma_{im} \Sigma_{nj} + \frac{1}{N^2} \Sigma_{in} \Sigma_{jm} \\ &+ \frac{1}{N} \Sigma_{ij} \bar{\mu}_m \bar{\mu}_n + \frac{1}{N} \Sigma_{im} \bar{\mu}_j \bar{\mu}_n + \frac{1}{N} \Sigma_{in} \bar{\mu}_j \bar{\mu}_m + \frac{1}{N} \Sigma_{jm} \bar{\mu}_i \bar{\mu}_n \\ &+ \frac{1}{N} \Sigma_{jn} \bar{\mu}_i \bar{\mu}_m + \frac{1}{N} \Sigma_{mn} \bar{\mu}_i \bar{\mu}_j \end{aligned} \right)
\end{aligned}$$

And finally, more cancellations produce the answer

$$\begin{aligned}
(130) \quad E(\mathbf{C}_{ij} \mathbf{C}_{mn}) &= \mathbf{C}_{0ij} \mathbf{C}_{0mn} + \mathbf{C}_{0ij} \Sigma_{mn} + \Sigma_{ij} \mathbf{C}_{0mn} + \Sigma_{ij} \Sigma_{mn} \\
&+ \frac{1}{(N-1)} \left(\Sigma_{im} \Sigma_{jn} + \Sigma_{in} \Sigma_{jm} + \Sigma_{im} \mathbf{C}_{0jn} + \Sigma_{in} \mathbf{C}_{0jm} + \Sigma_{jm} \mathbf{C}_{0in} + \mathbf{C}_{0im} \Sigma_{jn} \right)
\end{aligned}$$

Subtracting the square of the expectation of each produces the covariance:

$$(131) \quad Cov(\mathbf{C}_{ij}, \mathbf{C}_{mn}) = \frac{1}{(N-1)} \left(\Sigma_{im} \Sigma_{jn} + \Sigma_{in} \Sigma_{jm} + \Sigma_{im} \mathbf{C}_{0jn} + \Sigma_{in} \mathbf{C}_{0jm} + \Sigma_{jm} \mathbf{C}_{0in} + \mathbf{C}_{0im} \Sigma_{jn} \right)$$

The full matrix covariance then becomes

$$(132) \quad Cov(\mathbf{C}, \mathbf{C}^T) = \frac{1}{(N-1)} \left(\Sigma (\mathbf{C}_0 + Tr(\mathbf{C}_0)) + (\mathbf{C}_0 + \Sigma) (\Sigma + Tr(\Sigma)) \right)$$

If the measurement covariance is $\Sigma = \sigma^2 \mathbf{I}$ then the variance reduces to

$$(133) \quad Cov(\mathbf{C}, \mathbf{C}^T) = \frac{1}{(N-1)} \sigma^2 \left(\mathbf{C}_0 (2 + D) + Tr(\mathbf{C}_0) + \sigma^2 (1 + D) \right)$$

$$(134) \quad Tr(\mathbf{C})^2 = \sum_{i=1}^D \sum_{j=1}^D \mathbf{C}_{ii} \mathbf{C}_{jj}$$

$$(135) \quad E(Tr(\mathbf{C})^2) = \sum_{i=1}^D \sum_{j=1}^D \mathbf{C}_{0ii} \mathbf{C}_{0jj} + \mathbf{C}_{0ii} \Sigma_{jj} + \Sigma_{ii} \mathbf{C}_{0jj} + \Sigma_{ii} \Sigma_{jj} + \frac{1}{(N-1)} \left(2 \Sigma_{ij} \Sigma_{ij} + 4 \Sigma_{ij} \mathbf{C}_{0ij} \right)$$

$$(136) \quad E\left(\text{Tr}(\mathbf{C})^2\right) = \text{Tr}(\mathbf{C}_0)^2 + 2\text{Tr}(\mathbf{C}_0)\text{Tr}(\Sigma) + \text{Tr}(\Sigma)^2 + \frac{1}{(N-1)}2\text{Tr}(\Sigma^2) + \frac{1}{(N-1)}4\text{Tr}(\mathbf{C}_0\Sigma)$$

The inverse of the sample covariance is an important value to find its expectation.

