Axyz

Mathematics for Users

Authorship

Concept, text and diagrams: Dr. Stephen Kyle, Leica

Document reference data

Title:	*Axyz* Mathematics for Users
Order number:	722 238
Software version:	1.4
Document release date:	January 2000

Contents

1. Introduction

1.1 About this manual

This manual provides a descriptive explanation of a number of critical mathematical concepts and methods used by *Axyz*.

It does not provide a detailed specification of internal algorithms but the explanations should enable readers to understand how the functions work.

It is assumed that readers have a level of mathematical knowledge appropriate for a university-educated engineer. It is further assumed that they are familiar with the general techniques of optical triangulation and polar location.

2. Simple statistical ideas

2.1 Errors

2.1.1 Errors In brief

An error is the difference between the true value of the measurement and what you actually measure. Errors have 3 basic sources:
- Small random variations (a fact of life beyond the user's control)
- Systematic effects (can be compensated by suitable modelling methods)
- Mistakes (can often be detected by check procedures)

2.1.2 Introduction to errors

Measurements can never be exactly correct, which is an unavoidable fact of life. If the same angle or distance is measured many times, there will always be a variation in results. Additionally, no instrument can be perfectly manufactured, and some residual imperfections must remain, even if small. Both these effects introduce measurement *errors*. The term error is used in statistics to refer to the small deviations of measurements from their true value although error in the normal English sense means *mistake*. This term is reserved for more serious errors.

We try to keep errors a small as possible or adopt measuring strategies which reduce their effects and enable us to get good estimates of the true values. It is useful to identify different types of error.

Systematic errors arise, for example, when a theodolite's line of sight is not exactly perpendicular to its transit (trunnion) axis. These can be identified and largely eliminated by mathematical modelling and software compensation.
Outright *mistakes* might be caused by the operator pointing at the wrong target because it looks the same as the correct one. Even automated systems can make mistakes, for example when an image processing module locates a bright light source in the background instead of the correct target which happens to be dimmer. Additional measurements and diverse data filtering methods often detect mistakes which usually have large and isolated effects. To properly identify a true mistake it may also be called a *gross error* or *blunder*.

Quality issues are mainly concerned with *random errors*, such as a short-term temperature change which causes a small refraction error in a

pointing. These are beyond the control of the user. Fortunately we can reduce the effects of random errors by averaging repeated measurements or using more information than is strictly needed

2.1.3 Random errors

Random errors are small positive and negative variations in the value of a measurement which is repeated many times. This is a natural physical effect which cannot be completely eliminated by changing the design of the measuring system. Random measurement errors follow the Normal or Gaussian error distribution.

2.1.4 Systematic errors

Systematic errors follow a definite pattern caused by some particular physical effect. For example an electronic distance meter may give readings which are consistently too low or too high if the incorrect carrier frequency has been applied.

2.1.5 Mistakes (gross errors, blunders)

Mistakes are caused by some failure in the measuring procedure. For example an operator may accidentally sight the wrong target or an electrical spike may corrupt the reading from an instrument. In a properly designed measuring procedure mistakes rarely occur and are usually large compared to other errors. This makes it relatively easy to find them.

2.2 The normal distribution of random errors

Although individual random errors, as the name implies, do not seem to follow a pattern, in large numbers they do behave in a predictable way. Most people are aware of the simple concept that you repeat a measurement many times if you want to "average out" the variations in each one. This implies, correctly, that you can in the end get the true value of a quantity even if each separate attempt is close but not exact.

The whole science of statistics is built on this idea that there is indeed some predictable behaviour in random physical changes. However the rules strictly only apply when large numbers of measurements are involved. Naturally we try to get away with quite small numbers of measurements and in practice this works well. However it is also possible to make too few measurements, in which case any estimates of quality are based on insufficient information and produce bad results.

Imagine measuring a distance between the same two points many times. You might get the following random spread of distance measurements:

Value mm	Number of measurements	Relative frequency	Deviation from average value
810.11	1	0.005	-0.059
810.12	3	0.015	-0.049
810.13	7	0.035	-0.039
810.14	19	0.095	-0.029
810.15	20	0.100	-0.019
810.16	36	0.180	-0.009
810.17	38	0.190	0.001
810.18	29	0.145	0.011
810.19	24	0.120	0.021
810.20	10	0.050	0.031
810.21	11	0.055	0.041
810.22	0	0.000	
810.23	2	0.010	0.061

Average	Total
810.17	200

The table shows how many times a particular value came up. The values did not, of course, turn up in the order shown in the table, which involved some re-packaging for the presentation. The relative frequency is simply each measurement number divided by the total number of measurements and it offers a convenient way to compare tests with different measurement numbers.
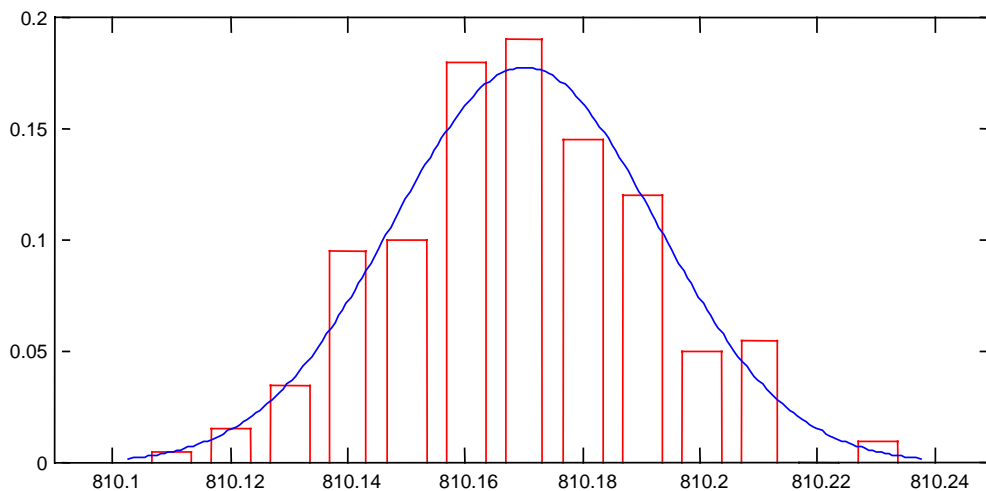
The diagram shows the same information as in the table, with the height of the red blocks indicating the relative frequency of each of the measurements. The blue, bell-shaped curved shows the pattern that the blocks fall into if you take very many measurements, although the red blocks already fit the curve quite well.

The curve indicates the probability of any particular measurement value occurring. In particular it shows that somewhere around the middle, where the average value lies, is the most probable region for obtaining measurements.

This curve is so common that it is called the *normal distribution* (error curve) where normal has its conventional meaning in English. It may also be called the *Gaussian distribution* after the German mathematician Karl Friedrich Gauss who made detailed investigations of the effects and properties of errors.

2.2.1 Probability density
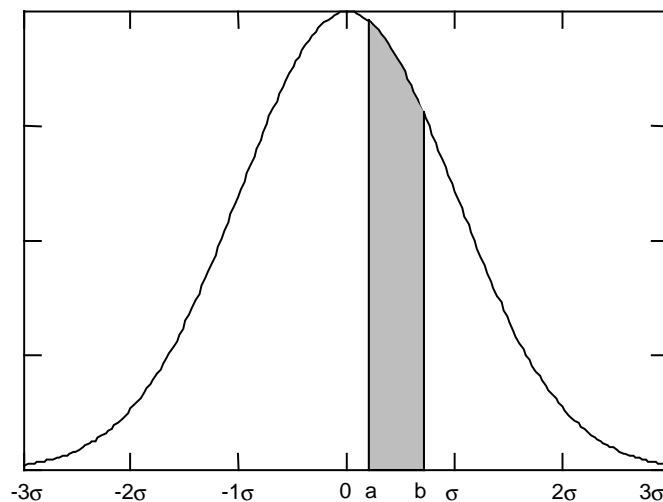
The curve describing the normal error distribution actually represents probability *density*, i.e. a probability per unit. It can really only be used by asking: "What is the probability that a measurement lies between value a and value b?" The shaded area under the curve then indicates the probability. You do not use the curve to ask: "What is the probability that a measurement has value c".

Imagine you have a bar made of some composite material whose density varies in the same bell-shaped way. In casual conversation you can say: "The bar is heavier in the middle than at the ends" and the fact that the

density of the material is highest in the middle clearly indicates this. However to be quite specific you would have to say that the weight of a section cut from the middle is heavier than the weight of a similar length section cut from the end. To obtain a section you must cut it out between positions a and b but it is meaningless to refer to the weight of a cross section at position c.

2.3 How good is the measurement?

In conversation we use several qualitative terms to explain how well we can measure:

- Accuracy
- Precision
- Repeatability
- Resolution

Without proper definition the conversation can lead to confusion For example:
"The robot can position its end effector with an accuracy of 0.1mm". Will it always return to the same position within 0.1mm, or is the separation of any 2 points within its entire workspace never in error by more than that amount?

"This measurement system has a resolution of 10 microns". Is that the smallest increment on the linear encoders, or the smallest increment at the object which can be reliably determined? Either way, how good is it?

2.3.1 Accuracy

Accuracy is a global effect, extending throughout the measurement field. It indicates how close measurements and derived quantities are to their true values. Although true values can never be found in practice, calibration and performance tests provide accuracy checks.

2.3.2 Precision

Precision is a local effect. It indicates how well measurements of a particular quantity agree with one another.

- If precision is high the spread of values is small.
- If precision is low the spread of values is large.

Repeatability tests can show the precision of individual instruments and complete systems.

2.3.3 Repeatability

Repeatability is effectively another term for precision. Precision tends to be used in connection with measuring instruments and repeatability with a complete measuring procedure or particular objects.

- A well maintained laser tracker is a *precise* instrument.
- An well maintained industrial theodolite has high angular *precision*.
- A dual theodolite triangulation system should give good *repeatability* when a set of test targets is measured on several occasions.
- A robot ideally shows good *repeatability* when returning to the same programmed pose.

2.3.4 Calibration

1) Not precise
   Not accurate

2) Precise
   Not accurate

3) Accurate
   Not precise

4) Accurate
   AND precise
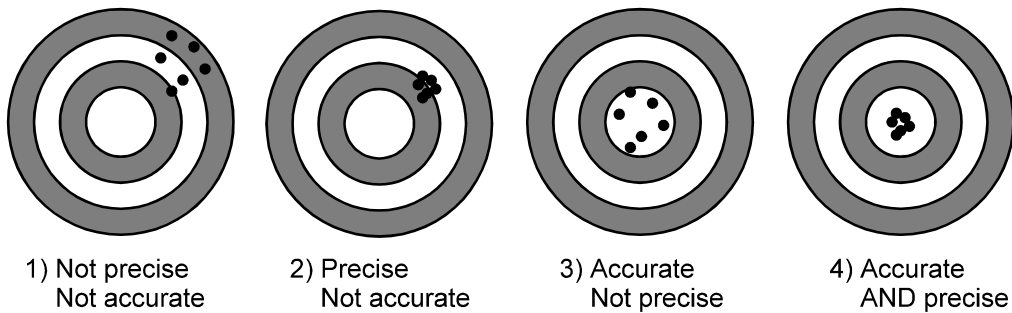
STATS01.WMF

Consider the results of the shooting practice session above. The precision of a process indicates the potential accuracy which it can achieve but only by making a suitable *compensation* or *calibration* can this accuracy be realized. In sessions (3) and (4) the errors in (1) and (2) respectively have been corrected.

Lack of accuracy is a good indicator that some systematic error is still present but once removed by calibration the random effects which cause the spread seen in the precision (repeatability) will still be present. Calibration cannot remove these as well.

2.3.5 Resolution

Resolution In brief

The resolution of a measuring component such as an angle encoder is the smallest incremental change which it can deliver. Resolution should be higher than the expected measurement accuracy so that the measurement is not degraded by an inefficient encoder.

The resolution of an optical system such as a telescope is a measure of its ability to reproduce fine detail in the image. Fine detail implies a sharp image and this contributes to a high quality pointing for both manual observation and electronic imaging.

Introduction to resolution

Optical resolution and instrument readings obviously influence precision and accuracy, but it should not be assumed that they indicate the limits. An angle read to 0.1" will not necessarily have a precision or accuracy of that order. These depend more on the ability of the observer or electronic sensor to centre on the target and on the quality and definition of the target itself.

Equally well, if a telescope can be reliably pointed to within ± 2 arc secs., then a system which only reads to 10 arc secs will not reflect the potential accuracy. Furthermore, repeated pointings to improve the result will not, in this case, be successful, since the small spread cannot be detected by the coarse reading. Generally, the minimum increment given by an instrument should be somewhat smaller than the expected precision of the quantity measured.

2.3.6 Significant figures

The number of significant figures indicates how many digits in a numerical value have any importance.

9.0075 m is a value with 5 significant figures. If this is a reading from an electronic distance meter it suggests the distance is accurate to about ½ mm but the instrument might display a reading to another place of decimals, e.g. 9.00753.

If the ½ mm is already in doubt, the final digit (9.0075**3**) contributes no information to the measurement.

However when making further computations with this value, small errors can arise which are purely due to mathematical manipulation and have nothing to do with the actual measurements. Rounding errors are an example of this effect.

To reduce errors caused by this source, the extra digit is often carried through the processing.

2.4 How are uncertainties quantified?

Now that general terms for discussing quality have been specified, concrete numbers need to be generated.

2.4.1 Average value

One of the most common ways of improving results is to take several measurements of the same quantity and use the *average value*, also called the *mean value*.

Average value In brief

For a set of N measurements of a quantity X:

$$X_{av} = \sum_{i=1}^{N} \frac{X_i}{N} \quad \text{or} \quad X_{av} = \frac{X_1 + X_2 + .. + X_N}{N}$$

More on average value

If the error quantities x_i are the deviations from the average value such that $X_i = X_{av} + x_i$, then

$$X_{av} = \frac{(X_{av} + x_1) + (X_{av} + x_2) + .. + (X_{av} + x_N)}{N}$$

which leads to

$$\frac{(x_1) + (x_2) + .. + (+ x_N)}{N} = 0$$

i.e. the average value of the error quantities is zero.
These error quantities x_i are not the true errors e_i since the average value is not the true value, although the more measurements you make the more likely the average is to be close to the true value. However the true errors will also have an average value of zero.

2.4.2 Variance and standard deviation

The spread of random variations is obviously an indicator of measurement quality. The quality of measurement is low if the spread is large and high if the spread is small.

Considering the error values themselves, they are sometimes positive and sometimes negative, with an average value of zero. The average error provides no information about the spread, also called *dispersion*. If instead the error terms are first squared and then averaged a positive number is always obtained. The average of the squared terms does indicate the spread, since it increases as the spread increases. This parameter is called the *variance*. (In normal English the term relates to some difference or lack of agreement as in the phrase: "Statement A is at variance with statement B".)

Variance is defined by the true errors e_i as:

$$\text{var} = \sum_{i=1}^{N} \frac{(e_i)^2}{N} \qquad \text{or} \qquad \text{var} = \frac{(e_1)^2 + (e_2)^2 + .. + (e_N)^2}{N}$$

Since the variance involves squared error terms it is not an easily recognized number. A more useful figure is the square root of the variance, called the standard deviation. It is commonly identified by the Greek symbol σ (sigma):

$$\sigma = \sqrt{\text{var}} \qquad \text{or} \qquad \sigma = \sqrt{\frac{(e_1)^2 + (e_2)^2 + .. + (e_N)^2}{N}}$$

(The variance itself is then usually identified by the squared term, σ^2.)
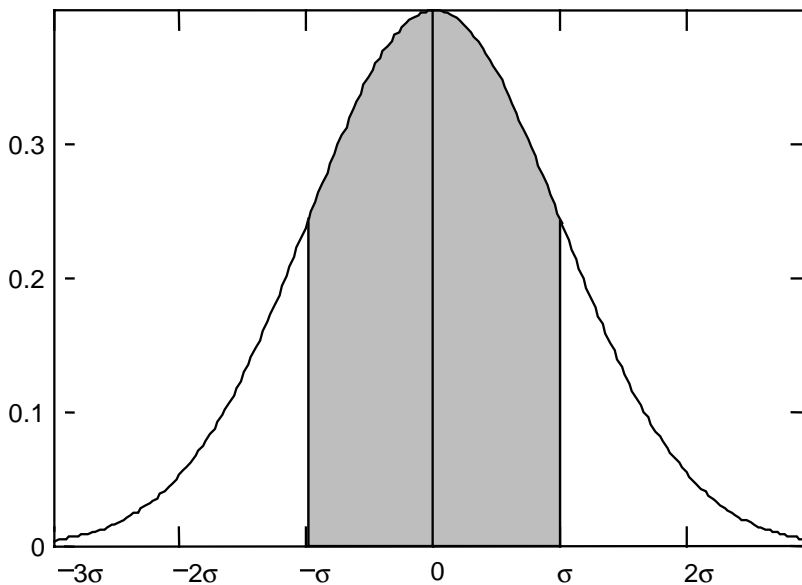
It can be shown that a measurement has a 68% chance of lying within $\pm\sigma$ of the true value, if the measurements follow the normal distribution.

Strictly speaking, these definitions only apply to an infinitely large set of measurements. Since most users have better things to do than take repeat measurements all day we must make do with a lot less. If instead the deviations x_i from the average value of a small number of measurements are used, then statisticians can prove that a good, unbiased approximation to the standard deviation is given by:

$$sd = \sqrt{\frac{(x_1)^2 + (x_2)^2 + .. + (x_N)^2}{N-1}}$$

The only practical difference is the use of (N-1) instead of N and the difference is small once you have, say, 10 or more measurements. In fact, if you take a less rigorous view of things and say:

$$\sigma = sd = \sqrt{\frac{(x_1)^2 + (x_2)^2 + .. + (x_N)^2}{N}}$$

then this becomes a definition of the *root mean square (RMS)* value. (See "Root Mean Square (RMS) error" on page 31.)

Very often we do not make a number of repeat measurements of the same quantity but do measure similar quantities many times. For example, in an orientation procedure involving a *bundle adjustment* many angular readings are taken but to a number of different points. The bundle adjustment can then produce quality figures such as the RMS value for angle measurements in general.

2.4.3 Correlation and covariance

Random errors in measurements are usually independent of each other. The random error in pointing at target A is not affected by the random error in pointing at target B.

Some measurements are not made directly but are derived entirely from others. In *Axyz*, angle measurements are used to derive coordinates. For example, pointings to a specific target from two theodolites can be intersected to compute the target's X,Y and Z coordinates. Errors in each coordinate value, dX, dY and dZ, are based on the angular errors of the two pointings. Although the individual coordinate errors are different because a different function is used to compute each one, they will still be

related since they all use the same starting information, i.e. the same errors in the same two angular pointings.

If errors are not independent but related in some way they are known as *correlated errors*. Just as a variance can be defined for any set of errors as a measure of their spread, so a *covariance* can be defined for any two sets of correlated errors and used as a measure of their correlation.

Correlations must always be taken into account when computing errors in one set of variables which are derived from another. In the example given, correlations appear in target coordinates even though the basic pointing information is uncorrelated. Any further processing of the coordinates must take the corresponding covariances into account, for example, when transforming the coordinates into another coordinate system. If this is not done the new variance values, which are usually the ones of interest, will not be correctly calculated.

Covariance is calculated for any two measurements in a similar way to variance. Suppose the two measurements A and B are repeated N times, generating two sets of deviations from mean values, a_i and b_i. Using the standard symbol σ_{ab} for the covariance between A and B:

$$\sigma_{ab} = \sum_{i=1}^{N} \frac{a_i \cdot b_i}{N} \qquad \text{or} \qquad \sigma_{ab} = \frac{a_1 \cdot b_1 + a_2 \cdot b_2 + .. + a_N \cdot b_N}{N}$$

Note that σ_{ba} will clearly be the same as σ_{ab}.

If the measurements are uncorrelated they will randomly have positive and negative values. The individual product terms will therefore also be sometimes positive and negative and the covariance will tend to average out to zero. However if a positive error in A tends to be associated with a positive error in B, then the covariance will tend to be positive. The same applies if a negative error in A correlates with a negative error in B. The signs can also go in opposite directions, which would result in a negative covariance.

2.4.4 The variance/covariance matrix

This is often called simply the *covariance matrix*. It is required in many of the procedures which make use of matrix analysis and is often one of the by-products. It provides a concise summary of the variances and

covariances between different measurements which need not be of the same type. A typical mixture would involve angle measurements of varying quality and distance measurements of varying quality.

Covariance matrix:

$$\begin{bmatrix} (\sigma_1)^2 & \sigma_{12} & .. & .. \\ \sigma_{21} & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}$$

For measurements 1,2 .. N the variances appear on the corresponding diagonal positions. The covariances between pairs of elements appear in the corresponding off-diagonal elements. Since the covariance between measurements j and k is the same as between k and j, this matrix is symmetrical.

In the simple case of uncorrelated measurements this matrix becomes a diagonal matrix of variance values.

$$\begin{bmatrix} (\sigma_1)^2 & 0 & .. & .. \\ 0 & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}$$

2.4.5 Tolerance

In manufacturing it is convenient to deal in numbers which cover all possible deviations from design values. Although this is literally impossible, a good practical approach is to quote error boundaries on measurements such that, say, more than 99% of all cases are covered. This range is known as the tolerance.

If a good estimate of the standard deviation of a particular measurement is known, a typical tolerance quote would be $\pm 3\ \sigma$. There is then a 99.7% chance that a measurement will fall between these limits.

The 3-sigma value is not a standard and other values are used. For example, $\pm 2.5\ \sigma$ is common and it still covers almost 99% of all the relevant measurements.

3. Least squares and modelling

3.1 Introduction to least squares

Using ***Axyz*** we make angle and distance measurements in order to calculate coordinates which describe the shapes of objects. The coordinates themselves can be further processed to find out more shape information. We therefore need a wide range of processing routines which convert measurements to coordinates and coordinates into shape information. If the original data source, the measurements, were free of error all the routines would look mathematically very different. Since they and therefore the coordinates are not perfect, all the routines at some point use a general mathematical technique called *least squares*. Once the particular set of equations have been constructed which describe the actual task, such as an orientation or shape fit, the least squares method solves for the values we want in an optimal way which resolves the conflicts caused by incompatible data.

Since measurements are subject to error, methods which process them must give us the best estimate of their true, error-free values. A crude method would be to make every measurement many times and take the average, but least squares offers a better solution.

The method of least squares is a very common technique for processing measurements which have small random variations and are therefore not *consistent*. A simple example is the measurement of all three angles of a triangle. They should add up to 180° but because of the random error in each this is very unlikely to happen. The 3 measured angles are therefore not consistent with the laws of geometry. Any two of the angles could be used to describe the shape of a plane triangle, and 3 slightly different triangles would be obtained in this way. Which one is the "true" triangle?

Using least squares the problem is resolved by creating a <u>single</u> mathematical model of a situation and deriving equivalent and <u>exact</u> mathematical measurements from it. These modelled values are compared with the actual measurements and the model altered step by step until a best fit between modelled measurements and actual measurements is obtained. The decision on the best fit is reached by examining the sum of the squares of the differences and altering the model until this sum is a minimum. The model is then assumed to be the best description of the actual measurement situation.

Amongst the many methods which could be used to find the parameters of a model (instrument locations, circle radius, etc.), it can be shown that the least squares technique gives the best unbiased estimates of these parameters.

In **Axyz** the principle of least squares is at the core of all the optimized methods of transformation, shape fitting, orientation, target location and calibration. In addition to providing an optimal answer such as the radius of a best fitting circle, the methods can also supply quality estimates for both measurements and modelled parameters.

> **Note**
> The procedure involving stepwise changes is the most common one although there are special cases where the model can be created in a single step.

3.1.1 Mathematical components

In mathematical terms, a least squares analysis has the following components:

1. The initial parameters of the model, e.g.
   Positions of instruments and targets in an orientation procedure
   The values which define a circle in a shape fitting procedure

2. Values derived from the model which can be compared with corresponding measurements, e.g.
   Horizontal and zenith angles (h,zn) from a modelled instrument to a modelled target position
   The modelled radius of the circle (r)

3. Corresponding values involving the known measurement quantities, e.g.
   Measured horizontal and zenith angles (H,ZN)
   The distances from measured points to the modelled circle centre (R)

4. Differences, known as *residuals*, between theoretical values from (2) and measured values from (3) i.e.
   $v_h = h - H$, $v_{zn} = zn - ZN$
   $v_r = r - R$

5. Equations which make the comparison between (2) and (3) in terms of the parameters in (1) and which also involve the residuals in (4).

In a simple least squares procedure, the quantity Φ is then minimized where

$$\Phi = (v_h{}^2 + v_{zn}{}^2 + v_r{}^2 + ...)$$

The minimization is done by altering the parameters in (1) which changes the values in (2) and possibly also in (3) until the residuals in (4) make Φ a minimum. The alteration of parameters is done in controlled steps using the equations in (5) to solve for improvements.

3.2 Iteration and non-linear solutions

In mathematics, repeating a process of comparison and modification until some value has been optimized is called *iterating* towards a solution and each of the steps is a single *iteration*. (In normal English we use the word reiterate meaning to repeat again, as in the sentence "Let me reiterate what I said before ..")

Iteration is needed when it is not possible to compute an optimized answer in a single step. This is because many mathematical formulations are *non-linear*, i.e. the modelled measurements and parameters appear in squared, cubic and higher powers or are multiplied together. With a few exceptions a one-step solution does not then exist.

A one-step solution mostly only exists when the formulation contains no product terms of parameters, i.e. parameters are only multiplied by constant and known values and not by themselves or by other parameters.

Suppose a calibration method uses least squares to find the axis tilt, A, and beam tilt, B, of an instrument, and that different models are possible to do this.

Typical equations which define the model	Type of model
$5 * A + 29.8 * B = 0.00032$	linear
$0.5 * A^2 + 7.66 * B = 0.0000097$	non-linear (note the squared term A^2)
$8.4 * A + 0.35 * A * B = 4.2$	non-linear (note the product term AB)

These are not real equations but simply illustrate the difference between linear and non-linear formulations.

When a model is non-linear a technique of *linearization* is possible. This assumes you already have starting values of the model parameters which are reasonably close to the values required for a best fit. Linearization then allows you to convert the exact non-linear formulation into an approximate linear formulation which is valid for small changes in the starting values. This approximate linear version of the model is then used for a <u>single</u> iteration which computes a small correction to the starting values. The original starting values are then updated by the corrections to give new starting values and the whole process is repeated (iterated). In the new iteration the linear approximation is slightly different because the starting values have changed by small amounts. When the corrections are so small that starting values are not significantly changed then you stop.

This is the mechanism of the stepwise sequence of changes which gradually improves the parameter values until you obtain your optimized answer.

Obviously with this method you need to get the ball rolling so somehow you need to know the answer approximately before you compute it! These approximate answers have various names such as *starting values, initial values, approximate values, trial values*.

Finding trial values can sometimes be very easy. In the case of instrument calibration the optimized parameter is often a deviation from a nominal design value and this deviation should be zero. For example in the laser tracker the laser beam should be parallel to the primary rotation axis (approximately vertical). Any residual beam tilt is deliberately manufactured to be very small, so the least squares calibration uses an initial value of zero for beam tilt.

Other situations are more complicated, as when trial instrument positions are needed before computing a bundle adjustment. However in all cases where **Axyz** uses least squares there are simple methods for finding initial values.

3.3 Redundancy

Least squares methods benefit from excess information, just as an average value will get closer to the true value the more times a measurement is repeated. The number of measurements above and beyond the absolute minimum necessary to compute some set of parameters is called the *redundancy* of the measurements. For example, if you need to calibrate a

scale bar, one measurement is the absolute minimum you must make. If you make 10 measurements the redundancy is 9.

To fully determine the shape and size of a triangle you need 1 side measurement for scale and 2 angles for shape, a total of 3 measured elements. If you measure 2 sides and all 3 angles you have 5 measurements which gives a redundancy of 2. Note that the type of measurement is critical since 3 angles alone appear to be sufficient but in fact they would not give a solution since they contain no scale information. The minimum information must contain both angle and distance data in this case.

It is also good practice to measure as many different elements as possible. In the triangle, if you measure 2 angles once each and then one side is measured 3 times, there are 5 measurements in total. The redundancy is then 2. However it would be more efficient, if possible, to measure each of the 3 sides once rather than the one side 3 times.

Redundancy can also be viewed as an excess of equations over unknowns. A least squares solution is often constructed so that each measurement provides one equation involving some or all of the parameters of a particular model. The parameters are the *unknowns* which must be determined and mathematical solutions demand that there be at least as many equations as unknowns.

With this approach the number of parameters or unknowns in the model is the minimum necessary to describe the situation. For example, when computing a best fitting sphere to a number of measured targets the sphere would be defined in terms of its centre (3 coordinates) and radius. The mathematical model therefore must therefore determine 4 parameters. To do this it must have a minimum of 4 equations. Each set of target coordinates generates one equation, so a minimum of 4 targets must be measured. If you measure, say, 10 targets, the redundancy is 6, i.e.

redundancy = (number of equations) - (number of unknowns)

3.4 Residuals and modelled observations

3.4.1 Residuals in brief

Residuals are the difference between a modelled value of a measurement and what you actually measure.

The least squares method finds a model which minimizes the sum of the squares of the residuals. The residuals therefore indicate how well a model of a situation fits the measurements.

A large residual may indicate a bad measurement or defective reference data, such as an incorrectly entered scale length or incorrect coordinates of a control point.

However, the least squares method does not guarantee to indicate bad measurements with large residuals. It sometimes happens that bad measurements have small residuals and good measurements are then assigned large residuals to compensate for the distortions introduced.

3.4.2 Residuals in detail

Residuals and modelled measurements, also called *observations*, are effectively correlated because they are based on a mathematical model. If one element of the model changes there must be changes in the other elements. This is a property of the modelling process.

Example of a triangle:
The 3 *modelled* angles must add up to 180 degs. If one angle is changed the others must change to preserve a consistent geometry.

Note that the 3 *measured* angles do not necessarily add up to 180 degs. This discrepancy is the reason for attempting to find a best fitting model which is internally consistent and can therefore be used to provide results. The errors in the measured angles (not the angles themselves) are assumed to be independent of one another. Only a good reason, such as systematic error in a theodolite, would invalidate this assumption.

The least squares solution can provide error estimates for the modelled observations. Since modelled observations are the best estimates of the true observations, their computed errors (in a covariance matrix) would be accepted as the best estimate of measurement quality.

If the modelling process is accurate and no errors are left unaccounted, then residuals are a good indicator of typical measurement errors. If the least squares procedure involves a reasonable number of measurements (good redundancy), the RMS residual value is a good indicator of measurement standard error. However, it tends to provide a smaller estimate of error than a rigorous analysis, i.e. it implies the measurements are better than they are.

3.5 Weights and weighted least squares solutions

3.5.1 Weights in brief

When calculating a least squares solution to a measurement problem, a measurement residual is assigned a value called a *weight* and the higher the weight the more influence the corresponding measurement will have on the final model of a situation. This means that the modelled and measured values will be much closer and the corresponding residual will therefore be small.

Weights provide two advantages:
- They can account for different measurement qualities
- They allow for processing mixed types of measurements such as distances and angles

The value of a weight is often related to a measurement's calculated or estimated quality, as defined by its standard error. If the standard error is small the weight value is high and vice versa. Calculating weights in this way is known as *weighting by standard error* (or *weighting by variance*).

Sometimes it is convenient to ignore differences between measurements and treat them all equally. In this case every measurement receives a value for weight = 1. This weighting scheme is known as *unit weighting*. You might choose unit weights in several situations:
- Relative measurement qualities are not very well known
- Measurement quality varies but only by small amounts
- For comparison with results from 3rd party software packages

Unit weighting represents the simplest form of least squares analysis and is very commonly used.

Weights in practice

Consider for example control points, which have known coordinates. They force an orientation into the coordinate system where the control points are located. During the analysis the control point reference coordinates are treated as measurements with very high weights so that they develop only very small residuals. If the weights are high enough the residuals are almost zero and the modelled control points are effectively *fixed* at the reference values. Since the modelled control points end up identical to the reference values, the rest of the model, including the instrument locations and tilts, must adjust to conform to this, i.e. they end up in the same coordinate system.

Weights provide a very flexible way to balance the influence of diverse types of measurement but this very flexibility can provide problems for unskilled users. Just as weights can be set sufficiently high to make some values effectively fixed, they can also be set so low that some measurements are effectively deleted from the solution. In this way a user might unintentionally remove a critical measurement and cause the solution to fail.

Axyz routines often provide simpler weighting schemes in which the user can treat values as *fixed* or *unknown*. Within the corresponding processing routine these assignments are actually treated as either very high or very low weights.

3.5.2 Introduction to weights

A simple least squares procedure minimizes a sum of squared residuals, i.e. an expression of the form

$$(v_1^2 + v_2^2 + v_3^2 + .. \; v_n^2)$$

where the procedure makes a total of n comparisons between model and measurement.

This implies that the measurements are equally good. Consider the task of calculating orientations in a measurement network. What happens if some pointings are made by instruments of lower precision than others or individual targets are poorly defined? Surely these pointings should not have as much influence on the final result as the higher quality pointings?

Another difficulty occurs with mixed types of measurement, for example with polar techniques or even basic triangulation. Suppose the user has created a very accurate 3m scaling length using an interferometer and wants to ensure that the analysis holds this length to within 3 microns of its interferometric value. This is a relative error of 1 micron per metre. Suppose also that the theodolite pointings are assumed good to 1 arc second. Since 1 arc second is approx. 5 microns per metre the scaling length should be given more influence in the analysis than the angular pointings.

Weights provide a way of taking into account different measurement qualities and allow for processing mixed types of measurements. A measurement residual is assigned a value called a *weight.* A high weight tends to make modelled and measured values closer and produces smaller residuals.

Weights are based on the fact that measurements which are known to have a high precision have a distribution with a small standard error, and measurements with a low precision have a distribution with a large standard error. Obviously high precision measurements should have a high weight and low precision measurements should have a low weight. It is mathematically more convenient to work with a variance rather than a standard deviation and so for a measurement with variance σ^2 the weight is defined as:

$$\text{weight} = \frac{\text{const}}{\sigma^2}$$

When the variance is small the weight is high and vice versa. This is the required mechanism.

The constant term has no natural value. It is simply a scaling factor which can be assigned any value to make the weights convenient to handle mathematically. Remember that the purpose of weights is to distinguish relative effects, for example to say:

    measurements of type A are twice as good as measurements of type B

If type A has an assigned $\text{weight}_A = 100$, then type B must have $\text{weight}_B = 50$. The weighting could equally well be $\text{weight}_A = 0.01$, $\text{weight}_B = 0.005$ or any other 2:1 combination.

In a weighted least squares analysis, the quantity to be minimized then becomes:

$$(w_1 \, v_1{}^2 + w_2 \, v_2{}^2 + w_3 \, v_3{}^2 + .. \; w_n \, v_n{}^2)$$

where w_1 etc. are the weights corresponding to the residuals.

If the measurements are all equally good and should logically have the same weight w, this then reduces to:

$$w \, (v_1{}^2 + v_2{}^2 + v_3{}^2 + .. \; v_n{}^2)$$

It is quite natural to assign the value 1 to w in this case and the result is then clearly identical to the original simple least squares concept. However if w had some other value, e.g. 100 or 0.000396, it would make no difference to the final result. The model which produces the minimum value would still be the same model. The actual value of the minimum might be bigger or smaller but it is still the smallest value which can be achieved with the chosen weight assignment.

3.5.3 The weight matrix

Every least squares analysis in *Axyz* uses matrix algebra and it is convenient to process the weights in a single *weight matrix*. In the simplest case where measurements are uncorrelated, this has the form:

$$w = k \cdot \begin{bmatrix} \dfrac{1}{(\sigma_1)^2} & 0 & .. & .. \\[2ex] 0 & \dfrac{1}{(\sigma_2)^2} & .. & .. \\[2ex] .. & .. & .. & .. \\ .. & .. & .. & \dfrac{1}{(\sigma_N)^2} \end{bmatrix}$$

where k is the constant scaling factor.

The matrix on the right-hand side is the inverse of the variance matrix of the measurements, and so the weight matrix can be written:

$$
w = k \cdot \begin{bmatrix} (\sigma_1)^2 & 0 & .. & .. \\ 0 & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}^{-1}
$$

When a full mathematical analysis is done it can be shown that the weight matrix is best defined as the inverse of the full covariance matrix when the measurements are correlated, so that the most general weight matrix is defined as:

$$
w = k \cdot \begin{bmatrix} (\sigma_1)^2 & \sigma_{12} & .. & .. \\ \sigma_{21} & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}^{-1}
$$

This form of weight matrix occurs often, for example when fitting shapes to measured points which normally have correlated coordinates. However users have some flexibility to choose a weighting scheme in these cases.

3.5.4 Selecting a weighting scheme

The actual value of the weights used in any *Axyz* procedure depends on the choice of scaling factor which is arbitrary. There are different ways to choose the scaling factor.

Unit scaling factor: Weight = inverse of variance.

In this case the scaling factor $k = 1$ and it has no dimensions.

For simple, uncorrelated measurements

$$
\text{weight} = \frac{1}{\sigma^2}
$$

This creates large numbers for weights. For example, if the weight is applied to angles which are assumed to have a standard error of 1 arc second and all calculations are done in radians, then:

$\sigma = 1" = 0.000004848$ rad   weight $= 42550 \times 1\,000\,000$ rad^{-2}

In practice this approach would define weights using the inverse of the full covariance matrix to generate a corresponding weight matrix, i.e.

$$w = \begin{bmatrix} (\sigma_1)^2 & \sigma_{12} & .. & .. \\ \sigma_{21} & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}^{-1}$$

Here the scaling factor is still dimensionless with a value 1, even if mixed types of measurements are involved.

Simple unit weighting

In the very simplest case, all measurements are uncorrelated, of the same type and equally good. The user need not even consider scaling factors and variances but simply makes the statement:

weight $= 1$

Here the <u>weight</u> is a dimensionless number of value 1, in contrast to the previous method where the <u>scale factor</u> was a dimensionless number of value 1. With unit weighting the scale factor has units of variance. Suppose the measurements in this case all have a standard error of σ_0 then:

$$\text{weight} = 1 = \frac{(\sigma_0)^2}{(\sigma_0)^2}$$

i.e. the scale factor has effectively been assigned the value $\sigma_0{}^2$ even though this may not have been a conscious choice by the user.

This scheme generates a diagonal unit weight matrix.

Approximate unit weighting

If there is some variation in measurement quality, a unit weighting scheme must be slightly modified to become:

$$\text{weight} = \frac{(\sigma_0)^2}{\sigma^2}$$

Now the user will have to consider some actual value for the scale factor σ_0^2 which will generate weights of around 1 when combined with the standard errors of the measurements.

For measurements of the same type the scale factor therefore has the same variance units as the measurements themselves, e.g. mm^2, rad^2. Its actual value is similar to a typical variance value within the measurements. (Mikhail[1] p77 has a good example of this.)

This method remains valid, but loses its simple concept, when measurements of mixed types are involved. The weight cannot then be a dimensionless number for every type of measurement. (Mikhail[1] demonstrates this with correlated measurements in his example 5-12.)

Weighting scheme used in Axyz

Axyz orientation methods use the inverse of the covariance matrix for weighting purposes. This is known as *weighting by variance* but within **Axyz** it is more loosely called *weighting by standard deviation.*

Axyz shape fitting routines permit the user to choose between this method and simple unit weighting. Shape fitting packages used by other measurement systems often employ unit weighting because these systems generate a fairly uniform error quality. **Axyz** users therefore have the option to use the same weighting method in order to get similar results for purposes of comparison.

3.5.5 Testing the weighting scheme after processing

When processing is complete the residuals can be examined to see if they correspond to the assumed quality of measurement. The scale factor used for weighting is also a test statistic known amongst other names as the *variance factor*. Using the residuals, it can be estimated purely analytically and if it does not agree with the original value assumed for weighting purposes then one of several problems is indicated. For example, there may

be one or two bad measurements, which should therefore have a low weighting, or a whole group of measurements has been weighted incorrectly relative to the others.

This is discussed in more detail later, see "Variance factor" on page 38.

3.5.6 Further use of the weighting factor

The weighting factor is generally known as the *variance factor* and is a component in *error propagation*. This technique computes errors in parameters due to errors in the measurements from which they were derived. For example the tolerance in the radius of a circle can be computed when the errors in the target points lying on the circle are known. Error propagation is a by-product of the least squares solution and since this involves weights based on the variance factor, the variance factor must be known in order to correctly scale the error estimates. See "Error propagation and the variance factor" on page 47.

4. Quality figures

4.1 Introduction

The section on simple statistical ideas outlines the sources of measurement errors and tolerances and introduces some commonly used statistical parameters such as mean values, standard errors and variances.

Once a measurement has been made, users will want to know how good it is, since quality is a major consideration in any manufacturing or analysis task. There is usually a requirement that the measurements meet or improve upon some quality goals.

Depending on the task, the processing routines in *Axyz* generate different types of statistics or quality figures. These provide information about the measurements which have been processed and the parameters such as target locations and dimensions of shapes which are the objective of the processing.

The routines are more complex than required for, say, the simple repeated measurement of the distance between two points. Some of the figures discussed in this section are also a little more complex than the simple figures already mentioned, but they are developed on the same statistical basis.

Note
The optimized processing routines in *Axyz* all use the method of least squares which mathematically generates the best estimates of the parameters they are designed to calculate.

4.2 Preliminary and calculated quality estimates

Least squares procedures often require estimates of measurement quality *before* data is processed. These are called *a priori* estimates which is a Latin term used to imply "before processing". These estimates are therefore *preliminary* values made manually by the user but are likely to be based on previous experience and should represent reliable values.

After processing is complete it may be possible to supply good *a posteriori* quality estimates. This is another Latin term used to imply "after processing". These estimates will be *calculated* by the processing program

and depend on the quality of the actual data supplied. Actual quality may be different from the user's a priori assumption!

For convenience, any further discussion will refer to preliminary and calculated estimates rather than a priori and a posteriori estimates. However the Latin terms are used in text books and reference sources.

It is important to realize that good statistical estimates can only be computed if a sensible number of measurements are made, i.e. if there are more measurements than strictly needed to compute a result. This is known as *redundancy*.

Example:
1. Toss a coin 10 times and you get maybe 8 heads.
2. Toss a coin 100 times and you get maybe 55 heads.
3. Toss a coin 1000 times and you get approximately 50% heads.

From the first test you might assume heads are much more likely than tails, but this would be wrong, as the improved tests show. It may not always be possible in a particular measurement situation to get a good estimate of measurement accuracy from the measurements themselves. For example, only very limited statistical information about a target's position and associated pointings is available from a dual theodolite intersection. In such cases it is better to use quality estimates based on experience and which are ultimately derived from other more valid sets of measurements.

The quality figures which are of primary interest are error estimates for computed target coordinates and elements derived from them. However, since the original source of coordinates are the measured angles and distances, quality estimates are also required for these.

To obtain preliminary error estimates for angular pointings and distance measurements users have the option of:
- Evaluating an earlier least squares analysis such as the results from a bundle adjustment
- Making multiple independent test measurements such as repeated pointings from a stable instrument position to a well defined fixed target

4.3 Summary of quality statistics provided by Axyz

4.3.1 RMS error

The RMS error has the following features:
- It applies to residuals of a selected type and not necessarily all the residuals in a particular least squares procedure
- It has the same units as the quantity evaluated, e.g. arc secs, mm
- It can act as an estimate of quality of the selected measurement type and is roughly equivalent to a standard error
- If larger than expected it may indicate the presence of bad measurements
- It requires redundant measurements (more than theoretically necessary) in order to be meaningful.

In **Axyz** shape fitting routines the RMS error is associated with unit weighting of the measurements. It is not a mathematical requirement to link RMS and unit weights but it is a convenient approach in the design of the routines. See "Quality results from unit weighting" on page 45.

4.3.2 Calculated variance factor (ManCAT "mean error")

The variance factor calculated after processing has the following features:
- It is statistic which involves all the residuals, possibly of different types, in a least squares analysis.
- In **Axyz** routines the factor is a dimensionless number which should be close to 1.
- If the factor is large this can indicate problems in the relative weighting of measurements. Amongst other possibilities, incorrect weighting may imply a bad measurement, .
- It requires redundant measurements (more than theoretically necessary) or the factor cannot be computed.

The variance factor is mainly useful for indicating weighting problems. Weights are assigned before processing with an preliminary variance factor of 1. Problems are detected after processing when the variance factor can be estimated again, using the actual results. This statistic always has an element of "before and after" comparison.

In **Axyz** shape fitting routines the variance factor is associated with weighting of the measurements according to their covariance matrices.

It is not a mathematical requirement to make this link and the factor can also be computed for measurements of unit weight. The link is made for convenience in program design. In fact, the variance factor is indeed computed for shape fits using unit weighting since it is required for error propagation but it is only used internally in this case. See "Error propagation with unit weighting" on page 52.

4.3.3 Variance/covariance estimates from error propagation

Variance/covariance estimates have the following features:
- Based on quality figures for the measurements, they provide estimates for standard errors and variances of specific calculated parameters, e.g. coordinates of a theodolite's position in a measurement network.
- In *Axyz* routines the calculated error estimates are ultimately derived from the user's preliminary estimates of measurement quality. It is up to the user to decide if estimates are consistent with the results returned.

4.4 Root Mean Square (RMS) error

4.4.1 RMS in brief

The Root Mean Square (RMS) error is derived from a set of measurement residuals produced by a least squares analysis such as a bundle adjustment or shape fit. It relates to a particular type of measurement quantity such as angles, distance measurements or coordinate offsets.

The RMS value provides a single quality figure in the units of the measurement concerned. It is an estimate of the spread of the measurements and/or an estimate of the closeness of a fit.

An RMS value larger than expected may indicate the presence of a small number of bad measurements. In this case the residuals may have to be examined, and other tests made, in order to track down the problem.

4.4.2 Simple definition of RMS

For a set of N residuals v_i the RMS value is defined as:

$$RMS_{resid} = \sqrt{\frac{\sum_{i=1}^{N} (v_i)^2}{N}}$$

4.4.3 Full definition of RMS

The RMS value can be applied to any set of N related measurements to generate a single value which is representative of the full set of measurements. It is defined as:

$$RMS_{resid} = \sqrt{\frac{\sum_{i=1}^{N}(X_i)^2}{N}} \qquad or \qquad RMS_{resid} = \sqrt{\frac{(X_1)^2 + (X_2)^2 + .. + (X_N)^2}{N}}$$

When estimating measurement quality the RMS is often applied to the deviations of measurements from their mean value, rather than to the measurements themselves. Since deviations are sometimes positive and sometimes negative, the squared terms enable a single quality figure to be calculated which represents the spread of the measurements, i.e.

$$RMS_{dev} = \sqrt{\frac{\sum_{i=1}^{N}(X_i - X_{mean})^2}{N}}$$

This is fine for evaluating the same physical quantity which has been measured a number of times, such as a distance between two particular targets but the statistic is most useful when applied to measurement residuals resulting from some least squares processing method. The definition is then:

$$v_i = X_i - Xe_i \qquad RMS_{resid} = \sqrt{\frac{\sum_{i=1}^{N}(v_i)^2}{N}}$$

Here Xe is the least squares estimate of the corresponding measured value X, and v is the residual.

To be meaningful the residuals must all be of the same type. However they relate to different physical elements such as the various offsets of a set of targets from a fitted surface, rather than repeat measurements of the same quantity. The following two examples of a circle fit and target intersection demonstrate this.

In these examples, the minimized least squares sum is the same one used to create the RMS value, i.e. the RMS is minimized. In more complex least squares solutions, such as a bundle adjustment involving angle, distance and control point measurements, the minimized least squares parameter involves mixed types of residuals and weighting factors.

Example: RMS residual for circle fit

A circle is fitted to a set of measured points on a plane. The analysis finds the RMS of the perpendicular offsets (d_1 .. d_5) from the circle.
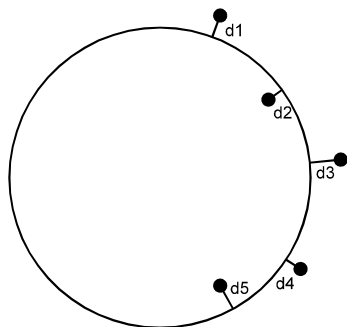
FIT-SH.WMF

Example: RMS residual for target intersection

A target is intersected from three theodolite positions. The method of analysis finds the RMS of the 3 perpendicular offsets (d_1 .. d_3)from the target position to each line of sight
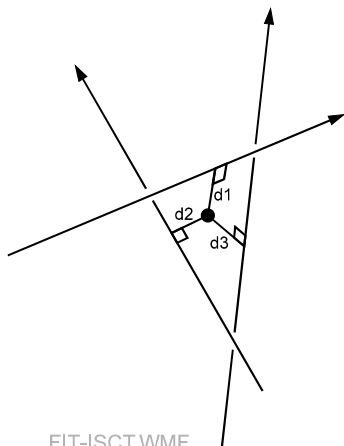
.   FIT-ISCT.WMF

4.4.4 RMS for zero redundancy

When there is no redundancy, residual values are zero and the RMS value is also zero.

4.4.5 Practical use of RMS

If the RMS is derived from a reasonably large number of residuals (e.g. >10) it is a reasonable estimate of the standard deviation (σ) which in turn indicates a typical measurement error.

An RMS which is larger than expected may indicate one or more bad measurements and the individual residuals should be reviewed. Measurements with large residuals are often in error. They can be removed from the least squares procedure provided there are still sufficient measurements left to generate a sensible result!

Alternatively if the large RMS results from a fit to design coordinates or a shape fit, the design data or assumption may be in error. For example, one of the design coordinates might be faulty or the points might not lie on a well defined shape.

If the RMS value is derived from a small number of residuals it is best regarded as indicating the closeness of fit rather than an estimate of measurement quality.

Example RMS: Bundle adjustment

Bundle adjustment with 50 targets and 2 theodolites (resolution of 1 arc second or better).

RMS angle residual = 1.2 arc secs
Good result. This figure should be a reasonable estimate of the standard deviation of a pointing.