The inverse of any matrix can be expanded by the Leontief Inverse to

$$(137) \quad A^{-1} = \sum_{k=0}^{\infty} \left(I - E(A)^{-1}A \right)^k E(A)^{-1}$$

as long as the spectral radius (largest absolute eigenvalue) is

$$(138) \quad \rho\left(\mathbf{I} - E(A)^{-1}A\right) < 1$$

Substituting A for the sample covariance produces

$$(139) \quad \mathbf{C}^{-1} = (\mathbf{C}_0 + \Sigma)^{-1} + \sum_{k=1}^{\infty} \left(\mathbf{I} - (\mathbf{C}_0 + \Sigma)^{-1}\mathbf{C} \right)^k (\mathbf{C}_0 + \Sigma)^{-1}$$

An intermediate expectation is needed to get the second order approximation

$$(140) \quad \begin{aligned} E(\mathbf{CAC}) &= \mathbf{C}_0\mathbf{A}\mathbf{C}_0 + \mathbf{C}_0\mathbf{A}\Sigma + \Sigma\mathbf{A}\mathbf{C}_0 + \Sigma\mathbf{A}\Sigma \\ &+ \frac{1}{(N-1)} \left(\begin{aligned} &\Sigma\mathbf{A}\Sigma + \Sigma\text{Tr}(\mathbf{A}\Sigma) + \Sigma\mathbf{A}\mathbf{C}_0 + \mathbf{C}_0\mathbf{A}\Sigma \\ &+ \Sigma\text{Tr}(\mathbf{A}\mathbf{C}_0) + \text{Tr}(\Sigma\mathbf{A})\mathbf{C}_0 \end{aligned} \right) \end{aligned}$$

Then, using this identity, the expectation of the inverse expands to

$$(141) \quad \begin{aligned} E(\mathbf{C}^{-1}) &= (\mathbf{C}_0 + \Sigma)^{-1} \\ &+ \frac{1}{(N-1)} (\mathbf{C}_0 + \Sigma)^{-1} \left(\Sigma(2 + D) + \text{Tr}(\Sigma(\mathbf{C}_0 + \Sigma)^{-1})\mathbf{C}_0 - \Sigma(\mathbf{C}_0 + \Sigma)^{-1}\Sigma \right) (\mathbf{C}_0 + \Sigma)^{-1} \\ &+ \sum_{k=3}^{\infty} E\left(\left(\mathbf{I} - (\mathbf{C}_0 + \Sigma)^{-1}\mathbf{C} \right)^k \right) (\mathbf{C}_0 + \Sigma)^{-1} \end{aligned}$$

So the second order approximation using the diagonal covariance is

$$(142) \quad E(\mathbf{C}^{-1}) = \mathbf{C}_0^{-1} + \frac{1}{N-1} \sigma^2 \mathbf{C}_0^{-1} \left((2 + D)\mathbf{C}_0^{-1} + \text{Tr}(\mathbf{C}_0^{-1}) - \mathbf{C}_0^{-2} \sigma^2 \right)$$

4.1.5.4 Expectation of Sample Third Central Moment

Theorem:

The Sample Third Central Moment is unbiased when the samples are independently measured from the multivariate Normal distribution with the same covariance but differing expectations.

Proof:

The Sample Third Central Moment for multivariates is defined as

$$(143) \quad \mathbf{S} = \frac{1}{N-1} \sum_{k=1}^N (x_k - \bar{x})(x_k - \bar{x})^T (x_k - \bar{x})$$

where the samples are independent of each other and come from the multivariate Normal distribution

$$(144) \quad x_i \sim N_D(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$$

$$(145) \quad x_i - \bar{x} \sim N_D(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}, \frac{N-1}{N} \boldsymbol{\Sigma})$$

Then expectation of the moment is

$$(146) \quad E(\mathbf{S}) = \frac{1}{N-1} \sum_{k=1}^N E\left((x_k - \bar{x})(x_k - \bar{x})^T (x_k - \bar{x})\right)$$