RMS angle residual = 4 arc secs
Worse than expected. Check the individual residuals to see if there are some obviously bad results.

Example RMS: Intersection (single point solution)

3-ray intersection from a baseline of 2m to a target 5m away using high resolution theodolites.

RMS intersection error = 50μ.
This is the RMS value of the offsets of the target from the pointings. Good result.

RMS intersection error = 250μ
Bad result.
Check the pointings to see if one is in error.

Best-fit coordinates

Fit of a set of measured target coordinates to a design set.

RMS coordinate offset =65μ
This is the RMS value of the lengths of the small space vectors between the design coordinates and the transformed target coordinates.
Good result.

RMS coordinate offset =315μ
Bad result.
Check if there are a small number of bad individual residuals. Either the measurements are bad or the design data is faulty at those points. If all residuals are bad, check if the correct set of design coordinates has been used.

Note
Residuals which are generally poor may indicate that the reference information is of lower quality than the measurements, rather than the other way around. A similar case can occur with shape fits. If points give generally poor results when fitted, say, to a cylinder, then perhaps the object does not define a very good cylinder. In this case the measurements could be used to map the deviation of the object from its nominal design shape.

4.4.6 RMS error (ECDS)

Intersection

The ECDS intersection method uses theodolite parameters derived from the ECDS bundle adjustment and finds the target position which minimizes the sum of the perpendicular offsets from the line of sight. The theodolite parameters are effectively regarded as fixed.

The RMS value of these offsets is provided.

Bundle adjustment

The ECDS bundle adjustment does not directly use theodolite pointings but converts them into "pseudo-photographs" with a nominal principal distance of 100mm. The least squares solution therefore processes photo observations in mm rather than angles.

The adjustment computes an RMS value for the photo residuals in mm. An RMS value for scale bar distances is not quoted.

4.4.7 RMS error (ManCAT)

The RMS error is not provided but the "mean error" displayed instead.

4.5 Mean error (ManCAT system)

4.5.1 Background to Mean error

The mean error is a name employed by the ManCAT system as an alternative to *variance factor*. For a full discussion see "Variance factor" on page 38.

Use of the term "mean error" gives rise to terminology problems which will be avoided in the *Axyz* system. The mean error is generally known in English textbooks under one of several other names:
- reference variance
- variance of a measurement of unit weight
- unit variance
- variance factor

The statistic can be defined in slightly different ways, so there is some justification for more than one name. However *Axyz* has chosen to adopt the term "variance factor" as being most appropriate to the type of definition in use. This avoids adding yet another name to the list and avoids other potential sources of confusion.

The term "mean error" was originally selected as more user-friendly than "reference variance", since many ManCAT users are not familiar with this specialist mathematical term.

However, German textbooks use a statistical figure called the "mittlerer Fehler". This translates as "mean error" and is the <u>square root</u> of the "mean error" defined in ManCAT, i.e. it is the square root of the variance factor.

Also the terms "mean" and "average" are interchangeable in English and a definition of "average error" exists in English literature. The "average error" is a quantity typical of the physical error values which can occur and in a triangulation system would have dimensions of angle or distance depending on the type of measurement.

In agreement with the definition of variance factor in *Axyz*, ManCAT's "mean error" is a dimensionless quality figure which should have the value 1 if preliminary estimates of measurement quality are close to their true values. It is not an error in the sense of having units such as mm, arc secs, etc.

4.5.2 Mean error in brief

The mean error is a dimensionless quality figure derived from <u>all</u> the measurement residuals produced by a weighted least squares analysis such as a bundle adjustment. It does <u>not</u> have units of measurement such as mm or arc secs.

The mean error is useful as an indicator of a weighting problem such as the presence of a bad measurement which should be given a low weight or eliminated from the processing.

In a well structured solution with good data of known quality the mean error should have the value 1. If this is not the case, the data may be affected by one of the following problems:

- One or two bad measurements with large residuals which should have a lower weight
- Unbalanced weights, for example control points which are given too much weight relative to theodolite pointings
- An error in a true element shape, for example measured points lie on an ellipse but you try to fit them to a circle
- A simple internal scaling error which does not affect the values of calculated parameters but which may cause incorrect error estimation of those parameters

Other quality figures may have to be investigated in order to identify the specific problem

4.6 Variance factor (calculated)

4.6.1 Variance factor in brief

In **Axyz** the calculated variance factor is a dimensionless scaling factor for weights in a least squares calculation. It involves <u>all</u> the measurement residuals in a particular analysis. If the weighting scheme is correct and there is good redundancy, it should theoretically have the value 1. If not, it may indicate an error in the weighting scheme due to one of several sources:

- One or two bad measurements with large residuals which should have a lower weight
- Unbalanced weights, for example control points which are given too much weight relative to theodolite pointings
- An error in a true element shape, for example measured points lie on an ellipse but you try to fit them to a circle
- A simple internal scaling error which does not affect the values of calculated parameters but which may cause incorrect error estimation of those parameters

The variance factor can only warn of a problem in a least squares solution but may not uniquely identify it. Other quality figures may have to be investigated, or individual residuals examined, in order to track down a problem.

The variance factor will only provide significant information if the measurement set has high redundancy. If the redundancy is low, a value different from 1 may not mean very much. Values between 0.3 and 1.8 would normally be considered acceptable.

4.6.2 Introduction to variance factor

The discussion on weights showed that a scale factor is required in order to generate suitable values for the weights of measurements used in a least squares analysis. (See "Weights and weighted least squares solutions" on page 20.) This scale factor is an arbitrary number defined before the least squares calculation is made. The scale factor is generally known as the *variance factor* and this is its preliminary value.

It can also be shown that the variance factor can be estimated in a different way after processing is complete. This value depends on both the chosen

weights and the actual residuals which result from the least squares analysis. The value is defined as

$$var\_factor = \frac{\mathbf{v^T} \cdot \mathbf{w} \cdot \mathbf{v}}{r}$$

(correlated measurements)

This is the calculated value of the variance factor and is the one reported by **Axyz** routines.

The weights involve an assumption about the quality of the measurements, based on an assumed variance. If this is correct the residuals should be consistent with the assumption, i.e. for measurements of a particular quality the spread of their residuals should roughly correspond to the assumed variance.

The calculated variance factor has two main uses. It checks the consistency of the weighting scheme and, because it involves the residuals, reflects the presence of any bad measurements. Bad measurements point to another weighting problem since they are effectively low quality measurements which have been initially assigned a weight which is too high.

Since an arbitrary (preliminary) variance factor is set before processing, problems are detected by comparing this with its value after processing (calculated). They should be roughly the same, i.e. their ratio should be approximately 1. When reporting this statistic, **Axyz** routines set the preliminary value to a dimensionless value of 1 so that the calculated value should also be 1.

Alternative ways to define the variance factor

It is a convenience to have the calculated value of the variance factor = 1, but there are other ways of handling the numbers depending on the values created for weights. This also gives rise to alternative names for the variance factor.

In the discussion on weights, a measurement of assumed variance σ^2 is given a weight defined as:

weight $= k / \sigma^2$ where k is a constant scale factor

This gives a high weight to measurements with a low spread and vice versa. The terms high and low are purely relative and have no real meaning if all measurements in a particular job are equally good or equally bad. In either of these cases it would be valid to assign the same weight to all the measurements. Purely for convenience, weights could all be assigned the dimensionless value 1 for each job.

Job A   Job B
weight $= 1 = k / \sigma_a^2$ weight $= 1 = k / \sigma_b^2$
$k = \sigma_a^2$ $k = \sigma_b^2$

The value of the constant then depends on the measurements used in the job. In this case the constant term is itself a variance with units of variance such as mm^2 or rad^2. In this weighting scheme, where the <u>weight</u> has the dimensionless value of 1, it is not surprising that the constant term is often called the *reference variance* or the *variance of a measurement of unit weight*.

Most cases are more complex than this and there is a need to use different weights, for example when mixed measurements are employed. A polar measuring system uses both angular and distance measurements. These have different units as well as different standard errors. Even if all the angles are equally good and all the distances equally good, angles and distances need to be weighted differently. When different measuring units are processed the reference variance can itself only be expressed in terms of one of these and the weights will end up having mixed dimensions (some of which may be dimensionless).

Since weighting simply ensures that the relative influence of different measurements is appropriate to their respective quality, the numbers can be set in an alternative way. Given that

$$\text{weight} = k / \sigma^2$$

the value 1 can be assigned to the scale factor k rather than the weight, i.e.

set $k = 1$ therefore weight $= 1 / \sigma^2$

Now the weight of a measurement is directly related to the dimensions of the measurement itself and is different for every measurement type. Its dimensions are inverse variance, such as mm^{-2}, rad^{-2}. With this approach the term *variance factor* is more appropriate. This is the approach used in

Axyz when reporting variance factors, but see also the comments in "Quality results from unit weighting" on page 45.

Finally, here are the different names for variance factor which may be used in other literature:
- "reference variance" also called "variance of a measurement of unit weight"
- "variance factor" also called "unit variance"

"reference variance" is a particularly common term.

4.6.3 Simple definition of variance factor (calculated)

The variance factor for uncorrelated measurements of the same quality, and which are weighted by variance, is defined as:

$$\text{var\_factor} = \frac{\sum_{i=1}^{N} \frac{(v_i)^2}{\sigma^2}}{r} \qquad\qquad r = N - u$$

N the total number of measurements made
v_i indicates the residual of measurement number i
σ is the preliminary estimation of the standard error of the measurements
r is the redundancy in the measurements
u is the minimum number of measurements to enable a solution.

If the residuals are typically of the same magnitude as the preliminary estimate of standard error, then this value tends to average to 1 for a large number of measurements.

To allow for the general case of correlated measurements of different quality, the following definition is made.

4.6.4 Full definition of variance factor

$$\text{var\_factor} = \frac{\mathbf{v}^\mathbf{T} \cdot \mathbf{w} \cdot \mathbf{v}}{r}$$

where

$$\mathbf{v}^\mathbf{T} = \begin{pmatrix} v_1 & v_2 & .. & v_N \end{pmatrix}$$

$$\mathbf{w} = \begin{bmatrix} (\sigma_1)^2 & \sigma_{12} & .. & .. \\ \sigma_{21} & (\sigma_2)^2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & (\sigma_N)^2 \end{bmatrix}^{-1}$$

(Weighting by variance.)

If there are no correlations then the covariances are zero and the weight matrix, w, simplifies to:

$$\mathbf{w} = \begin{bmatrix} \dfrac{1}{(\sigma_1)^2} & 0 & .. & .. \\ 0 & \dfrac{1}{(\sigma_2)^2} & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & \dfrac{1}{(\sigma_N)^2} \end{bmatrix}$$

If the individual standard errors are the same value σ, the variance factor then reduces to the same form as given in the simple definition.

4.6.5 Definition of variance factor in Axyz bundle adjustment

$$\text{var\_factor} = \frac{\mathbf{v}^\mathbf{T} \cdot \mathbf{w} \cdot \mathbf{v} + \mathbf{v'}^\mathbf{T} \cdot \mathbf{w'} \cdot \mathbf{v'} + \mathbf{v''}^\mathbf{T} \cdot \mathbf{w''} \cdot \mathbf{v''}}{r}$$

This definition is necessary because the bundle adjustment splits up matrices in order to permit a more efficient solution, but it is effectively the same definition as the full definition.

4.6.6 Variance factor for zero redundancy

When there is no redundancy all residuals are zero and r = 0. In this case the variance factor would be an indeterminate value 0/0. In practice **Axyz** routines forces the redundancy to have its actual value or 1, whichever is greater. As a consequence the variance factor would be given the value zero when redundancy is zero.

4.6.7 Practical use of variance factor

The variance factor is a convenient single parameter for expressing the quality of a least squares result. Many ManCAT users have standard measurement procedures where, after some initial experience, they routinely generate relatively small values for the factor but do not make special efforts to force it to the value 1. However when something goes wrong the value will usually become very large and this is used as a warning signal to check for errors.

There are many reasons why the variance factor has some value significantly different from 1. For example this may indicate that the preliminary estimates of measurement quality need some fine tuning, either because of some general imbalances in the weighting or because there are a few bad measurements.

In the comments which follow it may help to have a simplified and expanded definition of the calculated variance factor to see the reasons for a value different from 1. For uncorrelated measurements of different measurement quality:

$$\text{var\_factor} = \frac{1}{r} \cdot \left[\left(\frac{v_1}{\sigma_1} \right)^2 + \left(\frac{v_2}{\sigma_2} \right)^2 + .. + \left(\frac{v_N}{\sigma_N} \right)^2 \right]$$

where r is the redundancy in the measurements.

Low redundancy

The redundancy itself is important. If it is not a reasonably large value, which means the number of measurements N must be reasonably large, then the value of the variance factor will not be reliable.

For example, if there are not many terms within the brackets (N is low) the following situations can occur:

- A few residuals just happen to be at the tolerance limits (3σ). This makes the factor greater than 1.
- The few measurements happen to be consistent and therefore generate small residuals. This makes the factor less than 1.

In both cases there would be nothing actually wrong with the result. The further comments assume that there is reasonable redundancy.

Bad measurements

A small number of bad measurements in the set are likely to produce some large residuals, well outside the tolerance limit. This will increase the factor. Individual residuals need to be checked in this case.

Bad design data or assumption

Large residuals and a large variance factor are also generated when the measurements are perfectly good but some reference information is at fault. This occurs when you transform points onto design values but one of the reference coordinates is in error, or you try to fit points to a circle when they actually lie on an ellipse. Here the design data should be checked.

Unbalanced weights

Alternatively, if a sub-group of measurements have assumed standard errors which are either too low or too high, this will also affect the factor. Suppose the standard error is too low, i.e. these measurements are generally not as good as expected. Then typical residuals will generally be higher so that the terms v/σ will tend to be greater than 1 and will cause the factor to rise. Equally well, if the actual measurement quality is better than expected, this would tend to make the factor less than 1.

This can happen when mixing control coordinates with instrument measurements in a bundle adjustment. It may be difficult to estimate different variances for very different types of data and an imbalance results in favour of the control. Before making any changes to the assumed standard errors of the control it may be sensible to repeat the adjustment treating the control points as unknown. In a second step the calculated coordinates for the control points can be transformed onto their design values to see if there are one or two badly fitting points.

Generally high or low estimated errors

If measurements are all of the same type and their estimated standard error is too low or high, then the variance factor will correspondingly be greater or less than one. Re-scaling their standard errors to bring the calculated reference variance close to unity will not actually alter the results very much. This is because the relative weights stay the same and it is only relative weights which affect the calculation of the target coordinates or the circle radius or whatever the objective of the analysis is.

However the solution has told the user that the measurement quality, on the basis of the statistics present in the actual processed measurements, is different than the user originally thought. Unless changed, error estimates of the calculated parameters by *error propagation* will not be reliable. See "Linking the error type and weighting type" on page 47.

4.7 Quality results from unit weighting

When measurements are of the same quality and not correlated it is convenient to assign them weight values of 1. In a simple least squares analysis this is very common. In fact, the issue of a weight never explicitly appears in this case. The method simply minimizes:

$$(v_1^2 + v_2^2 + .. + v_n^2)$$

It is then quite likely that the RMS of the residuals would be produced as an approximation to the standard error of the measurements, i.e.

$$\text{RMS}_{\text{resid}} = \sigma_m = \sqrt{\frac{(v_1)^2 + (v_2)^2 + .. + (v_N)^2}{N}}$$

Viewed in terms of weights, measurements of standard error σ_m are here assigned weights according to the definition:

$$\text{weight} = k / \sigma_m^2$$

Since the weight is 1, k must be given the value σ_m^2, i.e.

$$\text{weight} = 1 = \sigma_m^2 / \sigma_m^2$$

k is the variance factor, although as explained earlier it would be better to call it the reference variance in this case since it is equivalent to the variance of the measurements.

It is not necessary to know the estimated measurement variance and variance factor before calculation, since the weight has been directly assigned. That is all that is needed to get the least squares answer. However, a subsequent calculation of the variance factor could obviously be used to provide the estimate of measurement variance and hence an estimated standard error.

By definition:

$$\text{var\_factor} = \frac{\mathbf{v}^{\mathbf{T}} \cdot \mathbf{w} \cdot \mathbf{v}}{r}$$

For uncorrelated measurements of unit weight this reduces in the current example to:

$$\text{var\_factor} = \left(\sigma_m\right)^2 = \frac{\left(v_1\right)^2 + \left(v_2\right)^2 + .. + \left(v_N\right)^2}{r}$$

The estimate of the standard error of the measurement is then:

$$\sigma_m = \sqrt{\frac{\left(v_1\right)^2 + \left(v_2\right)^2 + .. + \left(v_N\right)^2}{r}}$$

This is very nearly the RMS_{resid} value, but using the redundancy r instead of the number of observations N. Note that $r < N$ and the RMS_{resid} would be smaller, i.e. imply that the measurements were slightly better. However once the number of measurements becomes reasonably large this difference is not so important.

The method of unit weighting is offered in *Axyz* shape fitting routines. This is useful for comparison with the results from similar routines in other software packages which often use unit weights.

See also "Error propagation with unit weighting" on page 52.

4.8 Linking the error type and weighting type

When unit weighting is employed in shape fitting routines and transformations, only the RMS error is produced. The calculated variance factor could be offered but its value in this case would be a variance. In other cases using the alternative weighting scheme its theoretical value is 1, and it is better to present results in a consistent way.

In contrast, if standard errors are used for weighting purposes in shape fitting and transformation routines, then only the variance factor is offered as a statistic. In this case its value should be one.

See also "Error propagation with unit weighting" on page 52.

4.9 Error propagation and the variance factor

4.9.1 Error propagation in brief

Error propagation provides variance and covariance values for quantities which are derived from some set of measurements. It does this by evaluating the errors in these quantities due to the errors in the measurements.

Error propagation answers questions such as:

"If my angles are good to 0.7" and distances good to 2 microns, how good are the measured point coordinates?"

"If my point coordinates have standard errors of 50 microns, how good is the radius of the fitted circle?"

Propagation here means to carry through the error effects from start to finish. In *Axyz* elements build on other elements, and most originate at the instruments, i.e. angles and distances produce points which can create shapes which in turn can create further shapes and points. If you start off with an accurate knowledge of measurement quality for angles and distances it should be possible to generate good error estimates for point coordinates and shape parameters.

Very conveniently, error propagation is a by-product of least squares solutions. Since these involve weights based on the variance factor, the variance factor must be known in order to correctly scale the error estimates

For error propagation, ***Axyz*** always uses the preliminary variance factor = 1, except for shape fits where an RMS statistic has been requested. In these cases the calculated variance factor is used. This will only give meaningful values for the error estimates of shape parameters if reasonable redundancy is provided. You will not get a good value of the error in a fitted circle's radius if you only use 4 points to create the circle.

The ***Axyz*** routines therefore effectively rely on the user to provide good quality error estimates.

4.9.2 The scaling mechanism at work

In a simple least squares analysis, the following matrix equation is developed:

$$\mathbf{A}\,\mathbf{x} = \mathbf{l} + \mathbf{v}$$

where
- The vector $\mathbf{x}$ represents the unknown parameters
- The coefficient matrix $\mathbf{A}$ depends on the actual equations describing the model
- $\mathbf{l}$ is a known vector derived from the known measurements
- $\mathbf{v}$ is the vector of measurement residuals

A weight matrix $\mathbf{W}$ is associated with the residuals and the solution is given by:

$$[\mathbf{A}^T\,\mathbf{W}\,\mathbf{A}]\,\mathbf{x} = [\mathbf{A}^T\,\mathbf{W}]\,\mathbf{l}$$

which can be written as:
$$\mathbf{N}\,\mathbf{x} = \mathbf{t} \qquad \text{where } \mathbf{N} = [\mathbf{A}^T\,\mathbf{W}\,\mathbf{A}] \text{ and } \mathbf{t} = [\mathbf{A}^T\,\mathbf{W}]\,\mathbf{l} \tag{i}$$

The solution is:
$$\mathbf{x} = \mathbf{N}^{-1}\,\mathbf{t} \tag{ii}$$

To obtain statistical information about the unknowns, $\mathbf{x}$, two matrices are involved. $\mathbf{\Sigma_{xx}}$ is the required covariance matrix and $\mathbf{Q_{xx}}$, a scaled version of this, is the *co-factor matrix* (notation used by Mikhail). It is also necessary to make use of either the preliminary variance factor σ_0^2 or the calculated variance factor $(\hat{\sigma}_0)^2$ of the residuals.

Mikhail[1] proves
$$\mathbf{Q_{xx}} = \mathbf{N^{-1}}$$

By definition $\mathbf{Q_{xx}}$ is simply a version of $\mathbf{\Sigma_{xx}}$ scaled by the variance factor, i.e.
$$\mathbf{\Sigma_{xx}} = \sigma_0^2\, \mathbf{Q_{xx}} \qquad \text{or} \qquad \mathbf{\Sigma_{xx}} = (\hat{\sigma}_0)^2\, \mathbf{Q_{xx}}$$

therefore
$$\mathbf{\Sigma_{xx}} = \sigma_0^2\, \mathbf{N^{-1}} \qquad \text{or} \qquad \mathbf{\Sigma_{xx}} = \hat{\sigma}_0^{\,2}\, \mathbf{N^{-1}} \tag{iii}$$

This indicates how the least squares solution for the parameters also automatically provides the statistical information about the parameters via the matrix $\mathbf{N^{-1}}$.

Either expression should give the same result, since both versions of the variance factor should be the same.

Re-scale the variance factor with fixed measurement errors

The actual value of the variance factor is decided by the user's choice for σ_0^2 but this does not really affect results and in particular does not affect the values of the parameters in (ii) or the values in the covariance matrix in (iii). This is because a change in the value of σ_0^2 affects the weight matrix $\mathbf{W}$ and hence both $\mathbf{N}$ and $\mathbf{t}$ as follows.

Suppose σ_0^2 is doubled without changing estimated measurement errors, because the user thinks that makes better numbers for the weights. The following effects are then automatic:
* All the elements of $\mathbf{W}$ are doubled
* $\mathbf{N}$ and $\mathbf{t}$ are doubled, as indicated by (i).
* The inverse of $\mathbf{N}$, i.e. $\mathbf{N^{-1}}$, is <u>halved</u> (matrices often work just like ordinary numbers)

It is then obvious from (ii) that $\mathbf{x}$ remains unchanged and from (iii) that $\mathbf{\Sigma_{xx}}$ remains unchanged.

Re-scale the measurement errors with fixed variance factor

In contrast to re-scaling the variance factor, re-scaling the standard error estimates of the measurement residuals with a fixed variance factor has a slightly different effect.

Suppose the measurement residuals are uncorrelated and their estimated standard errors are <u>reduced</u> by a factor of 2.

- Each element of **W** then <u>increases</u> by a factor of 4 since the weight is the inverse of the square of the standard error.
- **N** and **t** <u>increase</u> by a factor of 4
- The inverse of **N** <u>reduces</u> by a factor of 4

Again it is obvious from (ii) that **x** remains unchanged. However equation (iii) shows that, using the preliminary variance factor, $\mathbf{\Sigma_{xx}}$ is now <u>reduced</u> by a factor of 4 since the inverse of **N** reduces by this factor but the variance factor stays the same.

Re-scale the variance factor but keep unit weighting

If the variance factor is re-scaled but a unit weight matrix is still used then, in effect, the estimated measurement errors are re-scaled by the same amount as the variance factor in order to keep the weights = 1.

A change of variance factor has no effect on the parameter values or their error estimates.

A change of estimated measurement error has no effect on parameter values but affects their error estimates. This is the net result.

4.9.3 Error propagation with preliminary or calculated variance factor?

Is there then any reason to choose either σ_0^2 or $\hat{\sigma}_0^2$ when calculating the variances and covariances of the computed parameters? With the exception of unit weighting for shape fits, ***Axyz*** always uses $\sigma_0^2 = 1$ in all cases of error propagation. This effectively means that the user's initial estimates of measurement quality are always carried through the analysis.

Several cases can be considered. Remember that $\hat{\sigma}_0^2$ is calculated as $\hat{\sigma}_0^2 = (\mathbf{v^T W v}) / r$, where r is the redundancy.

No redundancy

If there is no redundancy $\hat{\sigma}_0^2$ simply cannot be calculated. The least squares analysis will still give a perfectly correct solution for the parameters but residuals are zero, redundancy is zero and $\hat{\sigma}_0^2 = 0/0$ which is an undefined mathematical quantity. In this case the user's preliminary

estimation of measurement quality and relative weights is the best information available and so the only choice is to use $(\sigma_0)^2$.

Simulations

Something similar happens in a simulation if you feed the solution with an excess number of perfect measurements. Because everything fits exactly, all the residuals are zero although there are redundant measurements. The calculation of $\hat{\sigma}_0{}^2$ is then $\hat{\sigma}_0{}^2 = 0/r = 0$. If this value is used in (iii) then the parameters are also estimated to have zero variances and covariances, i.e. they are also perfect. This is mathematically logical and consistent but not of much practical use. Instead the user would arbitrarily set a value for $\sigma_0{}^2$, e.g. $\sigma_0{}^2 = 1$, and use this in combination with simulated variances for the measurements.

This would create a specific weight matrix and hence a specific estimate for the errors in parameters corresponding to this particular simulated measurement quality. If the simulated variances of measurements are then scaled up to some other value, the weight matrix is similarly re-scaled. This will not change the values of the simulated parameters since $\mathbf{N^{-1}}$ scales down by the same amount as $\mathbf{t}$ scales up (equation (ii)). However the simulated error estimates for the parameters will scale up, as they should do when you simulate the use of lower quality measurements. Note that re-scaling the measurement quality is not quite the same as re-scaling the variance factor $\sigma_0{}^2$, discussed earlier.

$\sigma_0{}^2$ *and* $\hat{\sigma}_0{}^2$ *are not very similar in value*

This implies one of the problems discussed in "Practical use of variance factor" on page 43.

If there is reasonable redundancy the lack of agreement indicates that the weights are not correctly balanced, either because of the presence of some bad measurements which should have a lower weight or because some group of measurements is assumed to be better or worse that it really is. Ideally the bad measurements should be found and removed or the weights re-assigned more appropriately and the solution run again.

If redundancy is low and there is no good reason to eliminate measurements or re-assign weights, then the best procedure is to use $\sigma_0{}^2$.

$\sigma_0{}^2$ and $\hat{\sigma}_0{}^2$ are consistent

When $\sigma_0{}^2$ and $\hat{\sigma}_0{}^2$ are consistent, Mikhail recommends always using $\sigma_0{}^2$ (see Mikhail[1] p250).

Directly assigned weights, e.g. unit weighting.

When weights are directly assigned, the user has not consciously chosen a value for $\sigma_0{}^2$ but $\hat{\sigma}_0{}^2$ can be calculated and used for error propagation. From the above discussion it can be seen that the best approach in error propagation is to obtain a good feel for the errors in the original data source and propagate this information by using $\sigma_0{}^2$, rather than using $\hat{\sigma}_0{}^2$.

4.9.4 Error propagation with unit weighting

In shape fitting routines ***Axyz*** permits the user to calculate shapes using unit weighting for the fitted points. As a result all coordinates are treated equally and no information about the actual measurement quality is used.

As explained in "The scaling mechanism at work" on page 48, the preliminary or calculated variance factors are required in order to obtain quality estimates of the computed parameters. When shape fitting uses existing variances to create weights then the preliminary variance factor (value = 1) is used.

In the case of unit weighting the calculated variance factor is used. In this case it has units of variance and its size should be typical of the residuals produced. Unfortunately, if there is very little redundancy the residuals will not be very meaningful in statistical terms and if there is zero redundancy there is no information at all from the results of processing.

If the redundancy is zero the variance factor cannot be calculated. In this case errors are propagated using a preliminary variance factor of 1, which is what is done for error propagation when weighting by variance. This is done simply in order to generate values for the data base. However since the weights are dimensionless in this case, the variance factor is not dimensionless. In fact, it is equivalent to a variance of 1 m^2. In effect you are estimating errors in shape parameters assuming that the fitted points have standard deviations of 1m.

4.10 Quality of angle and distance measurements

The quality of measured angles and distances determine the quality estimates for all measured coordinates and any further 3D or shape data derived from those coordinates.

Within **Axyz**, angles and distances are assumed to be uncorrelated. Their quality is defined by standard deviations <u>estimated</u> by the user from experience or analysis of previous work. These estimates are assumed to depend on a particular instrument.

Default values of standard deviation are therefore separately defined for horizontal angles, vertical angles and distances at each station during the station setup procedure. These values can be changed at any time. The default value cannot, however, be individually set for a specific measurement.

> **Note**
> If these estimates are incorrect then the calculated quality estimates of derived elements, particularly point coordinates, will also be incorrect. If you assume measurements are better than they actually are, then point coordinate quality will be estimated better than it actually is and vice versa. It is therefore worthwhile ensuring that these estimates are good.

4.10.1 Multiple and 2-face measurements

Within **Axyz** it is possible to make multiple, repeated measurements to the same target.

If a repeated single face pointing is made, an average value is calculated and this is then stored as the representative pointing.

Single pointings in both faces are separately stored, but combined into a single representative pointing when used in the Orientation Module and Single Point Solution.

Multiple pointings in both faces will generate an averaged value for each face which are then used in the same way as single face pointings.

Multiple pointings can also be made in the form of bolt hole measurements, which require evenly spaced pointings around the edge of a bolt hole. In this case the average horizontal and vertical angle is

calculated as the single representative pointing to the hole centre and stored in the job file.

In all cases of multiple measurements the same default standard deviation is used for the final single representative value, as defined in the station setup. An improved quality figure is not generated.

5. Coordinates and coordinate systems

5.1 Introduction

A 3D object is defined by points located in a particular coordinate system. A coordinate system is a set of 3 mutually perpendicular, intersecting axes and the coordinates of a point are defined with respect to these axes.

The axes intersect at the origin and a point is usually located by measuring its distance from the origin along each axis. This is not the only type of coordinate which is possible and the various types in use are described in this section.

The coordinate values will also depend on the location of the origin and the directions in which the axes point. Although the location and direction of a coordinate system's axes does not affect the object's shape, it is often convenient to work with coordinate values which have some direct meaning, for example where the origin is the centre of a drill hole.

Axyz permits users to define any number of coordinate systems and offers several methods to do this. Once defined, users can easily switch between coordinate systems and coordinate types.

5.2 Base coordinates and other coordinate systems

All 3D points on an object are derived from angle and distance measurements made by the network of instruments which surround it. The locations of the instruments are defined in a single common coordinate system calculated by the Orientation Module. By using a single system, any one point can be related to any other and this coordinate system represents the source of all measured 3D data. Until the system exists common coordinate data cannot be generated and further coordinate systems and shapes cannot be defined.

This fundamental system is known as the *base coordinate system*. Its parameters are stored in the "default" workpiece and given the name "BASE", which is also used as the type name. *Axyz* stores all 3D object coordinates here.

The base system has the following general properties
- Right-handed Cartesian coordinates
- Units of dimension in metres

However, the user frequently wants to view the data in a different way and can make the following alterations:

- Transform coordinates to a different origin and orientation, for example to force a particular point to represent the origin (0,0,0)
- Choose a different type of coordinates, such as cylindrical (r,α,h) rather than Cartesian (x,y,z)
- Select different dimensional units such as millimetres or inches

Any of these changes do not affect internal coordinate values and coordinate data remains stored in its standard base format. Instead such changes requested by the user are stored, together with any computed parameters necessary to implement them. Whenever coordinate data is displayed on screen, printed out or transferred to an external file it is temporarily modified according to the defined coordinate type, measurement units and parameters of the chosen output coordinate system. In effect the internal data is passed through a mathematical filter every time the user views it.

5.2.1 Where is the base coordinate system?

The base coordinate system is defined by the Orientation Module, which calculates the station locations and any orientation points in a common network.

Briefly, in the final optimization of the network, the base coordinate system ends up in one of two places.

1. On the object
This is a feature of a controlled (object) orientation. Reference coordinates on the object, defined by *control points,* are included in the orientation procedure. The origin and axes which define the control coordinates are used as the origin and axes of the base system.

2. At, or close to, one of the stations
This is a feature of a relative (local) orientation. In this case control points are <u>not</u> included in the orientation procedures. Since a meaningful origin and axes are not provided, the origin and axes of the base system are arbitrarily located at one of the stations. Depending on a further optimizing option the base axes may drift slightly away from this initial position.

5.2.2 Can existing base coordinates change?

Base coordinates of existing points can change when the number of measurements to a point change and/or the number of stations in the network changes. In both cases there is different information available which would normally result in a change of optimized coordinate values. Typically this occurs when additional measurements are made to a point or a new station is added to the network, but deletion of measurements or stations has the same effect.

If a new measurement is made to a point from an existing station, the point's coordinates are automatically re-calculated. These optimized point coordinates must take account of this new measurement and will change slightly from the old values.

If a new station, with its new measurements, is brought into a network by re-calculating the orientation, the optimal shape of the entire network will change slightly to accommodate the new data. Users have the option to fix existing stations in their current locations but this is not an optimized result. If an optimized result is chosen all existing points will be re-calculated to reflect the changes in station locations and angular orientations.

If the re-calculation of orientation uses the same optimizing procedure, then changes in coordinate values are expected to be small. However a change in optimizing procedure can have much larger effects on existing base coordinate values. For example, switching from a relative orientation to a controlled orientation will move the base system origin to a completely different place and cause base coordinate values to change significantly.

5.2.3 Definition of other coordinate systems

All other coordinate systems created by the user are defined with respect to the base origin and axes by a shift, rotation and optional scale change.

For a definition of the parameters, see "Transformation parameters" on page 88.

5.3 Classification of coordinate systems

5.3.1 Classification in the database (job file)

In the database, coordinate systems are identified in two ways:

Coordinate system name

In *Axyz* each coordinate system is given a name to uniquely identify it.

Coordinate system type

Coordinate systems are assigned a type which indicates the method used to create them.

Example 1
A system of type "base" is created by the Orientation Module. This is the coordinate system used for storing all measured data and derived elements and only one of these exists.

Example 2
A system of type "best-fit" is created when a best fitting 3D transformation is calculated in order to display coordinates in a coordinate system used to design the object.

Example 3
If the axes defining a shape are used as a coordinate system the shape's type will be given as the coordinate system type, e.g. "circle" if a circle has been used.

5.3.2 Descriptive classification

Three additional terms are often used in discussions and explanatory text and which indicate what coordinates and axes relate to.

Object coordinate system

The most meaningful coordinates which describe an object are generally defined by selected features on the object itself. By suitably locating the origin and axes of the coordinate system, measured coordinate values can correspond very closely to the design coordinates used to manufacture the object. This is very convenient for building and inspecting features.

Object coordinate systems can be created in two ways:
- By orienting the measurement network to control points
- By transforming an existing system onto design values, see 5.7 Transformation of coordinates.

Local coordinates and axes

The term *local* is often used to indicate that a coordinate system is most relevant to the feature which defines its origin, i.e. the coordinates and axes are primarily of local interest rather than relevant to all measured features.

Example 1: Local shape coordinate system
A circle, for example, is created with its centre defining the origin of a local coordinate system. The xy axes of this system are in the plane of the circle and the z axis is perpendicular to the plane.

Example 2: Local station coordinates
If a theodolite, for example, occupies a station in the measurement network, its local axes, with origin at the centre of rotation, sometimes define the base coordinate system within which 3D data are recorded [xref to base system].

> **Note**
> Although every oriented station in a network can, in principle, display local station coordinates, **Axyz** does not have a function which permits every station to be used in this way.

Instrument coordinates

Individual polar measuring instruments such as Total Stations and laser trackers directly provide enough information to locate points in 3D space. It is sometimes convenient to see these coordinates before the station occupied by the instrument has been oriented into a network of two or more stations.

Like local station coordinates, instrument coordinates are referred to the internal axes of an instrument. However, unlike local station coordinates, only points measured from the relevant instrument are displayed in its own instrument coordinates. Other points in the network, measured from instruments located at other stations, cannot be displayed in a different instrument's coordinate system. Instrument coordinates are intended for

use at stations not yet oriented and, by definition, the necessary link does not then exist between different instruments.

> **Note**
> There is a special case where a single Total Station or laser tracker is used to measure an object and the base system is defined by the measuring station. In this case base system coordinates are equivalent to local station coordinates which are identical to instrument coordinates

ECDS local and object systems

ECDS users are familiar with the terms "local" and "object" coordinates and coordinate systems. These are closely related to the ECDS terms "local" (relative) and "object" (controlled) orientation, as well as the ECDS "local to object" transformation.

ECDS allowed only two systems to exist, one located in a station which was called "local" and generated by a "local" (relative) orientation. The other was located in the object. This was named "object" and generated directly by an "object" (controlled) orientation or by a 3D transformation from a "local" system onto the object.

Axyz has adopted the more flexible approach introduced in ManCAT in which there can be many coordinate systems. Since more than two are possible, the terms "local" and "object" cannot easily indicate a specific coordinate system but the use of unique names makes clear which system is being identified.

In **Axyz** the terms "local" and "object" are therefore used in a descriptive way as outlined earlier.

5.4 Examples of multiple coordinate systems

5.4.1 Example 1: Single instrument, relative orientation
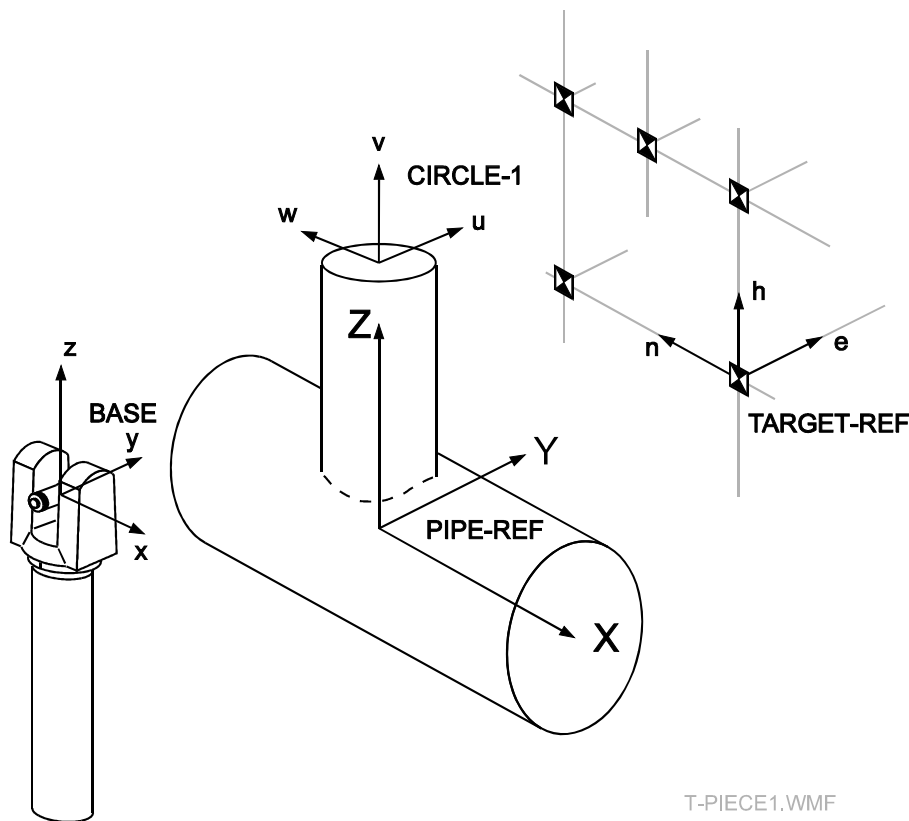
T-PIECE1.WMF

The diagram shows a section of pipework measured by a single Total Station with several defined coordinate systems. (The axes look parallel but this is simply a convenience in drawing the diagram. In practice they may point in any direction.)

Original coordinates were computed in a *relative* coordinate system called "BASE", located by default at the centre of the Total Station and labelled xyz.

The coordinate system used to design the pipework has main axes along the cylinder axes and origin at their intersection. These are labelled XYZ. This is the main *object* coordinate system. By measuring points on the cylinders from the instrument and fitting cylindrical surfaces to these data, a coordinate system called "PIPE-REF" has been created which is equivalent to this original object coordinate system.

An circular section has also been measured as "CIRCLE-1". It has its own *local* reference axes labelled uvw which can also be used for viewing other data.

Finally the user has established an array of fixed targets with accurately known coordinates in a coordinate system labelled "TARGET-REF". In effect, these define a second object in the measurement field. The user could measure the targets and calculate a best-fit 3D transformation which would locate the coordinate system "TARGET-REF" with respect to the base system. This is not particularly useful. It is more likely that this configuration would be used to control the orientation of a triangulation network, as the next example shows.

5.4.2 Example 2: Multiple instruments, controlled orientation
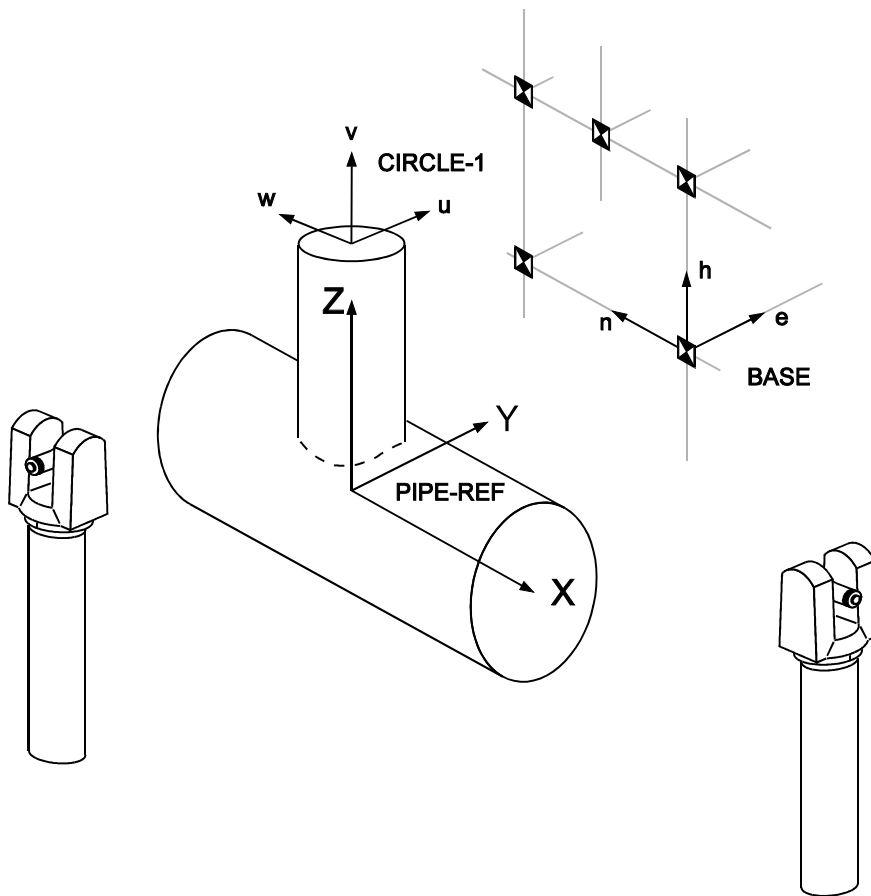
T-PIECE2.WMF

In this configuration the fixed reference targets are used as control points in a second measurement of the pipework, for example using a triangulation technique. As a result of the *controlled* orientation, the coordinate system defining the reference targets becomes the base system. This is labelled "BASE" but is equivalent to the system labelled "TARGET-REF" in the earlier example.

Once again shape fitting methods can establish the coordinate systems "PIPE-REF" and "CIRCLE-1" which are now defined relative to the new base coordinate system.

5.5 In brief: changing coordinate types and systems
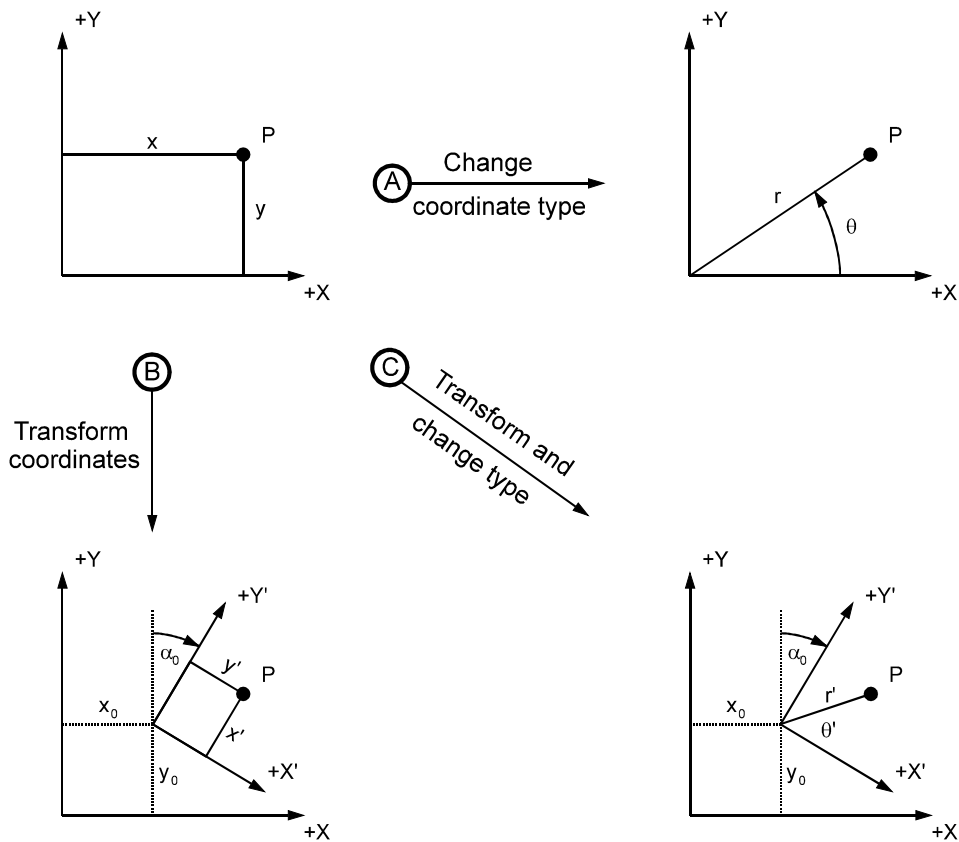
CRDSYS01.WMF

For a 2D situation the diagram summarizes two ways of presenting coordinates. The situation in 3D is essentially the same.

(A) Changing the type of coordinate

Coordinate systems are normally based on orthogonal (rectangular) axes and this is the case in **Axyz**. To specify the position of a point with respect to the axes it is most common to record its distance from the origin along each axis. Coordinates of point P are then (x,y).
However this is not the only way of defining the position of P with respect to the axes. It could also be located by defining its distance from the origin along a specified direction. In the example the two elements defining the location of P are called *polar coordinates* and expressed as (r,θ).

Note that there is no change of origin or reference axes. A different coordinate type is simply an alternative way to locate a point with respect to the same coordinate system. In fact, it is necessary to keep the axes in order to define uniquely any rotational values, for example θ.

Converting between coordinate types is a simple mathematical procedure. In the 2D case above:

From (x,y) to (r,θ):

$r = x^2 + y^2$
$\theta = \arctan (y/x)$
(must use a function which finds the correct quadrant where P is located)

From (r,θ) to (x,y):

$\quad x = r \cos(\theta)$
$\quad y = r \sin(\theta)$

(B) Transforming coordinates to a different coordinate system.

Sometimes the position of the coordinate system is not convenient and you want to transform coordinates into another coordinate system.

The diagram shows how the coordinates of P would be defined in a new coordinate system X',Y'. If you know the transformation parameters x_0, y_0, α_0 you can easily convert (x',y') to (x,y) or (x,y) to (x',y'). The user has several options for defining the transformation parameters:

- Manually specify the shift (x_0, y_0) and rotation (α_0)
- Select a point to be the new origin and another point to define the direction of x' or y'. Transformation parameters can then be automatically computed from the existing coordinates of these two points.
- Supply coordinates in the new system for at least two points measured in the old system. Transformation parameters can again be automatically computed.

Transforming between coordinate systems is again a simple mathematical procedure. In the 2D case above:

From (x,y) to (x',y'):

$\quad x' = (x - x_0) \cos(\alpha_0) - (y - y_0) \sin(\alpha_0)$
$\quad y' = (x - x_0) \sin(\alpha_0) + (y - y_0) \cos(\alpha_0)$

From (x',y') to (x,y)

$$x = x_0 + x' \cos(\alpha_0) + y' \sin(\alpha_0)$$
$$y = y_0 - x' \sin(\alpha_0) + y' \cos(\alpha_0)$$

(C) Transforming coordinates and changing type

Both techniques can be combined. Once the transformation has been specified the transformed data can then be displayed in any of the optional coordinate types.

5.6 Mathematical rules

It is difficult to discuss coordinate systems and transformations without establishing some conventions, two of which are briefly reviewed here.

5.6.1 Right and left handed axes

The order in which axes are defined is important. A simple 2D case illustrates the point.
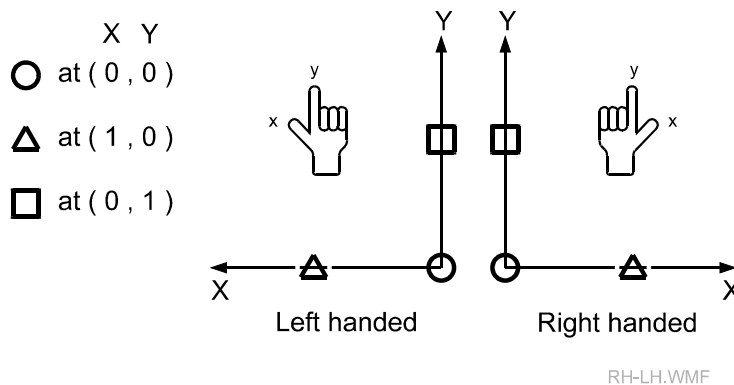
If axes are defined in an order say X then Y, the set of coordinates defining a point are assumed to correspond to this order. For example, a point with coordinates (3.8,4.5) has X value = 3.8 and Y value = 4.5. However, the axes can be physically drawn in two different ways and the shapes form a different pattern in each. One pattern is a mirror image of the other.