Then using the expectation and covariance from (144) gives

$$(147) \quad E(\mathbf{S}_i) = \frac{1}{N-1} \sum_{k=1}^N \sum_{j=1}^D (\boldsymbol{\mu}_{ki} - \bar{\boldsymbol{\mu}}_i)(\boldsymbol{\mu}_{kj} - \bar{\boldsymbol{\mu}}_j)^2 + (\boldsymbol{\mu}_{ki} - \bar{\boldsymbol{\mu}}_i) \frac{N-1}{N} \boldsymbol{\Sigma}_{jj} + 2 \frac{N-1}{N} \boldsymbol{\Sigma}_{ij} (\boldsymbol{\mu}_{kj} - \bar{\boldsymbol{\mu}}_j)$$

and the last two parts disappear because of the definition of the average, leaving behind

$$(148) \quad E(\mathbf{S}) = \mathbf{S}_0 = \frac{1}{N-1} \sum_{i=1}^N (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T (\mu_i - \bar{\mu})$$

As can be seen, the sample third central moment is an unbiased estimator in the general case of constant covariance but different mean.

Another important property is the covariance between the sample third central moment and the sample mean. This value starts with expanding the sum as

$$(149) \quad \mathbf{S}\bar{\mathbf{x}}^T = \frac{1}{N-1} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \mathbf{x}_k \bar{\mathbf{x}}^T - \frac{2}{N-1} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \bar{\mathbf{x}} \bar{\mathbf{x}}^T - \frac{1}{N-1} \sum_{k=1}^N \bar{\mathbf{x}} \mathbf{x}_k^T \mathbf{x}_k \bar{\mathbf{x}}^T + \frac{2N}{N-1} \bar{\mathbf{x}} \bar{\mathbf{x}}^T \bar{\mathbf{x}} \bar{\mathbf{x}}^T$$

And expanding the means produce

$$(150) \quad \mathbf{S}\bar{\mathbf{x}}^T = \frac{1}{(N-1)N} \sum_{j=1}^N \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \mathbf{x}_k \mathbf{x}_j^T - \frac{1}{(N-1)N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N (2\mathbf{x}_k \mathbf{x}_k^T \mathbf{x}_i \mathbf{x}_j^T + \mathbf{x}_i \mathbf{x}_k^T \mathbf{x}_k \mathbf{x}_j^T) + \frac{2N}{N-1} \bar{\mathbf{x}} \bar{\mathbf{x}}^T \bar{\mathbf{x}} \bar{\mathbf{x}}^T$$

Separating the independent values produce the expanded sums

$$\begin{aligned}
E(\mathbf{S}\bar{x}^T) &= \frac{1}{(N-1)N} \sum_{j=1}^N \sum_{k=1}^N E(x_k x_k^T x_k) E(x_j^T) \\
&+ \frac{1}{(N-1)N} \sum_{k=1}^N E(x_k x_k^T x_k x_k^T) - E(x_k x_k^T x_k) E(x_k^T) \\
&- \frac{1}{(N-1)N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N 2E(x_k x_k^T) E(x_i) E(x_j^T) + E(x_i) E(x_k^T x_k) E(x_j^T) \\
&- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 3E(x_k x_k^T x_k) E(x_j^T) \\
&+ \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 2E(x_k x_k^T) E(x_k) E(x_j^T) + E(x_k) E(x_k^T x_k) E(x_j^T) \\
(151) \quad &- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 2E(x_k x_k^T \mu_j x_k^T) + E(x_j) E(x_k^T x_k x_k^T) \\
&+ \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 2E(x_k x_k^T) E(x_j) E(x_k^T) + E(x_j) E(x_k^T x_k) E(x_k^T) \\
&- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 2E(x_k x_k^T) E(x_j x_j^T) + E(x_j x_j^T) E(x_k^T x_k) \\
&+ \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N 2E(x_k x_k^T) E(x_j) E(x_j^T) + E(x_j) E(x_k^T x_k) E(x_j^T) \\
&- \frac{1}{(N-1)N^2} \sum_{k=1}^N 3E(x_k x_k^T x_k x_k^T) - 2E(x_k x_k^T) E(x_k) E(x_k^T) - E(x_k) E(x_k^T x_k) E(x_k^T) \\
&+ \frac{2N}{N-1} E(\bar{x} \bar{x}^T \bar{x} \bar{x}^T)
\end{aligned}$$