Your hands provide a simple way to remember the arrangement. With palms up and thumb and forefinger of the left and right hands extended as shown, they point in the positive directions of the first and second axes respectively (here called X and Y). If the third finger is extended upwards, it will point in the positive direction of the third axis of a 3D system (typically called Z).

5.6.2 Direction of rotation

Transformations usually involve either rotating objects or their reference axes into new positions. Directions for positive rotations follow a right handed rule.
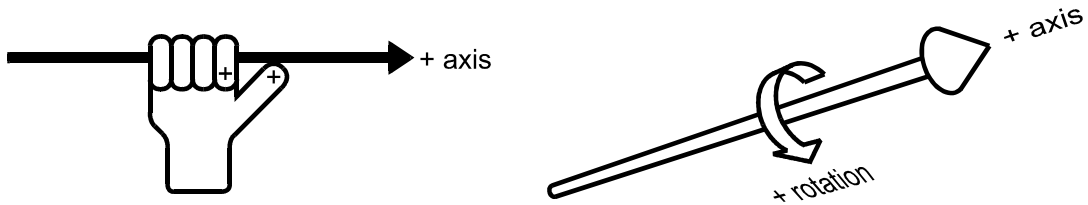
Right hand rule for determining the positive direction of rotation about an axis

RH-ROTNa.WMF

Right handed systems
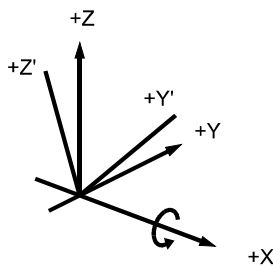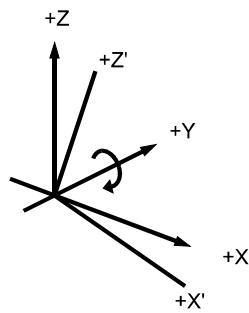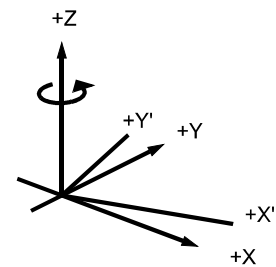
Positive rotation of axes about X   Positive rotation of axes about Y   Positive rotation of axes about Z

RH-ROTNb.WMF

Left handed systems
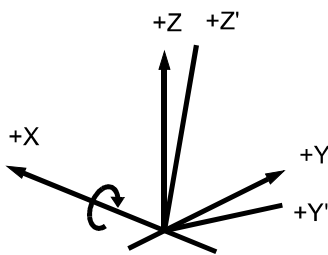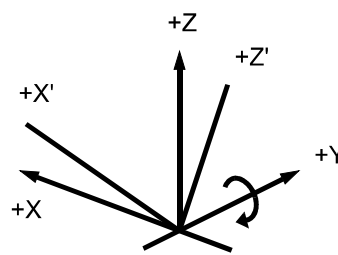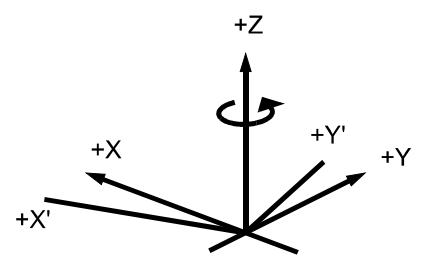
Positive rotation of axes about X   Positive rotation of axes about Y   Positive rotation of axes about Z

RH-ROTNc.WMF

The upper diagram shows the right hand rule for positive rotations. Imagine you grip the axis in your right hand with the thumb pointing towards the positive direction of the axis. The curl of your fingers towards the tips then shows the positive direction of rotation. Note that this definition is not linked to right and left handed axes and is simply a way of defining a positive angular direction about a single axis.

The lower two diagrams show the creation of new axes by applying positive rotations about each of the existing axes, for both right and left handed systems. A positive rotation about X moves Y and Z to Y' and Z' respectively. The X axis itself does not change in this case. Similar effects are seen for positive rotations about Y and Z.

In summary, right handed axes:
+ x rotation moves +y axis towards +z axis
+y rotation moves +z axis towards +x axis
+z rotation moves +x axis towards +y axis

In summary, left handed axes
+ x rotation moves +z axis towards +y axis
+y rotation moves +x axis towards +z axis
+z rotation moves +y axis towards +x axis

Note
Rotations about the X, Y and Z axes are often labelled with the Greek letters ω/Ω (omega), φ/Φ (phi) and κ/K (kappa) respectively.

5.7 Transformation of coordinates

5.7.1 Overview

The purpose of a transformation is to view coordinates in a different coordinate system from the one they are stored in. If the origin and direction of axes of the new system is known with respect to the old system, it is a simple mathematical procedure to convert coordinates from one to the other (see "In brief: changing coordinate types and " on page 63).

Transformation of the coordinate system is often used if measurements have initially been made in a relative system defined by an instrument but it is more convenient to view data in an object system which relates to design or blueprint coordinates.

A new coordinate system can be defined by 4 different transformation methods.

In brief: Manual transformation

At the simplest level coordinate transformation involves re-scaling, shifting and rotating the existing reference axes. If you manually specify

the scale factor, shift or rotation it is then a simple matter to define coordinates with respect to the new reference axes rather than the old ones.

Axyz provides functions to separately:
- Re-scale current coordinates
- Translate (shift along each axis) the current origin
- Rotate current axes

In brief: Alignment of axes

The new reference axes can be aligned to a particular set of points. For example:
Point A should be the origin (0,0,0), the X axis should pass through point B and the XY plane should pass through point C (which must form a triangle with A and B)

In brief: Transformation onto reference coordinates

The new coordinate system can be defined by the design values of selected measured points.
This method requires at least 3 measured points which form a triangle to have nominal design values in the object's own coordinate system. Since the measured points will not conform exactly to the design shape due to manufacturing tolerances and measurement uncertainties a best fit using least squares is used for the transformation.

In brief: Transformation onto computed shapes

The new coordinate system can be the local coordinate system of any computed shape

5.7.2 Manual transformation

Manual transformations require the definition of an initial coordinate system since scale factors, shifts or rotations are relative values which will have different effects depending on the starting point. In ***Axyz*** therefore the functions below need to know the starting system.

Manual transformations can be applied successively, with each new definition based on the previous one. However the complete transformation sequence is referenced directly back to the base system, in which coordinates are actually stored, by combining the individual transformations into a single transformation.

Starting system	Transformation	New system	Action
BASE	t1➜	SYST1	Create a new coordinate system SYST1 by some method.
SYST1	t2➜	SYST2	Apply scaling, shifting or rotation to SYST1 to create SYST2
			Axyz computes: t1 x t2 = tA
			Stored information: BASE tA➜ SYST2
SYST2	t3➜	SYST3	Apply scaling, shifting or rotation to SYST2 to create SYST3
			Axyz computes tA x t3 = tB
			Stored information: BASE tB➜ SYST3

etc.

Re-scale

a) Apply scale factor = 3

x'=3x

b) Apply scale factor to to make hole separation =10

10.0

3.14
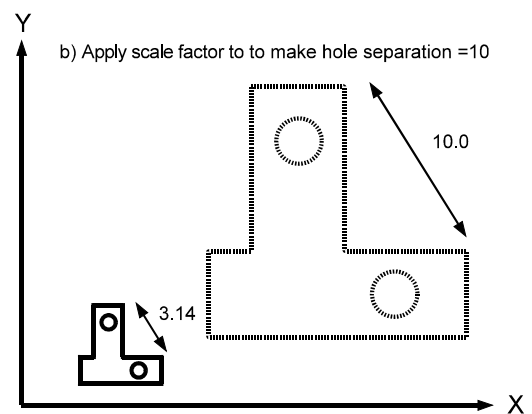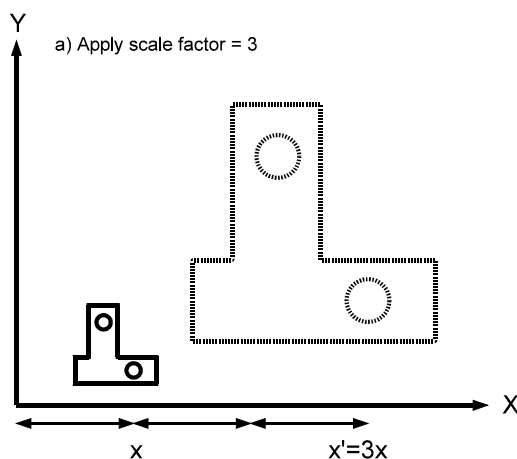
TRF-SF.WMF

A coordinate system can be re-scaled by either defining the scale factor directly or by re-defining the separation between any two points to have a certain value.

The 2D example shows:
a) Re-scaling by a factor of 3, i.e. all coordinates are multiplied by the scale factor.
b) Re-scaling by defining the distance between the hole centres to be 10 units. The current separation is 3.14 units and the program derives a scale factor of 10/3.14 = 3.1847. All coordinates are then multiplied by this factor.

The 3D procedure functions in exactly the same way.

Note
Defining a new coordinate system by re-scaling an existing system can be useful if you want to work in model coordinates. Suppose you are constructing a 1/5 scale model of a component and you directly created a BASE system using control points with design values of the actual component. BASE values are therefore in an object system. You would also like to see the coordinates at the model scale. Define a new coordinate system, e.g. MODEL, with a scale factor of 0.2 applied to BASE and switch between them as required.

Rotate
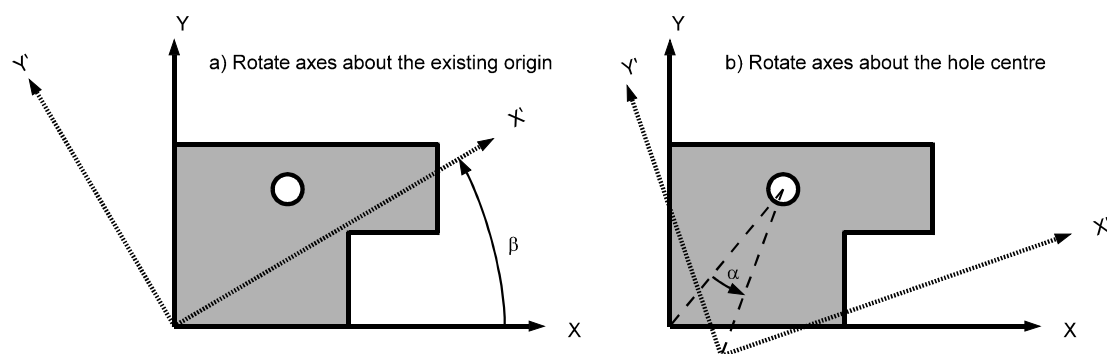
TRF-ROTN.WMF

New coordinates can be defined by rotating the axes of the current system. ***Axyz*** permits a single rotation about any one of the 3 existing coordinate axes using a right-handed system. This causes only the other two axes to move. Additionally the fixed point of rotation can be either the existing origin or any other named point.

Rotation about origin

A positive rotation β about the Z direction which is not indicated but is positive out of the page and towards the reader. The rotation is applied through the existing origin. The bottom left corner of the object is located at this origin and it remains at the origin in the new system.

Rotation about specific point, e.g. hole centre

A positive rotation α about the Z direction, applied through the hole centre. This time the origin changes. The hole centre will have the same coordinate values in the new system but the bottom left corner will have new coordinate values because it is no longer located at the origin in the new system.

Note that the axes are rotated, not the object.

Shift

a) Shift axes by sx and sy   b) Shift origin to hole centre   c) Shift hole centre to coords. (20,10)
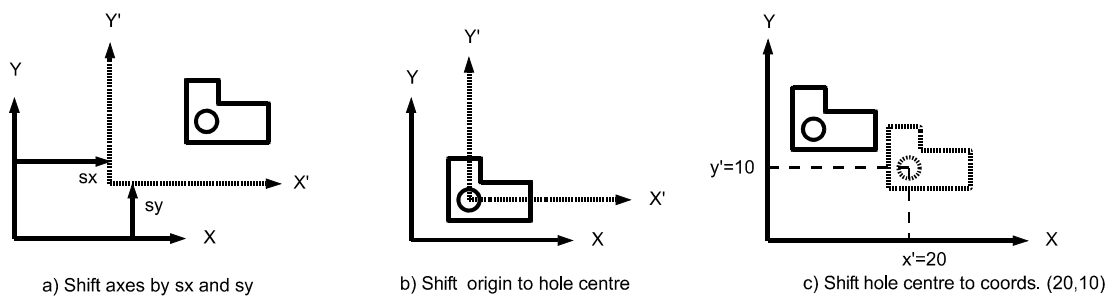
TRF-SHFT.WMF

Shifts can be applied in 3 different ways. The 2D example is shown for convenience but the operation in 3D is identical. Shifts are applied using a right handed system.

Shift axes by defined amount

The axes can be directly shifted in any axial direction. All X coordinates are <u>reduced</u> by the value sx and all Y coordinates are <u>reduced</u> by the value sy.

Shift axes to defined point

The shift is calculated by placing the new origin of coordinates at a defined point. In the example the hole centre becomes the origin. If the old point coordinates are (px,py) then these effectively become the axial shifts, i.e. all X coordinates are <u>reduced</u> by the value px and all Y coordinates are <u>reduced</u> by the value py.

Assign coordinates to named point

This is easier to visualize as a shift of the object. The example shows the hole centre coordinates increased by an amount in X to 20 and reduced by an amount in Y to 10. All X coordinates are increased by the same amount and all Y coordinates reduced by the same amount.

5.7.3 Axis alignment

This technique creates a new object coordinate system by aligning new axes to directions and planes defined by the object points. The new axes may be freely oriented to object points or they may be placed in the object but additionally oriented to gravity. The technique defines a new right-handed system of axes.

Axis alignment does not require a starting system since the new system axes are not defined relative to an existing system but defined directly on the object. However for gravity oriented systems it is <u>assumed</u> that the Z axis of the base system represents the direction of gravity.

Note

It is not possible to tell from coordinate values if a Z axis represents the direction of gravity and the ***Axyz*** database does not explicitly record if the base system has been levelled. For this reason the algorithms can only assume that this is the case if the user requests an alignment to gravity. No errors are introduced if a new system is "aligned to gravity" when the base system is not levelled. In reality, the new system is then only aligned to the Z axis of the base system .

Axis alignment by free orientation: Summary
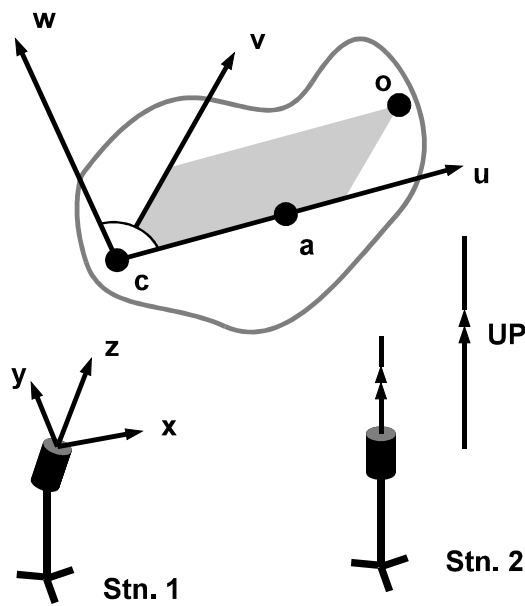
Axis alignment: free orientation   ALIGN01.WMF

The example shows a base coordinate system xyz established at station 1 and which is not levelled. A second station 2 is shown which is levelled, i.e. its primary axis of rotation is parallel to the direction of gravity with the upwards direction shown by the double headed arrow. This merely illustrates a mixture of possibilities. It is irrelevant to the free orientation technique if there are one or more stations and if any, all or none are levelled.

New axes are defined in the object by first selecting a *controlling point c.* This point acts like an origin but it does not need to have coordinates (0,0,0) in the new system and can take any value such as (100, 100, 50).

A second *axis point a* is selected so that the line c-a indicates the direction of the first new axis. In the example this is labelled u and shown positive from c to a in the example.

Finally a third *offset point o* is selected which creates a plane c-a-o, shown shaded in the example. The purpose of the plane is to define the second axis labelled v in the diagram. This lies in the plane and is perpendicular to the first axis. In the example it is shown positive towards point o.

The third axis labelled w in the example can then be automatically calculated. It is made perpendicular to the plane and creates a right-handed system with the other two. The diagram assumes that the right-handed order of axes is (u,v,w).

Axis alignment by free orientation: Steps in procedure
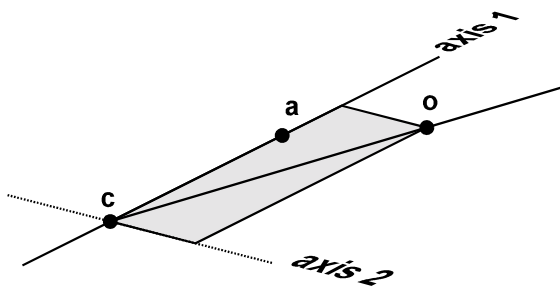
Free Axis alignment:
Definition of axes 1 and 2
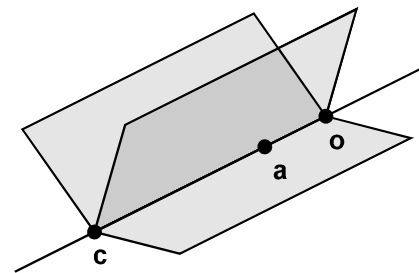
Offset point o must not lie on the line
joining the controlling and axial points

ALIGN03.WMF

In the following discussion the labels "axis 1", "axis 2" and "axis 3" only
indicate the sequence in which axes are defined and do not define the order
of coordinates.

The line between controlling point c and axis point a defines axis 1, which
can be made positive in either direction.

The offset point o must form a triangle with c and a to ensure that only one
plane and a unique second axis can be defined. If the offset point is on the
same line as c and a, an infinite number of planes can be generated through
the line, three of which are shown in the example.

Axis 2 lies in the plane and is perpendicular to axis 1. The direction from
axis 1 towards the offset point o can be chosen as either the positive or
negative direction of axis 2.
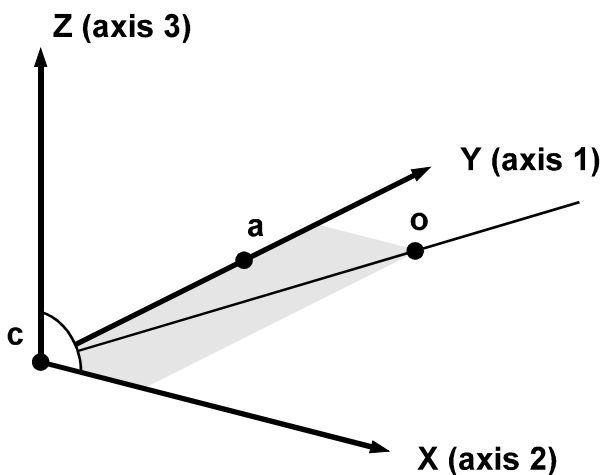
Free axis alignment:
Definition of axis 3 (example).   ALIGN04.WMF

Axis 3 is automatically created to make a right-handed rectangular set of axes with axes 1 and 2. In the example, right-handed coordinates are assumed to be labelled X, Y and Z. The Y coordinates have been assigned to axis 1 and made positive from c to a. X coordinates were assigned to axis 2 and made positive towards the offset point. Z coordinates were then automatically assigned to axis 3 such that their positive direction makes a right-handed rectangular system with the other two axes.

Axis alignment by free orientation: Offsets
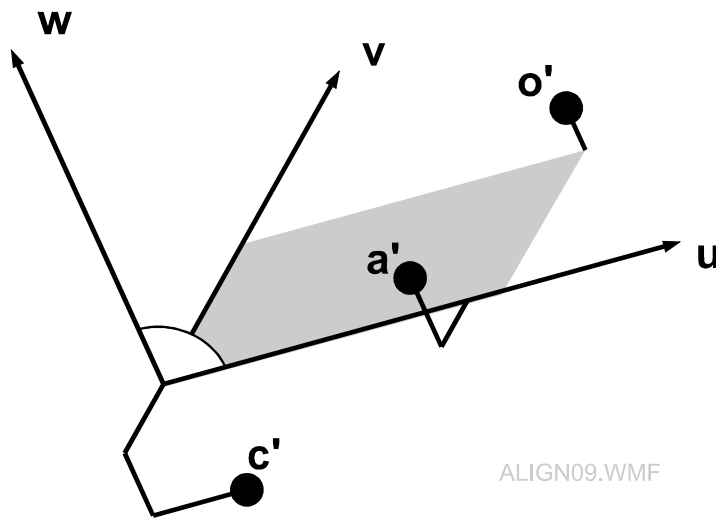
ALIGN09.WMF

It may not be convenient in practice to measure points which lie exactly on the new axes or planes. Instead, nearby points with known offsets in the new coordinate system can be used.

The example shows how the controlling point c may be substituted by point c', with 3 axial offsets along u, v and w.
Point a may be substituted by point a' with two offsets (along v and w in the example).
Point o may be substituted by point o' with a single offset (along w in the example).

Axis alignment by orientation to gravity: Summary

Partial alignment to gravity ensures that the new first and second axes lie in a vertical plane.

Full alignment to gravity ensures that:
- The new first and third axes define a horizontal plane
- The new second axis is parallel to the direction of gravity (positive UP or DOWN).
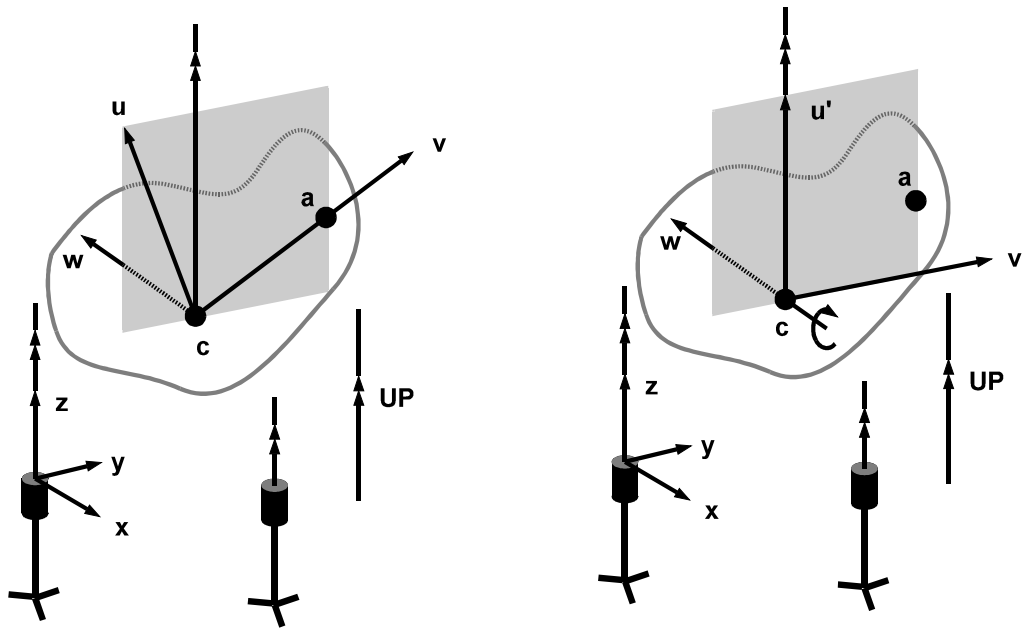
a) Axis alignment: **Partially** oriented to gravity   b) Axis alignment: **Fully** oriented to gravity

ALIGN02.WMF

In example (a) a base coordinate system xyz has been established using levelled instruments. The base z axis is therefore parallel to the direction of gravity and positive up.

A controlling point c is selected as an origin for the new axes in the object.

The axis point a defines the direction of the first axis labelled v along the line c-a.
A vertical plane is then set through c and a. This plane defines the second axis labelled u which lies in the vertical plane and is perpendicular to the first axis. In the example it is positive in an upwards direction.

The third axis labelled w in the example can then be automatically calculated. It is made perpendicular to the vertical plane and creates a right-handed system with the other two axes. (The example assumes the right-handed order of axes is u,v,w.)

In example (b) the above procedure is taken one stage further. When aligning axes to gravity, it is very likely that the existing coordinates are very nearly levelled and the new axes should show this. When axes are fully aligned to gravity it is therefore assumed that the line c-a is approximately horizontal and the new second axis u is intended to be vertical. This also means that the plane defined by the first and third axes (v and w) is also approximately horizontal.

However since the line c-a may not be exactly horizontal a correction must be made to the directions of the new axes in order to level the new system exactly. By rotating the system about the third axis w, the plane containing the v and w axes can be set exactly horizontal.

> **Note**
> When axes are fully aligned to gravity, the first axis v no longer passes through the axis point a.