Substituting the expectations produce the sums

$$\begin{aligned}
E(\mathbf{S}\bar{x}^T) &= \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k=1}^N (N\mu_k \mu_k^T \mu_k \mu_j^T + NTr(\Sigma)\mu_k \mu_j^T + 2N\Sigma\mu_k \mu_j^T) \\
&+ \frac{1}{(N-1)N^2} \sum_{k=1}^N (2N\Sigma^2 + N\Sigma Tr(\Sigma) + 2N\mu_k \mu_k^T \Sigma + N\Sigma \mu_k^T \mu_k) \\
&- \frac{1}{(N-1)N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \left(2\mu_k \mu_k^T \mu_i \mu_j^T + 2\Sigma \mu_i \mu_j^T \right. \\
&\quad \left. + \mu_i \mu_k^T \mu_k \mu_j^T + Tr(\Sigma)\mu_i \mu_j^T \right) \\
&- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N (2Tr(\Sigma)\mu_k \mu_j^T + 4\Sigma\mu_k \mu_j^T) \\
&- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N (2\Sigma \mu_k^T \mu_j + 2\mu_k \mu_j^T \Sigma + 2\mu_j \mu_k^T \Sigma) \\
&- \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N \left(2\mu_k \mu_k^T \Sigma + 2\Sigma \Sigma + \mu_j \mu_j^T \mu_k^T \mu_k \right. \\
&\quad \left. + \Sigma \mu_k^T \mu_k + Tr(\Sigma)\Sigma \right) \\
&+ \frac{1}{(N-1)N^2} \sum_{j=1}^N \sum_{k \neq j}^N (\mu_j \mu_k^T \mu_k \mu_j^T) \\
&- \frac{1}{(N-1)N^2} \sum_{k=1}^N \left(6\Sigma^2 + 3\Sigma Tr(\Sigma) + 2Tr(\Sigma)\mu_k \mu_k^T + 6\mu_k \mu_k^T \Sigma \right) \\
&\quad \left(+4\Sigma\mu_k \mu_k^T + 3\Sigma \mu_k^T \mu_k \right) \\
&+ \frac{2N}{N-1} \left(\bar{\mu} \bar{\mu}^T \bar{\mu} \bar{\mu}^T + 2\frac{1}{N^2} \Sigma^2 + \frac{1}{N^2} \Sigma Tr(\Sigma) + \frac{1}{N} Tr(\Sigma) \bar{\mu} \bar{\mu}^T \right) \\
&\quad \left(+2\frac{1}{N} \bar{\mu} \bar{\mu}^T \Sigma + 2\frac{1}{N} \Sigma \bar{\mu} \bar{\mu}^T + \frac{1}{N} \Sigma \bar{\mu}^T \bar{\mu} \right)
\end{aligned} \tag{152}$$

This rather large sum simplifies all the way down to

$$E(\mathbf{S}\bar{x}^T) = \mathbf{S}_0 \bar{\mu}^T \tag{153}$$

thus proving that the sample third central moment is independent of the sample mean.

The covariance of the sample third central moment is much more difficult involving the multiplication of six multivariate normals together. To determine the covariance, the following matrix must be analyzed.

$$\mathbf{SS}^T = \frac{1}{(N-1)^2} \sum_{j=1}^N \sum_{k=1}^N (x_j - \bar{x})(x_j - \bar{x})^T (x_j - \bar{x})(x_k - \bar{x})^T (x_k - \bar{x})(x_k - \bar{x})^T \tag{154}$$

The expectation of this matrix is

$$\begin{aligned}
(155) \quad E(\mathbf{S}\mathbf{S}^T) &= \frac{1}{(N-1)^2} \sum_{j=1}^N \sum_{k \neq j}^N E\left(\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^T \left(x_j - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T\right) \\
&\quad + \frac{1}{(N-1)^2} \sum_{k=1}^N E\left(\left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T\right)
\end{aligned}$$