Axis alignment by orientation to gravity: Steps in procedure

Axis alignment to gravity:
Definition of axes 1 and 2

Controlling and axial points
must not form a vertical line
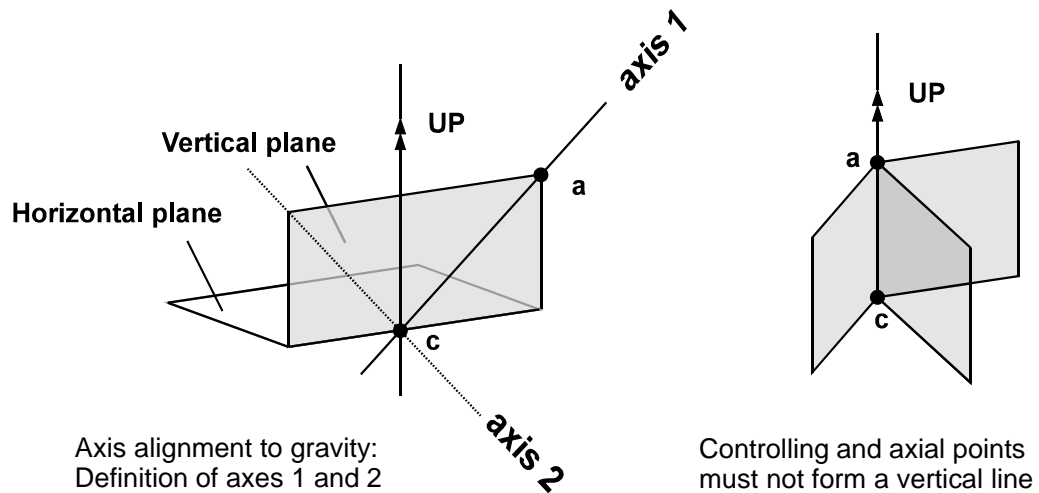
ALIGN05.WMF

In the following discussion the labels "axis 1", "axis 2" and "axis 3" only indicate the sequence in which axes are defined and do not define the order of coordinates.
The line between controlling point c and axis point a defines axis 1, which can be made positive in either direction.

A vertical plane is then placed through c and a which will define axis 2. Axis 2 lies in the vertical plane and is perpendicular to axis 1. Its positive direction may point either up or down.

The axis point a must have some horizontal separation from controlling point c. If it lies vertically above or below c then an infinite number of vertical planes can be placed through c and a. Three of these are shown in the diagram. If the line c-a is reasonably close to the horizontal it is also easy to recognize the up and down directions for specifying the positive direction of axis 2.
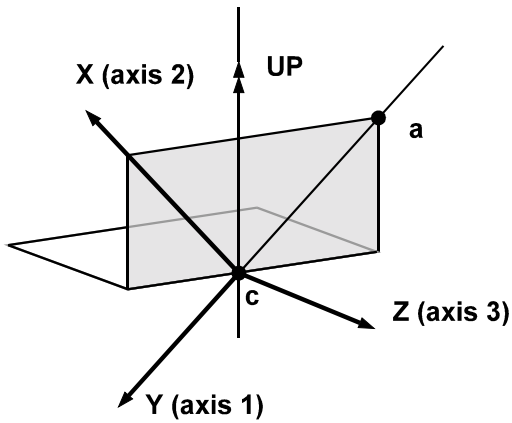
Axis alignment to gravity:
Definition of axis 3 (example).   ALIGN06.WMF

Axis 3 is automatically created to make a right-handed rectangular set of axes with axes 1 and 2. In the example, right-handed coordinates are assumed to be labelled X, Y and Z. The Y coordinates have been assigned to axis 1 and made positive in the direction from a to c. X coordinates were assigned to axis 2 and made positive upwards. Z coordinates were then automatically assigned to axis 3 such that their positive direction makes a right-handed rectangular system with the other two axes.
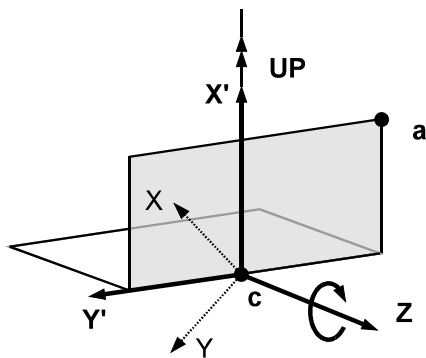
Axis alignment to gravity (X up):
Set YZ plane horizontal
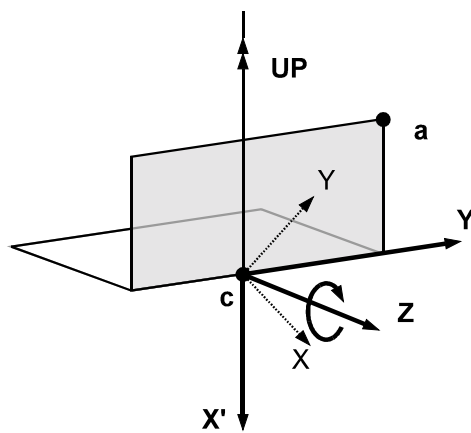
Axis alignment to gravity (X down):
Set YZ plane horizontal

         ALIGN07.WMF

Full orientation to gravity is possible with a final rotation about the axis which is exactly horizontal, i.e. the axis labelled Z in the example. This rotation makes the plane of the defined first and third axes exactly horizontal and the defined second axis exactly vertical. In the example the first axis was defined as Y, the second as X and the third as Z. The example also defined X as pointing up. The diagram shows what happens if X is either positive up or down. In both cases the YZ plane is made exactly horizontal.

Axis alignment by orientation to gravity: offsets
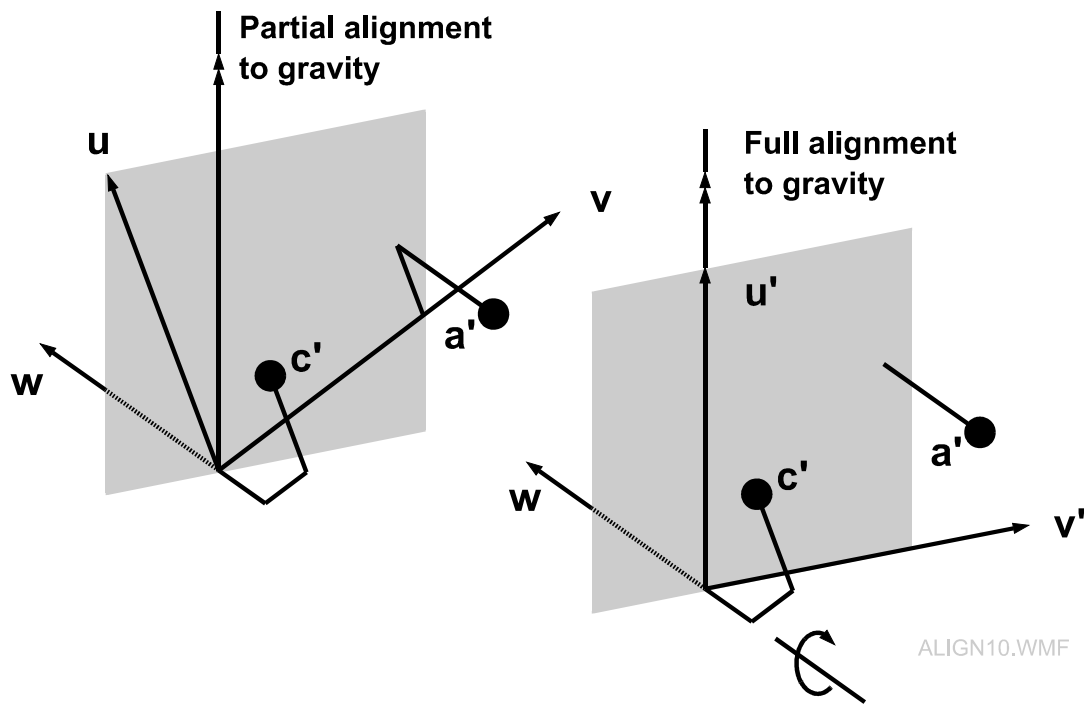
As in alignment by free orientation, it may not be convenient in practice to measure points which lie exactly on the new axes or planes. Instead, nearby points with known offsets in the new coordinate system can be used.

The example for partial alignment shows how the controlling point c may be substituted by point c', with 3 axial offsets along u, v and w.
Point a may be substituted by point a' with two offsets (along v and w in the example).

If the system if fully aligned to gravity, then point a' can only have one offset (along w in the example).

5.7.4 Transformation onto reference coordinates

The final way of defining a new coordinate system is to do the mathematical equivalent of picking up an object and moving it into its correct position by attaching it to fixed locating points. This is known as a *transformation onto reference coordinates* and in some textbooks is called a *similarity transform.*
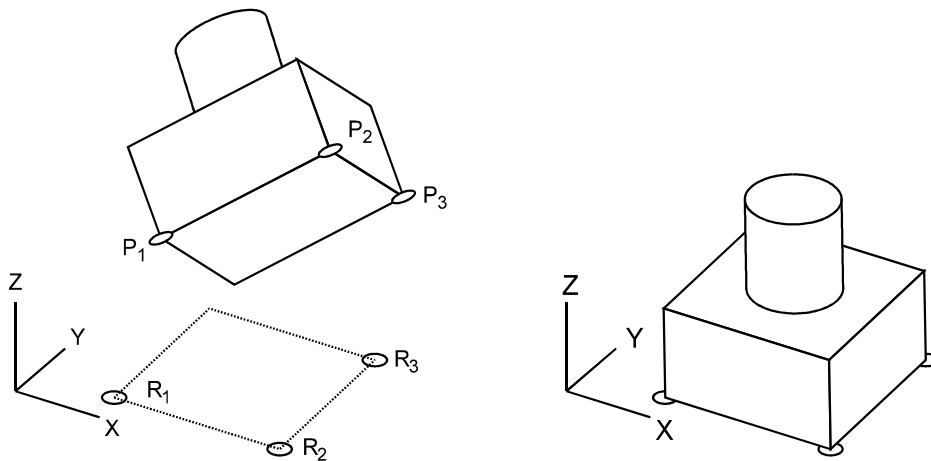
Basic principle of 3D transfromation   TRF3D.WMF

The components of this action are a shift, coupled with angular tilts and twists which match up object points P_1, P_2, P_3 with their design counterparts R_1, R_2, R_3. If the object points are not quite at the same scale as the design data then a scale factor may also be required to make the fit as exact as possible.

Axyz uses design values located in a *reference file*. The set of measured points selected for calculating the transformation must contain at least 3 points which form a triangle. It is not necessary to have full coordinate information at the reference points. The transformation calculates 3 shifts, 3 rotational values and an optional scale factor, making 6 or 7 elements in all. Corresponding to this there must be at least 6 or 7 pieces of reference coordinate information. This could, for example, be

- Full XYZ information at R_1
- YZ information at R_2 (or full XYZ information if a scale factor is calculated)
- Z information at R_3

If minimum reference data is used, see the additional comments in "3-2-1 Transformation" on page 83. Normally more than the minimum 7 reference elements will be specified by using full coordinate information at more than 3 points. Since the measured points will not conform exactly to the design shape, due to manufacturing tolerances and measurement uncertainties, the transformation is calculated by a best fit using least squares.

The algorithm provides options to hold the transformation parameters fixed at values defined by the user. A typical option, for example, is to request no change of scale by fixing the scale factor = 1.

Use of FIXED, NOT FIXED, WEIGHTED

The *Axyz* transformation program allows a degree of uncertainty in both the transformed points and the reference points. This permits a high degree of flexibility in tailoring the solution to operate in different ways and to accommodate different measurement configurations.

Transformed points are typically measured points which are recorded with a covariance matrix. This matrix defines the measurement quality of the point. The square root of the diagonal elements (variances) are the standard deviations of the point with respect to the coordinate system axes. Reference data can also be defined with a measurement quality and each individual coordinate can be assigned a standard deviation. (This can be used to create a covariance matrix although there is no provision for defining correlations between coordinate values and hence covariances are automatically zero.)

This facility enables transformed and reference points to be assigned a weighting and in the solution each will develop small residuals. However it would be common to ensure that the reference points have very small standard deviations and hence very high weights so that in practice their calculated values are almost identical to their nominal values.

It has also been explained how reference points may sometimes only be partially known. For example, a reference point may only define a Z coordinate and be equivalent to stating that a particular point lies on a specific XY plane. In this case there is no need for X and Y values. However, a partially known coordinate can be regarded as a known coordinate of very low quality.

Finally it is sometimes convenient to treat all transformed points equally, by assigning them unit weights and ignoring any covariance values or standard deviations associated with them.

In order to accommodate the different features described above, weight flags are defined for reference points.

FIXED

If a reference coordinate is defined as fixed, the user is stating that the value is either very accurately known or is a nominal (design) value. Internally a very small standard deviation is assigned to the value which will ensure that it has a very high weight when a transformation is calculated. This will ensure that the values do not change significantly during the solution. In effect they are treated as known values.

NOT FIXED

If a reference coordinate is defined as not fixed, the user is stating that the coordinate value is unknown. Internally a very large standard deviation is assigned to the coordinate which gives it a very low weight when a transformation is calculated. This will permit the values to change significantly. In effect they are treated as unknowns.

Note
If points are used as setup points in the transformation and their reference coordinates are only partially known, then the unknown elements must be given reasonable estimates or the algorithm may fail.

WEIGHTED

If a reference coordinate is defined as weighted, the user is stating that the coordinate is accurately known but there is some degree of uncertainty which is significant compared to the uncertainty in the points to be transformed. For example, the reference values might not be defined by a CAD system or blueprint but may have been measured by some other system. In this case the user will also define a standard error to express this degree of uncertainty.

Note
If the reference standard errors are much smaller than the standard errors of the transformed values, it may make more sense to define them as FIXED.
If the reference standard errors are much larger than the standard errors of the transformed values then you should carefully consider if it makes sense to use these values as reference values.

Flags for unit weighting

If unit weighting has been selected then all points to be transformed onto reference will be treated equally and given a weight value of 1. Their covariance matrices will be ignored in this case.

If the corresponding reference values are defined when unit weighting is active, only the flags FIXED and NOT FIXED are available. This allows for the possibility that a reference value may not exist (NOT FIXED) but if it does exist (FIXED) then the reference value is given a very high weight (much larger than 1) which ensures that it does not change in the solution.

Suppose however that you have switched between unit weighting and weighting by standard deviation, and some reference values already have the WEIGHTED flag . In this case the reference value will be treated as FIXED and given a very high weight. This may not produce the effect you want so it is always advisable to check the flags before computing the transformation.

5.7.5 3-2-1 Transformation

A 3:2:1 transformation is a normal 3D coordinate transformation which uses minimum reference elements, i.e. reference points whose coordinates are only partially known.

Although full reference coordinate information is not required for calculating the transformation parameters, good approximate values are needed for any unknown elements at 3 points defining less than 9 reference values. There are two reasons for this:

1. The algorithm requires 3 setup points with full coordinate information in order to calculate approximate transformation parameters. In such cases the *Axyz* software will <u>assume</u> that any unknown coordinates in the setup points have been provided with values which are good enough to make the approximate calculation. If this is not the case the approximation may be bad and the transformation may not work.

2. If the algorithm worked directly with the minimum information, more than one solution would be possible. See "Ambiguities in transformation" on page 84. However if reasonable approximations are available for the missing elements this effectively forces the choice of only one solution.

The use of minimum reference information (6 or 7 coordinate values) is very similar to axis alignment and produces an exact fit at the reference values.

Assisting the solution

You may have difficulties in getting a 3-2-1 transformation to work. If the solution does not actually fail, it may appear to be slow or it may appear to converge to an inaccurate solution. The following techniques may help:
* Improve the estimates of the unknown coordinate values
* Add more points in case the minimum data contains an error
* Use an axis alignment to generate initial transformation parameters
* Increase the number of iterations in the solution to between 5 and 10. (Although more iterations takes time, only the minimum amount of data is processed.)
* Set the percentage convergence change to a smaller value. A slowly converging solution may stop too early otherwise.

Ambiguities in transformation

When minimum data is used to calculate the transformation, more than one solution is theoretically possible. This could, for example, cause an object to be transformed into the upside down position. As a result, the unknown reference values must be approximately known in order to force the solution to find the correct position.

The first example shows a reference Z plane. Measured points Pa, Pb, Pc are transformed onto reference points Ra, Rb, Rc. The reference data is defined as follows:

* Only Ra is fully known in X,Y,Z.
* Rb has only a reference Z value which is the same as Ra, i.e it lies in the same Z plane as Ra but its XY values are unknown.
* Rc has only a reference XY position in the plane, i.e. its height above or below the Z plane is unknown.

Clearly point Pc could also transform to Rc' which is as far below the plane as Rc is above it. See example 1.

Yet another alternative is shown in example 2. Here the transformed triangle can be rotated about line Ra/Rc to move point Pb from reference

position Rb to Rb'. Both Rb and Rb' have the same Z value so that the conditions of transformation are still met.

If the approximate XY position of Rb and the approximate height of Rc were known, this would make it easy to choose which of the possible positions is the correct one.

These examples are special cases, chosen because they are easy to visualize. In general, however, several solutions are always possible. In order to resolve the ambiguities, unknown coordinate elements must be approximately known.

Example 1: Two possible positions
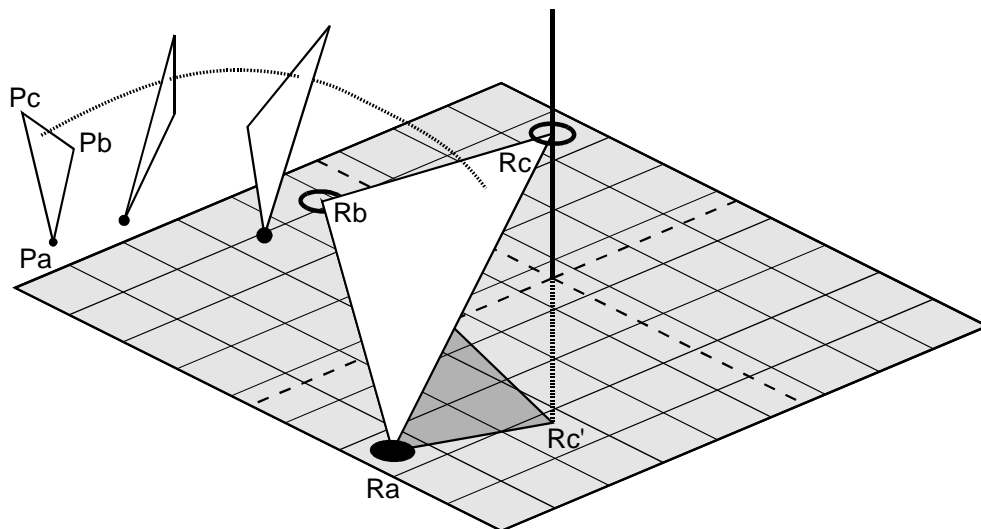
TRF-321a.WMF

Example 2: Another possible position
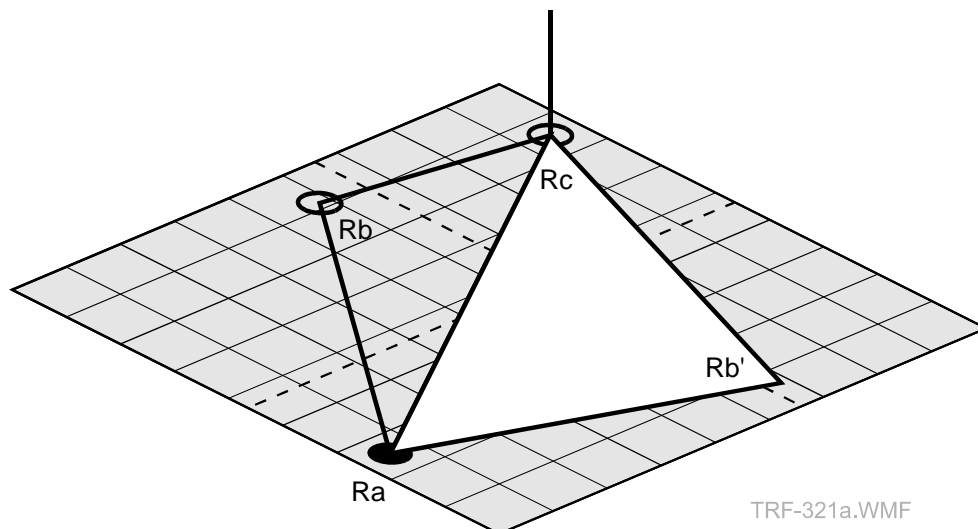
TRF-321a.WMF

5.7.6 Coordinate systems defined by shape fits

Every standard shape in the **Axyz** shape fitting routines has a coordinate system associated with it. It is therefore also possible to view coordinates in the coordinate system of any shape calculated by **Axyz**.

Note.
Since the definition of a coordinate system is effectively a sub-set of the definition required for a shape, **Axyz** stores coordinate systems under shapes although the Internal Data Manager displays them separately.

5.7.7 ECDS "local to object" transformation

Axyz classifies object reference information according to its purpose and handles it slightly differently. *Control points* are used specifically to control orientations *Reference files* are used to calculate 3D *transformations* and provide the reference data for *building* and *inspecting* points.

In contrast, ECDS has a single type of object reference information known as "control points" These are used for:
- Control of orientation
- 3D transformation

In ECDS only two coordinate systems can exist. The "local" system is created by relative orientation and is defined by the local axes of one of the theodolites. The "object" system is defined by "control" coordinates on the measured object. The object system can be created directly in the orientation process by inclusion of the control coordinates. Alternatively a local system can first be created. Then a least-squares, best-fit 3D transformation can be used to transform coordinates from the local to the object coordinate system.

To generate trial transformation values, all 3 coordinates of at least 3 points must be approximately known in both systems, even if the minimum 7 fixed elements are to be used.

The following flags can be assigned to the individual control coordinates:
- FIX (fixed)
- UNK (unknown)
- APX (approximate)

FIX flag - fixed control coordinate

The value of this coordinate is known and should not change. This is achieved internally by treating it as a variable quantity with a very high weight.

This corresponds to *Axyz* flag FIXED.

UNK flag - unknown control coordinate

Used for partial "control" points in which not all coordinates are known. For example a certain control point may only define a Z coordinate. In this case, X and Y are defined as UNKNOWN.

This corresponds to *Axyz* flag NOT FIXED.

APX flag - approximate control coordinate

This may be required to ensure a there is a minimum of 9 coordinate values (full information at 3 points) which the program requires in order to estimate the transformation parameters. From the users point of view the coordinate is unknown but a good estimate of its value is available. However, once the information has been used to estimate the transformation parameters, coordinates flagged as APX are subsequently treated as UKN (unknown).

There is no corresponding *Axyz* flag. If *Axyz* needs approximate data it simply assumes than the current coordinate values are reasonable estimates and a specific flag is not required. The *Axyz* flag WEIGHTED has no corresponding value in ECDS since ECDS does not allow for variation in quality of reference values.

ECDS 3:2:1 and 3:2:2 transformation

The ECDS techniques based on the use of minimum control can also be achieved by providing equivalent minimum control to the "Transformation" function in *Axyz*. See:
"Transformation onto reference coordinates" on page 79.
"3-2-1 Transformation" on page 83.

Alternatively a similar effect can be achieved using the "Axis Alignment" function in *Axyz.* See "Axis alignment" on page 72.

5.7.8 Transformation parameters

A transformation provides a new origin point and new axial directions. The purpose of the transformation is to take coordinates known in one system, typically the base coordinate system, and display them with respect to the new origin and axes.

The transformation makes use of the following parameters:
1) Location of the new origin with respect to the current coordinate system
2) 3 rotations from the current system into the new orientation
3) A scale change

The 3 rotations are implemented as a rotation matrix **R**.

Let new origin coordinates = (xn,yn,zn)
Let scale change = s
For current coordinates (x,y,z), corresponding transformed coordinates (xt,yt,zt) are then given by:

$$
\begin{pmatrix} xt \\ yt \\ zt \end{pmatrix} = s \cdot \mathbf{R} \cdot \begin{pmatrix} x - xn \\ y - yn \\ z - zn \end{pmatrix}
$$

Rotation matrix R

The rotational parameters ω, φ, κ, are given about the current x, y and z axes. Each produces an individual transformation matrix as follows:

$$
\mathbf{Rx}(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega) & \sin(\omega) \\ 0 & -\sin(\omega) & \cos(\omega) \end{pmatrix}
$$

$$
\mathbf{Ry}(\phi) = \begin{pmatrix} \cos(\phi) & 0 & -\sin(\phi) \\ 0 & 1 & 0 \\ \sin(\phi) & 0 & \cos(\phi) \end{pmatrix}
$$

$$
\mathbf{Rz}(\kappa) = \begin{pmatrix} \cos(\kappa) & \sin(\kappa) & 0 \\ -\sin(\kappa) & \cos(\kappa) & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

They are combined into a single matrix **R = Rz * Ry * Rx** where the
elements of **R** are given by:

r11 = cos (κ) * cos (φ)
r12 = sin (κ) * cos (ω) + cos (κ) * sin (φ) * sin (ω)
r13 = sin (κ) * sin (ω) - cos (κ) * sin (φ) * cos (ω)

r21 = - sin (κ) * cos (φ)
r22 = cos (κ) * cos (ω) - sin (κ) * sin (φ) * sin (ω)
r23 = cos (κ) * sin (ω) + sin (κ) * sin (φ) * cos (ω)

r31 = sin (φ)
r32 = - cos (φ) * sin (ω)
r33 = cos (φ) * cos (ω)

5.8 Types of coordinates

Coordinate systems are normally defined by 3 mutually perpendicular axes
through an origin. The axes are conventionally labelled X,Y,Z in a right
handed sequence although alternative labels are used in certain industries.
With respect to these coordinate systems a point is usually located by its
distance from the origin along each axis. The resulting triplet of numbers
(x,y,z) is known as a point's *Cartesian coordinates* (named after the French
mathematician René Descartes) or *rectangular coordinates*.