Separating out the covariance produces

$$\begin{aligned}
(156) \quad E(\mathbf{S}\mathbf{S}^T) &= \frac{1}{(N-1)^2} \sum_{j=1}^N \sum_{k \neq j}^N E\left(\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^T \left(x_j - \bar{x}\right)\right) E\left(\left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\right) \\
&\quad + \frac{1}{(N-1)^2} \sum_{k=1}^N E\left(\left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T\right) \\
&\quad + \frac{1}{(N-1)^2} \sum_{j=1}^N \sum_{k \neq j}^N Cov\left(\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^T \left(x_j - \bar{x}\right), \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\right)
\end{aligned}$$

These expectations are found in Section (7.1.1) and produces the covariance as

$$\begin{aligned}
(157) \quad Cov(\mathbf{S}, \mathbf{S}^T) &= \frac{N-1}{N^2} \left(\begin{aligned} &\Sigma Tr(\mathbf{F}_0) + 2\mathbf{F}_0 \Sigma + 2\Sigma \mathbf{F}_0 \\ &+ 4 \frac{N}{(N-1)^2} \sum_{k=1}^N (\mu_k - \bar{\mu})(\mu_k - \bar{\mu})^T \Sigma (\mu_k - \bar{\mu})(\mu_k - \bar{\mu})^T \end{aligned} \right) \\
&\quad + \frac{N-1}{N^2} \left(\begin{aligned} &4\Sigma \mathbf{C}_0 \Sigma + 8\Sigma^2 \mathbf{C}_0 + 8\mathbf{C}_0 \Sigma^2 + 2\mathbf{C}_0 \Sigma Tr(\Sigma) + 2\Sigma Tr(\Sigma) \mathbf{C}_0 \\ &+ 2Tr(\Sigma^2) \mathbf{C}_0 + 4\Sigma^2 Tr(\mathbf{C}_0) + 2\Sigma Tr(\Sigma) Tr(\mathbf{C}_0) \\ &+ 4\Sigma \frac{1}{N-1} \sum_{k=1}^N (\mu_k - \bar{\mu})^T \Sigma (\mu_k - \bar{\mu}) \end{aligned} \right) \\
&\quad + \frac{N-1}{N^2} \left(\Sigma Tr(\Sigma)^2 + 2\Sigma Tr(\Sigma^2) + 8\Sigma^3 + 4\Sigma^2 Tr(\Sigma) \right) \\
&\quad + \frac{1}{(N-1)^2} \sum_{j=1}^N \sum_{k \neq j}^N Cov\left(\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^T \left(x_j - \bar{x}\right), \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T \left(x_k - \bar{x}\right)\right)
\end{aligned}$$

The j and k vectors are not quite independent with their covariance being

$$(158) \quad Cov\left(\left(x_j - \bar{x}\right), \left(x_k - \bar{x}\right)^T\right) = -\frac{1}{N} \Sigma$$

This means the larger covariance in Equation (157) can be approximated by

$$(159) \quad \text{Cov}\left(\left(x_j - \bar{x}\right)\left(x_j - \bar{x}\right)^T, \left(x_k - \bar{x}\right)\left(x_k - \bar{x}\right)^T\right) = O\left(\frac{1}{N}\Sigma\mathbf{C}_0^2\right)$$

Using this and the special case of diagonal covariance

$$(160) \quad \Sigma = \sigma^2\mathbf{I}$$

produces the simplified equation for the covariance of the sample third vector moment as

$$(161) \quad \text{Cov}(\mathbf{S}, \mathbf{S}^T) = \frac{1}{N}\sigma^2\left(8\mathbf{F}_0 + \text{Tr}(\mathbf{F}_0) - 4\mathbf{C}_0^2 - 4\text{Tr}(\mathbf{C}_0)\mathbf{C}_0 - \text{Tr}(\mathbf{C}_0)^2 + 38\sigma^2\mathbf{C}_0 + 14\sigma^2\text{Tr}(\mathbf{C}_0)\right) + O\left(\frac{1}{N}\sigma^6\right)$$

4.1.5.5 Center Equation Theorem

This section proves that the GDKE estimator of the center of the hypersphere is truly the center of the sphere when the points all lie on the sphere.