However this is not the only way of expressing coordinate values and
alternative ways may be more convenient. For example, when measuring a
cylindrical object such as a storage tank it may be useful to define a point
by its position along the cylinder's axis, the perpendicular distance out
from the axis and angle around the axis from some defined zero direction.

The right handed rectangular system is used by **Axyz** as the starting point
for conversion to other coordinate types, since the **Axyz** base system is of
this type.

5.8.1 Rectangular coordinates

3D rectangular coordinates are defined by 3 mutually perpendicular axes
(each is at right angles to the others). The axes are often called X, Y and Z
and a set of coordinates is conventionally given in the order (X,Y,Z).
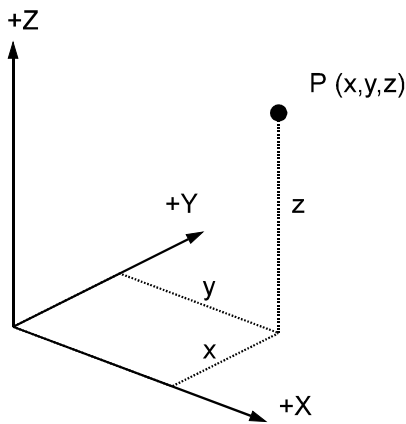
Right-handed Cartesian (rectangular)   Left-handed Cartesian (rectangular)
AXREC-R.WMF   AXREC-L.WMF

Right handed axes follow the right handed convention for the first, second and third axes.

Left handed axes follow the left handed convention for the first, second and third axes. When ***Axyz*** changes from a right handed to a left-handed system, the signs of the X coordinates change but Y and Z values remain the same.

5.8.2 Cylindrical coordinates

In ***Axyz*** cylindrical coordinates are derived from a right handed rectangular system. In a cylindrical system the X and Y values are expressed in terms of a radial (distance) offset from the Z axis and a horizontal angle of rotation. The Z coordinate remains the same.

This is a convenient system for measuring cylindrical objects where the cylinder axis corresponds to the Z axis of the rectangular system. If the cylinder is standing upright, the Z axis corresponds to a height measurement, which is the alternative name implied in the diagram. Angles are then horizontal angles of rotation around this axis.

A set of coordinates is conventionally given in the order (radius, horizontal angle, height).

In two dimensions only the angle and radial distance of the point from the origin are required. In this case coordinates are conventionally known as *polar coordinates* (but see also Spherical)
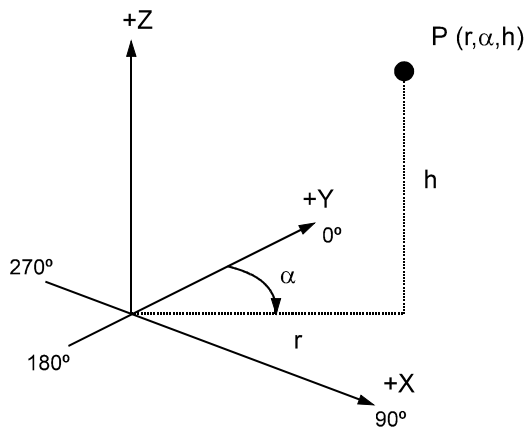
Cylindrical clockwise   Cylindrical counter clockwise
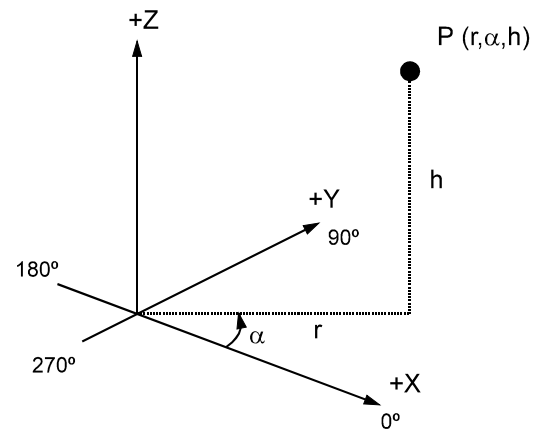
In a clockwise system horizontal angles are measured starting at zero on the Y axis and increasing towards X which is at 90°. This is a clockwise increase when viewed from above.

In a counterclockwise (anticlockwise) system horizontal angles are measured starting at zero on the X axis and increasing towards Y which is at 90°. This is a counterclockwise increase when viewed from above.

Note
It can be useful to define a cylindrical coordinate system where you have a number of points which should be approximately the same distance from a particular axis. The offsets from the axis should then show up as similar radial coordinate elements.

5.8.3 Spherical coordinates

In *Axyz* spherical coordinates are derived from a right handed rectangular system. In a spherical system a point is located by a distance and two angles rather than the 3 coordinate values along the rectangular axes. For axes labelled XYZ, and Z regarded as vertical, the point is located by its distance from the origin, horizontal angle in the XY plane and zenith angle measured from the Z axis.

A set of coordinates is conventionally given in the order (distance, horizontal angle, zenith angle).

Coordinates in this format are directly produced by instruments such as Total Stations and laser trackers. These measure the distance to a point and

the corresponding horizontal and vertical angles. Spherical coordinates may sometimes be confused with polar coordinates which are the two components (distance, horizontal angle) recorded in the 2D case (see also "Cylindrical coordinates" on page 90). This is because the descriptive term *polar* is often applied to these types of instruments.

The naming of the second angle in a spherical system is also potentially confusing. It is mathematically known as a *zenith angle* whose zero direction is vertically up. It is **not** the angle of elevation or depression with respect to the XY plane, both of which are also called *vertical angles* (positive if elevation, negative if depression). However the encoder on a polar instrument which measures this value is known as the *vertical encoder* or *vertical circle* and its output is sometimes called the vertical angle even if it is actually a zenith angle, as is the case with all Leica instruments.
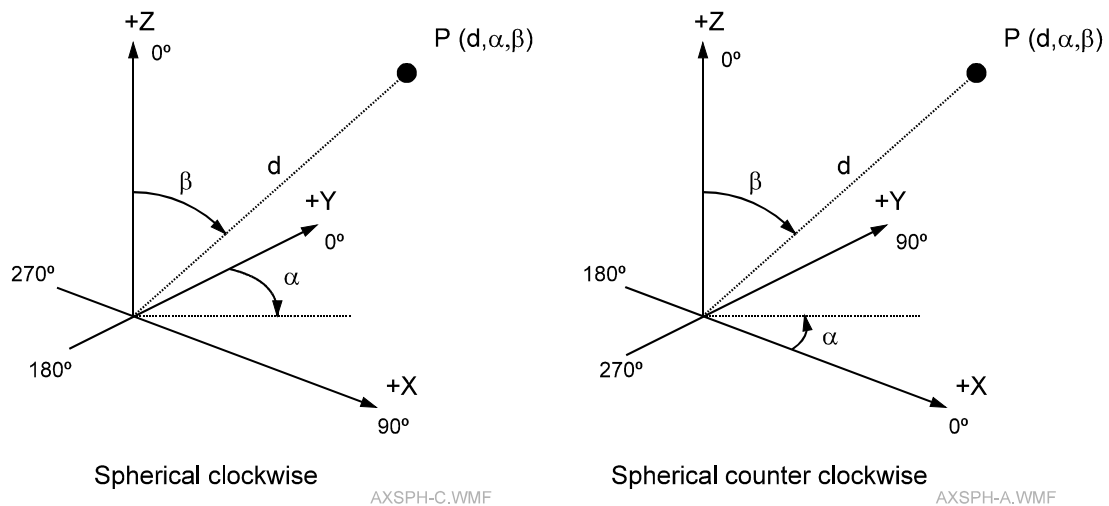
Spherical clockwise   Spherical counter clockwise

In a clockwise system horizontal angles are measured starting at zero on the Y axis and increasing towards X which is at 90°. This is a clockwise increase when viewed from above.

In a counterclockwise (anticlockwise) system horizontal angles are measured starting at zero on the X axis and increasing towards Y which is at 90°. This is a counterclockwise increase when viewed from above.

> **Note**
> It can be useful to define a spherical coordinate system where you have a number of points which should be approximately the same distance from a particular point. These should then show up with similar distance coordinate elements.

5.8.4 Theodolite and Total Station readings (spherical coordinates)

There is a lack of standardization regarding coordinate elements, the way they are specified and the way they are grouped. This is not simply an issue for Leica. For example, standard mathematical conventions are slightly different from those used in surveying and navigation.

Leica's theodolites and Total Stations generate horizontal angles, zenith angles and distances. A reading from either instrument type will generate a measurement group which is very similar to a set of spherical coordinates. However they may be presented in a slightly different way.

Conventionally, readings from a theodolite are given in the order (horizontal angle, zenith angle). By extension, readings from a Total Station simply add the distance to this group, i.e. the readings are given in the order (horizontal angle, zenith angle, distance).

To keep both types consistent, *Axyz* adds a dummy zero distance value to theodolite readings, so that each produces the set (horizontal angle, zenith angle, 0.0).

The horizontal angle is measured positive clockwise, in the same way as a bearing in navigation. In industrial metrology the local axes of a theodolite or Total Station are treated as right handed. They therefore follow the convention:

- y axis on horizontal zero
- x axis on horizontal 90°
- z axis points vertically up when the xy plane is horizontal (i.e. when the instrument is levelled).

The origin of the axes is at the centre of rotation.

This is identical with the definition for a coordinate system defined as "spherical clockwise" **except** for the order of coordinate elements.

Instrument reading: (horizontal angle, zenith angle, distance)
Spherical coordinate: (distance, horizontal angle, zenith angle)

In practice, instrument readings are presented by *Axyz* using a grid in the style of a spreadsheet. If wished, users can re-arrange the order of the columns to match the pattern of a spherical coordinate set.

6. Orientation Module

6.1 Orientation Module in brief

The Orientation Module takes all the current measurements, or a sub-set edited by the user, and mathematically builds a *measurement network* of stations from which further object points can be measured. This network is located in a common coordinate system known as the *base coordinate system*. All object coordinate data is stored in this system although they can subsequently be viewed and output in any other coordinate system defined by the user.

The Orientation Module operates in two distinct stages.
1. Initial and approximate estimation of station parameters (location and tilt), orientation point positions and scale bar locations using standard *orientation methods*
2. Final and optimized estimation of these values, also known as a *bundle adjustment*.

6.1.1 Orientation methods, measurements and points

The initial estimation of values makes use of *orientation measurements* and basic *orientation methods* such as *collimation* or *ECDS object orientation*. It is up to the user to ensure that there are sufficient orientation measurements to enable the network to be approximately constructed, although the module will report difficulties if this is not the case.

Some orientation measurements are only approximate and not used in the final optimization stage, the bundle adjustment. Again, therefore, the user must ensure that there are sufficient properly configured point measurements to ensure a solution, although a check is made that there are sufficient accurate measurements to calculate, in principle, the unknown orientation parameters and any associated point locations.

Sometimes points may be measured purely to ensure that an orientation is successful or to improve its accuracy. These are not necessarily part of the measured object and may be called *orientation points*. However normal object points may be sufficiently good for orientation purposes. There is no distinction in the database between "orientation" and "normal" points.

6.1.2 Optimization by bundle adjustment

The bundle adjustment is a least squares optimization. It takes the "bundles" of instrument pointings and makes successive "adjustments" to the network parameters until there is a best fit between the mathematical model of the network and the actual measurements.

6.1.3 Changes as network is updated

The Orientation Module can be used repeatedly as a measurement network is extended with further stations and measurements. Every time its results are accepted the base coordinate system may be re-defined and existing base coordinates may be modified.

6.2 Building a measurement network

In order to measure an *object* you need a *measurement network*. This is the arrangement of instrument stations from where most of the object points are measured and located in 3D. In order to obtain meaningful results, the instrument stations, object points and any other relevant targets and ancillary equipment must be located in a common *frame of reference* or *coordinate system*. The task here is therefore the mathematical construction of the measurement network in a single coordinate system. This is known as the *base coordinate system* and the task is carried out by the *Orientation Module.*

6.2.1 Building a network for on-line measurement

The basic stages in measuring an object on line are these:
1) Construct a network. This itself is a 2-stage process
- Find approximate initial values for the locations of stations and ancillary devices and targets
- Optimize this approximate information to obtain an accurate picture of the network
2) Measure the required object point coordinates on line.

A large object may have to be measured in parts, so the sequence may be repeated as:
   1a) Construct part of the network
   1b) Measure part of the object on line

   2a) Add another section of network
   2b) Measure another part of the object on line

6.2.2 Single step network and object measurement (off-line analysis)

An alternative to constructing an optimized network from which object points are measured, a single stage approach is possible. With this technique you measure all the required object points at the same time as any measurements needed to construct the network, and then compute everything simultaneously.

In contrast to the 2-stage approach, all station and point locations are optimized in one procedure. In the 2-stage approach two completely separate optimizing procedures are active - one mainly for the stations (some object points included) and one for individual points measured on line.

The alternative, single stage approach is often used in photogrammetry and videogrammetry. It is effectively an off-line calculation of object points.

Example of single stage object measurement

- Place targets on a radar dish
- Take a dozen images (conventional or electronic) from different locations
- Process the images to obtain camera locations and target coordinates in a common frame of reference
- Do any further off line processing, e.g. fit a paraboloid to the target points

6.2.3 Single step or 2-stage method in Axyz?

Axyz users generally require on-line coordinates, so many features of ***Axyz*** accommodate the two stage process. However, the alternative technique will still work. You could take a single theodolite or Total Station, move it to different locations, make all possible pointings at each location and then process all measurements off line.

The single stage process is really equivalent to first stage in the 2-stage process. To construct a network you must make measurements to some targets in the object space. All that happens in a single stage process is that extra points, i.e. all the object points, are added into this set of measurements.

6.2.4 Orientation methods

When constructing the network *Axyz* makes use of "network building blocks". These provide different ways of positioning one station with respect to the existing stations in the network or directly into the coordinate system which has been established. These building blocks are known as *orientation methods*. They enable the initial and approximate *orientation parameters* for each station to be calculated.

A station requires 6 orientation parameters to locate it with respect to another station or a frame of reference. These are the 3 coordinates of its position and 3 rotational or tilt parameters. The rotational parameters express its angular attitude, just as the angles of roll, pitch and yaw tell you what angle an aircraft has with respect to the ground.

Targets are involved in the orientation procedures and they also require initial coordinate values in the common coordinate system. They may be *scale bar targets*, *control points* with known coordinates or just additional *orientation targets* with unknown coordinates whose sole purpose is to ensure that the network has a strong geometry and can deliver the required accuracy. Control points and orientation targets may also be points on the measured object.

Orientation methods depend on the underlying technique for locating points which in *Axyz* can be one of two types:
- Polar measurement
- Triangulation

These concepts are developed more fully in the following sub-sections and further details are given in sections:
- "Initial orientation and target location" on page 112.
- "Optimized orientation: the bundle adjustment" on page 121.

6.3 Polar measurement

Polar measurement records the angular pointing and distance from the measuring instrument to the target point. Total Stations and laser trackers work in this way. Polar measurements are equivalent to spatial coordinates in a spherical coordinate system and are easily converted into orthogonal 3D values.

With this technique full 3D measurement is possible from a single instrument. However it is common to require more than one polar measuring station in order to fully cover all points of interest on an object. This can be achieved by using several instruments or one instrument moved to several locations. Individual stations must therefore be linked together into a common network in order to obtain object coordinates in a single coordinate system.

6.3.1 Diagram: Principle of polar measurement
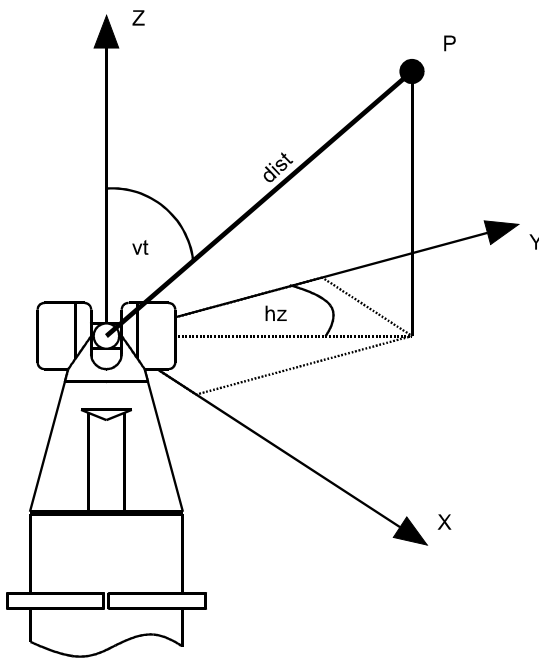
Target location by polar coordinates

POLARFIX.WMF

6.3.2 Orientation with polar measurements

This linking or orientation procedure can be achieved using a simple 2D transformation if stations are levelled.

The example shows a levelled station being oriented into an existing levelled coordinate system by a 2D transformation. If not all stations are levelled, or control points in a tilted system are used to link the station, then at least 3 existing points must be measured. This technique is mathematically identical to a 3D coordinate transformation.

Although this technique can produce good results it can still be improved by optimizing with the bundle adjustment. Optimization will also deal with

redundant and *overdetermined* cases where, for example, more than one polar measurement is made to a particular point.

Example: Orientation by polar measurements

Existing measurements

Orient new polar station by transformation using measurements to existing points
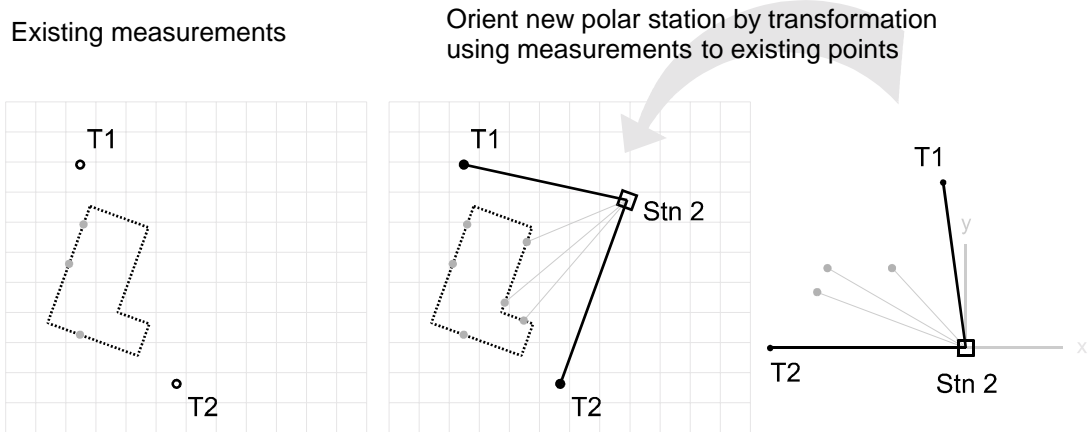
OR-POL-a.WMF

6.4 Triangulation

Theodolites, which measure angle only, employ the principle of triangulation to locate points in 3D space. The diagram shows how the technique takes intersecting rays from at least 2 locations in order to generate 3D target coordinates.

6.4.1 Diagram: Principle of triangulation

Principle of triangulation

TRIGFIX1.WMF

6.4.2 Diagram: Analysis of triangulation

Horizontal plane   Vertical planes

TRIGFIX3.WMF

6.4.3 Simple analysis of triangulation

The diagram showing the principle of triangulation represents a very simplified form of triangulation in which both instruments are levelled and measure horizontal angles from the common connecting line between them.

Evaluation of the triangles gives the following:

Assuming the horizontal base length b between the instruments is known, the following information is easily derived:

The horizontal triangle is fully defined by the base length b and measured horizontal angles αA and αB. The horizontal ranges rA and rB can be calculated and the horizontal position of the target point, Nh, found.

In the left hand vertical triangle rA is now known, as is the measured vertical angle βA. This is sufficient to calculate the target height hA.

In fact the target height can also be computed from the right hand vertical triangle in a similar way, assuming that the instrument height difference h is known. This gives a second calculation for target height, hB.

This simple case demonstrates redundancy and the need for an optimized solution. Clearly it is also necessary to know the instrument separation and the fact that both are levelled. In general, the orientation of one with respect to the other is required.

6.4.4 Simple orientation methods in triangulation

Here are some simple techniques which can provide fairly accurate orientation.

Diagram: Accurate collimation

Existing station measures scale points and collimation pointing
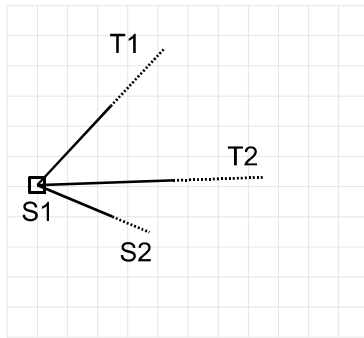
Orient new station by aligning collimation pointings and fitting to scale points
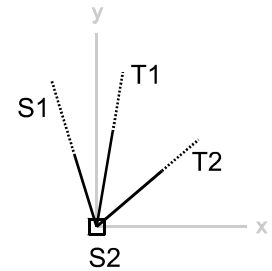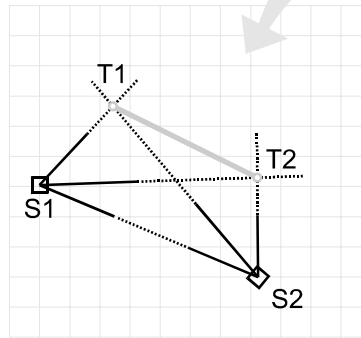
OR-COL-a.WMF

Accurate collimation

Collimation means the sighting of one instruments rotation centre from another, i.e. the base line between them is directly measured. This makes it particularly simple to line one up with the other.

The diagram shows the procedure with two levelled stations which have additionally measured a scale bar identified by the targets T1 and T2.

Diagram: Horizontal resection

Existing measurements
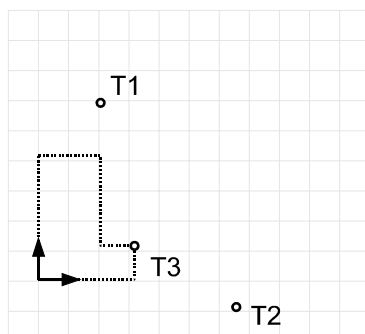
Orient new triangulation station by horizontal resection onto existing points
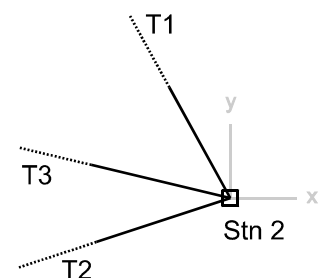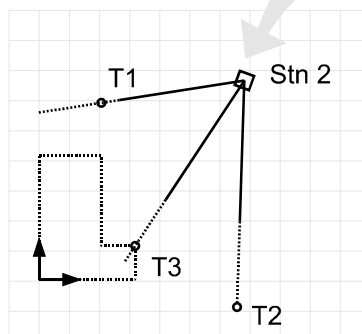
OR-HOR-a.WMF

Horizontal resection

Existing stations oriented, for example, by accurate collimation can create some target points in a levelled system. A new levelled station can *resect* its location with respect to these targets, which links it into the existing network.

Accurate orientation methods

Although reasonably accurate, these simple orientation procedures still benefit from optimization which in any case is required in order to deal with redundant information.

In some cases, particularly where stations are not levelled, additional orientation techniques are needed.

6.4.5 Scale requirements in pure triangulation

Since theodolites do not supply distance information, scale must be introduced in some other way.

Normally correct scale is introduced by sighting the ends of one or more *scale bars* or by including orientation points with correctly scaled coordinates. Known distances between two target points are then either directly available or can be derived.

Alternatively the inclusion of *control points* can be used instead of, or in addition to, scale bars. Control points have known coordinates which means that they inherently include scale information..

If scaling information is not available the orientation can be calculated at an *assumed scale* or *model scale,* for example by assigning the value "10 cm" to the distance between two instruments. An unscaled network can be built up until such time as scale information appears, at which point everything can be re-scaled to the true scale.

6.5 Control points

Control points are locations with accurately known coordinates. It is very common for control points to be object features which have known *design (CAD)* or *blueprint* coordinates.

Control points are included in orientation methods for two purposes.
- They force the results of the orientation into the coordinate system defined by the control, for example the CAD coordinate system used to manufacture the object. This is convenient for further data comparison.
- They influence or *control* the results of an orientation, i.e. they affect the shape of a measurement network. This can improve accuracy when the geometry of the measurement network is not ideal.

Control points can have various sources, for example:
- Selected object features such as critical drill holes.
- Specially designed and fixed targets in a reserved measurement area into which objects are brought for meas.
- Points on a reference frame or object inserted into the measurement area

When used to define a coordinate system it is not necessary to have full coordinate information at every control point. For example, a set of drill holes might only be required to lie in a particular XY plane. Only their Z values would be strictly used to define this aspect of the coordinate system. It is therefore possible to have *partial control points*.