Theorem:

The Center Equation (86) calculates the true center of the hypersphere if all points are equidistant from the center, i.e. they all lie on the surface of the same hypersphere.

Proof:

The Center Equation can be rearranged to be

$$(162) \quad \mathbf{0} = 2\mathbf{C}(\bar{x} - c_0) + \mathbf{S}.$$

Expanding this into the summation produces

$$(163) \quad \mathbf{0} = 2\sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T (\bar{x} - c_0) + \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T (x_i - \bar{x})$$

Combining the two summations gives

$$(164) \quad \mathbf{0} = \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T (x_i + \bar{x} - 2c_0)$$

The difference can be changed to be around the center instead of the mean thus

$$(165) \quad \mathbf{0} = \sum_{i=1}^N (x_i - \bar{x})(x_i - c_0 + c_0 - \bar{x})^T (x_i - c_0 + \bar{x} - c_0)$$

Then expanding the second and third term under the summation produces

$$(166) \quad \mathbf{0} = \sum_{i=1}^N (x_i - \bar{x}) \left((x_i - c_0)^T (x_i - c_0) - (\bar{x} - c_0)^T (\bar{x} - c_0) \right)$$

The square of the distance of point i from the center is defined as

$$(167) \quad r_i^2 \equiv (x_i - c_0)^T (x_i - c_0)$$

The summation can be then simplified to

$$(168) \quad \mathbf{0} = \sum_{i=1}^N (x_i - \bar{x}) r_i^2 - (N\bar{x} - N\bar{x})(\bar{x} - c_0)^T (\bar{x} - c_0)$$

The second term is then eliminated, producing the simple summation

$$(169) \quad \mathbf{0} = \sum_{i=1}^N (x_i - \bar{x}) r_i^2$$

It is easy to see that if all points are equidistant (i.e. $r_i=r$) from the center then this equation becomes true. There are other ways to produce zero with this sum, but that is where the GDKE steps in. The GDKE solves this equation when the distances are not equal.

4.2 Skeleton Approaches

Producing a skeleton involves finding the centers of joint rotations, which was addressed in Section 4.1, the hierarchy of segmentation, and the orientation of each segment. This thesis does not rely on novel ideas for the remaining steps of producing a skeleton. The hierarchy of segmentation is known ahead of time. The orientation of a segment is determined by one of two methods. Either it is given in the raw data, e.g. magnetic trackers, or it is calculated by the fastest technique known. The fastest way to calculate the orientation of a segment from positional information of markers is quite intuitive and has been in use for a long time. One of the earliest uses found in motion capture were published in 2000 by Herda, et al. [50]. Unfortunately, many authors don't realize that this speedy method of orienting the segment can also be used to calculate the entire skeleton. The most common approach to calculating a skeleton from motion capture data is through minimization until the skeleton fits where the rotation points have been approximated. The minimization involves squishing segments and moving joints until all joints are nearest to the calculated rotation points. O'Brien, et al. [84] uses a linear least-squares minimization that produces the rotation points from a collection of time frames. Their method relies on the constraint that two connected segments have a common point between them during rotation. Their solution calculates a point from the knowledge of the orientation of both the parent and the child segments and calculates the best fit point that remains the most still relative to the two segments. The solution involves the Singular Value Decomposition of a $3N \times 6$ matrix to produce the common point that is the rotation point. If the solution fails to come up with an answer, as in the case of no motion or planar motion, then the closest point between the two coordinate systems is

used. This technique relies heavily on orientation information that is not always available. In their study, magnetic motion tracking devices are used which contain both position and orientation. This is akin to solving for the best-fit sphere around a center. The state of the science for sphere fitting was presented in Section 4.1 but the improvement upon the state is now presented in Section 5.1.