The fact that control points influence an orientation can be advantageous when a measurement configuration, for good reasons, has some geometrical weakness. (Unless you have a good excuse for a weak setup you should design a better one!) Adding control points to a weak measurement network will help preserve accuracy in new measured positions.

Ideally control point coordinates should be known to a higher accuracy than they can be measured by **Axyz**.

Since control points have known coordinates they also supply scale information to a network, although it is not normally good practice to rely exclusively on control points for this purpose.

When control points influence a network's shape this means that there is an <u>excess</u> of control information.

6.6 Levelling constraints

Theodolites and Total Stations are normally accurately levelled before use. This makes their primary (standing) axes parallel to the vertical.

Some laser trackers can be referenced to the vertical by attaching a separate tilt sensor. In this case the standing axis is not set parallel to the vertical but its deviation from the vertical is accurately measured.

Many laser trackers are operated without a tilt sensor and so cannot be considered levelled. It is also possible to switch off a theodolite's tilt compensator and operate the instrument in a tilted position.

The Orientation Module will take account of all these situations. For example, a sub-group of levelled stations will have their standing axes forced to be parallel.

Where instruments have been levelled or referenced to the vertical this will impose an additional constraint on the network which may simplify the orientation procedure and improve the accuracy of the measuring network.

6.7 Optimized orientation (bundle adjustment)

The bundle adjustment brings together all the items discussed so far:
• Polar measurements (horizontal angle, vertical angle, distance)
• Triangulation measurements (horizontal angle, vertical angle,)
• Calibrated scale lengths (distance)
• Control points (coordinate value)
• Levelling constraints

The bundle adjustment must process this diverse range of equations and constraints to generate an optimized configuration of stations and targets. This configuration ensures that angular pointings and polar measurements meet as closely as possible at the corresponding target positions.

Unfortunately, the equations in a bundle adjustment are *non-linear*. This means that the optimized parameters which are computed must first be approximately known. Hence the need for mathematically simpler orientation techniques in order to "prime" the bundle adjustment with its starting values.

Some of these orientation techniques have already been mentioned. They may, in fact, generate very good starting values. In fact, you might consider them so good that further optimization is unnecessary. However they are not truly optimal and in any case other methods may be used which generate much less accurate starting values. In fact, even a rough manual estimate of relative positions by use of a tape measure may be sufficient to start the optimization.

Once the optimization starts, any approximate measurements used for the trial orientation will be discarded. It is important to ensure that the remaining measurements are sufficient to compute the optimized result.

6.7.1 The minimum measurements required

The minimum number and distribution of measurements needed for a successful calculation depend on the particular measurement configuration. In all cases only a relatively small number of measurements are required. The most difficult situation is also the most general case where two theodolites have not been accurately levelled and have only made sightings to common points. In this case a minimum of 5 common points must be measured in order to calculate the location and tilt of one instrument with respect to the other.

A mathematical argument can explain how a minimum of five measurements provide sufficient information to locate one theodolite in the coordinate system of the first, as well as finding the positions of the five unknown points.

One instrument arbitrarily defines the coordinate system. Its coordinates and rotational parameters are <u>given</u> the value zero. Targets and second instrument position are found with respect to this coordinate system.

Unknown quantities	Total
Each of the five targets has three unknown positional elements (X,Y,Z).	15
The second instrument has three positional unknowns and three rotational unknowns, e.g. roll, pitch and yaw as in an aircraft.	16
Total number of unknown quantities:	21

Equations	Total
Each pointing generates **two** angle equations (one horizontal, one vertical). There are 10 pointings, 5 from each instrument	20
The separation between two targets (scale bar) provides an additional distance equation	1
Total number of equations:	21

With as many equations as unknowns, an analytical solution for the unknowns can normally be found. *Failure cases* exist, for example, if all five targets lie on a straight line, but in practice these are easily avoided. Adding constraints to the situation will generally reduce the minimum number of measurements required. For example if both theodolites are levelled, a minimum of 3 common points will provide a solution.

6.7.2 Classification of orientation

Optimized orientations have a descriptive classification to indicate if control points were used or not. The following descriptions also emphasize the optimization methods in use.

Controlled (object) orientation

A controlled orientation makes use of control points which force the results into the coordinate system of the control and influences (controls) the shape of the measurement network.

It may be called an object orientation since the control coordinate system generally has some direct meaning to the object. For example it represents the coordinate system used in the design and manufacture of the object.

Relative (local) orientation

A relative orientation does not make use of control points and the coordinate system is initially arbitrarily defined by the first station processed, which is the lowest numbered station.

It may be called a local orientation because the local axes of a station define the final coordinate system which is "local" to the network itself.

Relative orientation with a balanced station network

This option produces a very similar result to a standard relative orientation but distributes the estimated coordinate errors in a more even-handed way.

The optimization starts with the origin defined by the first (lowest numbered) station processed. However as it proceeds the origin drifts away from this initial position.

Relative and absolute orientation (Photogrammetry)

Many of the concepts in *Axyz* are derived from photogrammetry where the following terms are common:

- Relative orientation
  An orientation of one instrument with respect to another
- Absolute orientation
  An orientation with respect to the coordinate system of an object

In photogrammetry, absolute orientation could imply either the use of control in the orientation process itself or the subsequent transformation of a relative system onto "control" coordinates. In *Axyz* control points are exclusively used in the orientation procedure and the transformation function is implemented as a transformation onto reference coordinates. There is therefore a clear separation between these two similar but different functions. Furthermore, absolute orientation implies orientation to the measured object's internal frame of reference but control points for *Axyz* could have a source other than the main object of interest, such as targets in a test field. For this reason *Axyz* makes use of the explicit term *orientation to control.*

6.8 Base origin and coordinates

The location of the coordinate axes and the values of the associated station and point coordinates depend on the optimizing technique used by the bundle adjustment. This defines the base coordinate system for all point measurements. There may be small or even large changes in this coordinate system when the orientation is re-calculated with new or different measurements, or a new optimizing method.

6.8.1 Axes on the object (controlled orientation)

If reference coordinates on the object, defined by *control points,* are included in the orientation procedure then the origin and axes which define the control coordinates are used.

This is a natural consequence of the fact that the measured control points must end up with coordinate values very close to their nominal values and this can only be achieved by forcing the measuring network into the same coordinate system.

6.8.2 Axes at the lowest numbered station (relative orientation)

In this case control points are <u>not</u> included in the orientation procedures.

Since a meaningful origin and axes are not automatically provided, the origin and axes of the base system are arbitrarily located at the lowest numbered station.

6.8.3 Axes near the lowest numbered station

If the option for a balanced station network is chosen, then a relative orientation is calculated with base axes initially defined by the local axes of the lowest numbered station.

As the optimization proceeds, the base axes drift away from this initial position.

6.8.4 How the base axes and coordinate values can change

The location of the base coordinate system, and the coordinate values of any existing points, can change by small or large amounts when the orientation is re-calculated.

Same optimization, new information

If an orientation is repeated with the same optimization technique but new information, typically another station is added to the network, then the base system will change. This happens because the optimization must accommodate different data than in the previous calculation and the results <u>must</u> be different although the changes should not be large.

To some extent this can be avoided by fixing existing stations at their previous locations. However, this is does not then provide a fully optimal

inclusion of a new station. Also, if a the new station has measured existing points, the coordinates of these points will still change because of the additional measurements to them.

Change from relative to relative plus balanced orientation

This change will cause the base axes to move from local axes at the lowest numbered station to some location nearby. Users have reported movements between 200μm and 60mm.

Change between relative and controlled orientation

This will cause a very significant switch of base system axes between the local axes of the lowest numbered station and the axes defining the coordinates of the control points.

The shape of the measurement network may also change slightly because of the influence of the control points.

6.9 Transformation onto reference coordinates

Instead of employing an orientation to control, there is an alternative way to use known design values to define a system of coordinates. This requires a *3D transformation*. For more information see "Transformation onto reference coordinates" on page 79.

With this method you first complete a relative orientation, i.e. an orientation which does not include control points. The relative coordinate system is then transformed onto the coordinate system defined by the design data using the 3D transformation. This design data is contained in a *reference file*.

Since the design data is not included in the orientation procedure it cannot influence the shape of the measurement network. Clearly this is <u>not</u> the same effect as the use of design data in the form of control points.

This technique is used to compare the effects of including or not including control points and may be helpful in tracking down problems associated with design data.

6.10 Good and bad geometry

In triangulation, the geometrical configuration of the network strongly influences the quality of point locations. This is particularly important for triangulation techniques. Particular requirements are that orientation measurements provide a strong network and that points are located with good intersection angles.

Diagram: Intersection geometry
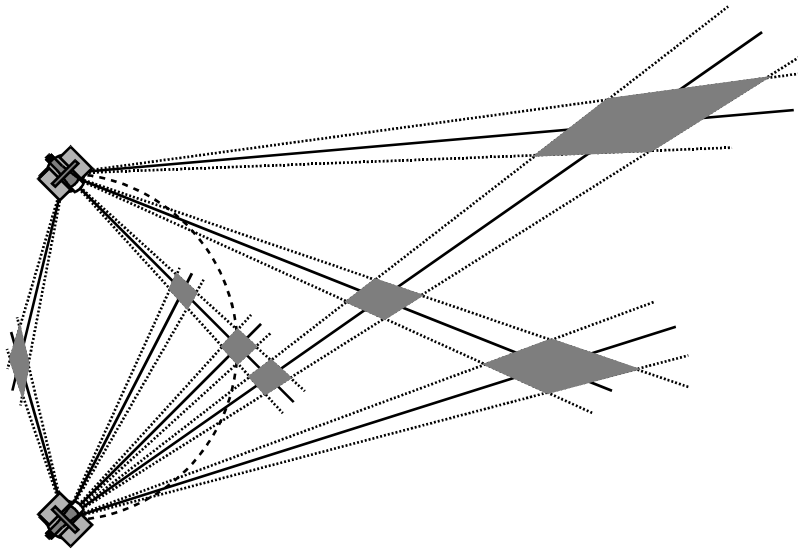
Effect of intersection angle on intersection error

INTSECER.WMF

Intersection geometry

In triangulation, points must be intersected by at least 2 rays. This will not pick up a bad pointing in the plane of the intersecting rays, since the rays will still successfully intersect. Only an error out of the plane can be detected in this case.

Multiple pointings (3 or more) are effective in detecting bad pointings and identifying the specific pointings which are in error. They are most effective when the rays are not all in the same plane. Placing stations at different heights will provide this condition.

The intersections themselves create minimum error when the intersection angle is close to 90°. The diagram shows how fixed angular tolerances give rise to different errors at the intersected points. As the intersection angle gets narrow, the 3D positional error gets larger.

Failure cases

Orientation methods require certain minimum conditions and if these are not met then an orientation will not, of course, be possible. However, even when the minimum conditions are apparently met, some geometrical configurations can still produce ambiguous results.

When a measurement network is constructed, angular pointings and polar measurements should come together very close to the corresponding targets. In the end, this is how you know that the network is correct and can be used for further target location. If you can achieve this condition for different instrument locations then you have a problem known as a *failure case*. Obviously these situations should be avoided, as should a situation which is close to a failure case.

Example: Connecting Total Stations
Failure case if connecting points on a vertical line.

If a new levelled Total Station is added into the network by measuring 3 existing points lying on a vertical line, the station will not define a unique position. The station can lie anywhere on a circle around the vertical line.

Example: Connecting theodolites in a relative orientation:
Failure case if connecting points all on a line.

If a new instrument is positioned with respect to an existing instrument by measuring 5 or more common points all lying on a straight line, the station will not define a unique position. The station can lie anywhere on a circle around the line of points.

7. Initial orientation and target location

7.1 Relative orientation methods

A number of relative orientation methods assume that instruments are levelled or approximately levelled. (Mathematically it is sufficient if their primary (standing) axes are approximately parallel, which is easily achieved when the instruments are levelled.)

This simplifies many of the algorithms whose purpose is to obtain approximate instrument locations and angular attitudes. This does not mean that measured objects must be levelled or that instruments must be oriented to gravity in the final bundle adjustment.

7.1.1 Collimation (accurate)
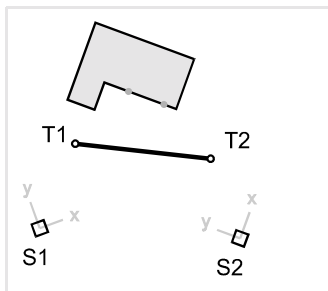
Situation on site
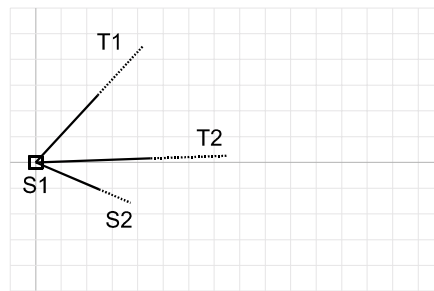
Station 1 defines the coordinate system.
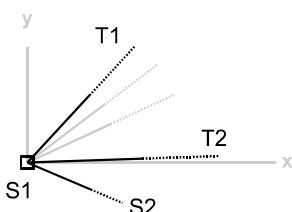
Measurements at station 1
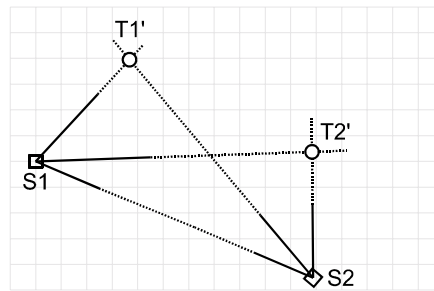
Align collimation pointings to locate station 2. Provisional scale.
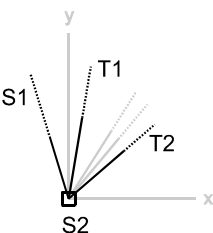
Measurements at station 2
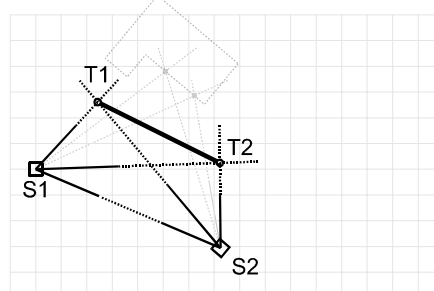
Re-set the scale to fit the scale bar length. Further points by intersection.

OR-COL-1.WMF

This calculation assumes that the instruments are approximately levelled.

Accurate collimation pointings are sufficiently good to be used in the optimized orientation.

7.1.2 Collimation (approximate)

This is essentially the same technique as accurate collimation except that the collimation pointings are only made approximately. These measurements are therefore not used in the optimized orientation.

This calculation assumes that the instruments are approximately levelled.

7.1.3 Collimation to two instruments

Situation on site
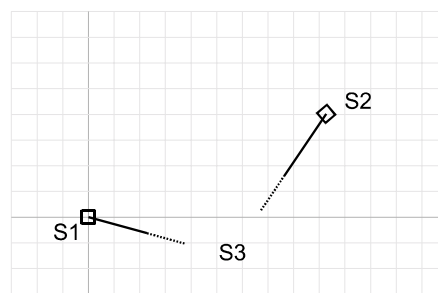
Stations 1 and 2 exist in a local system at 1.
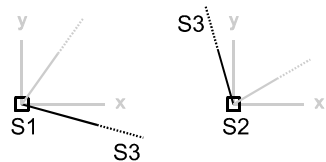
Measurements at stations 1 and 2
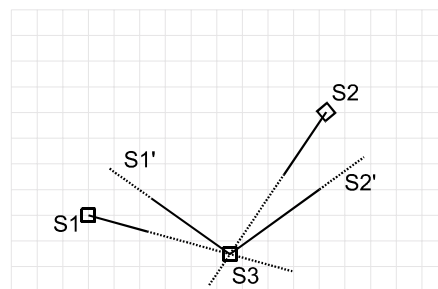
Locate S3 by intersection from S1 and S2.
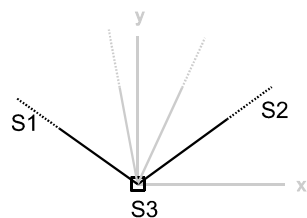
Measurements at station 3
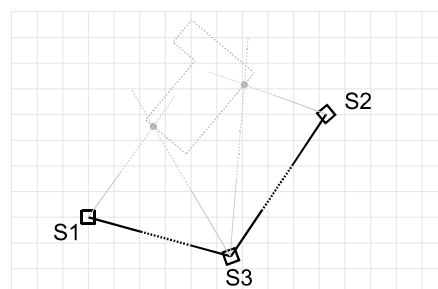
Rotate S3 to align the collimation pointings.
Further points by intersection.

OR-COL-2.WMF

This calculation assumes that the instruments are approximately levelled.

Any collimation measurements which are accurate will be further used in the optimized orientation.

7.1.4 Collimation with common point
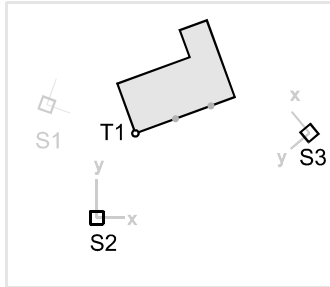
Situation on site
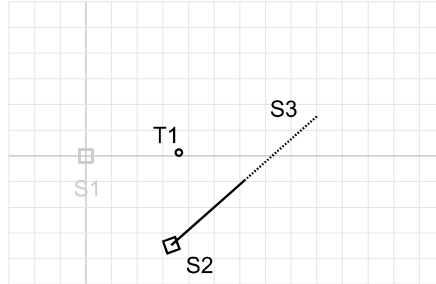
Station S2 and target T1 exist in a local coordinate system defined by station S1.
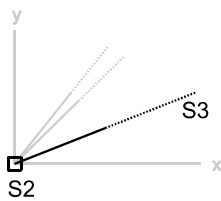
Measurements at station 2
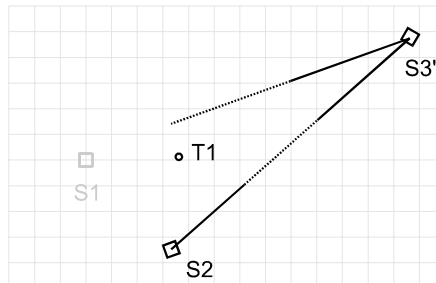
Align collimation pointings between 2 and 3.
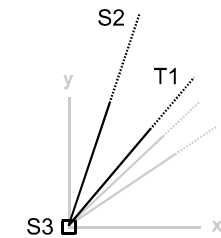
Measurements at station 3
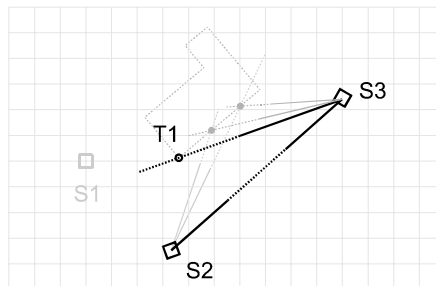
Adjust location of S3 to intersect target T1. Further points by intersection.

OR-COL-3.WMF

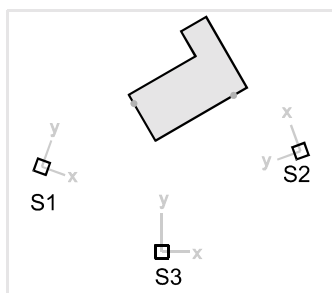This calculation assumes that the instruments are approximately levelled.

If the collimation measurement is accurate it is included in the optimized orientation.

The target measurement should be accurate and will be included in the optimized orientation.

7.1.5 Horizontal resection
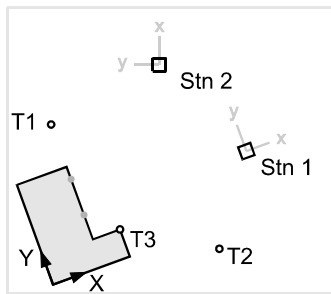
Situation on site
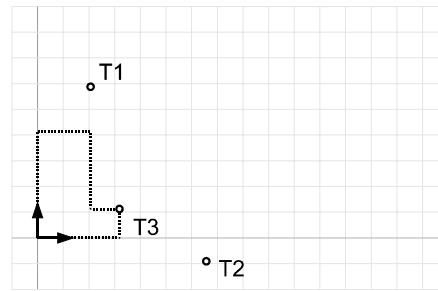
Existing points (shown in an object system).
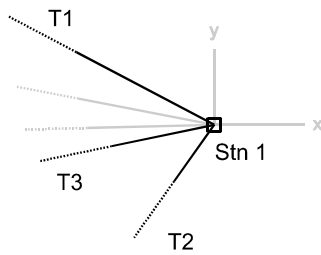
Measurements at station 1
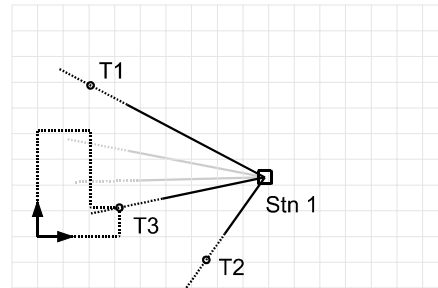
Station 1 located by horizontal resection.
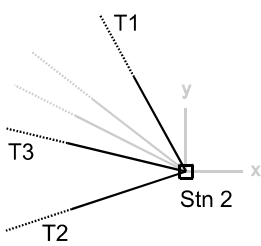
Measurements at station 2
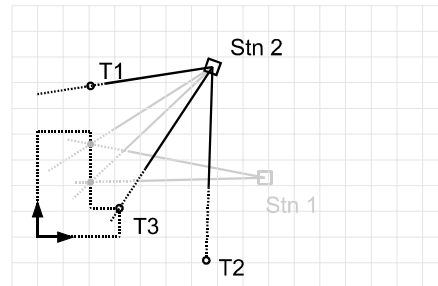
Station 2 located by horizontal resection.
(Further object points by intersection.)

OR-HOR-1.WMF

This calculation assumes that the network is approximately levelled by inclusion of approximately levelled instruments at one or more previously oriented stations. Any existing target coordinates are therefore located in a coordinate system referenced to gravity.

The target measurements are assumed to be accurate and are included in the optimized orientation.

7.1.6 Orientation of polar measuring instrument

This technique applies to Total Stations and laser trackers.

Simple orientation with levelled instruments
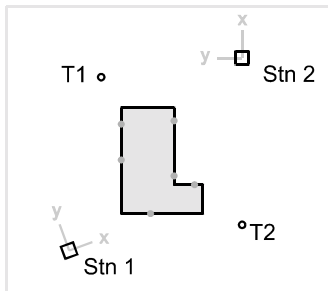
Situation on site
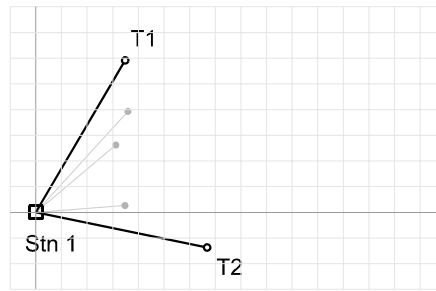
Station 1 defines the coordinate system.
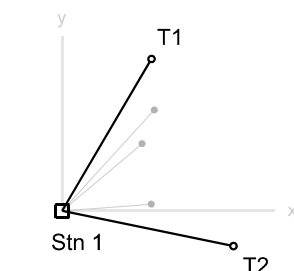
Measurements at station 1
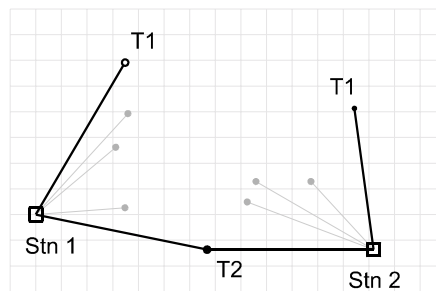
Station 2 provisionally attatched at T2.
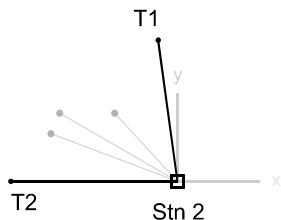
Measurements at station 2
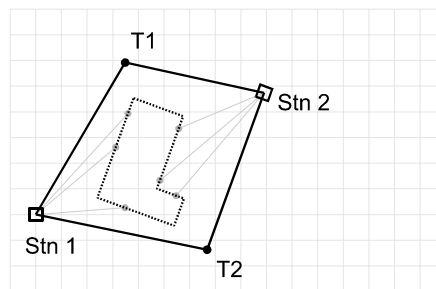
Rotation of station 2 to attach at T1.

OR-POL-1.WMF

The diagram shows the common situation of linking one polar station to another via two existing target points. This works when the network is approximately levelled.

The existing contact points can be measured by any other instruments in the network, or can be control points in a levelled reference system. The only general requirement is that the contact points and station must be approximately referenced to gravity.
The target measurements are assumed to be accurate and are included in the optimized orientation.

Orientation with non-levelled instruments

In this case 3 existing contact points must be measured and the station is linked into the existing network by the same method as used for a 3D transformation.

7.1.7 ECDS "local" orientation

This technique corresponds to the ECDS method of "local orientation". It is a relative orientation method which can be used in its original ECDS format. However some modifications have been made which can make the method more flexible.

Simple local orientation

Situation on site

Station S1 defines the coordinate system.

Measurements at station 1

S2 aligned with X pointing parallel to X axis.

Measurements at station 2

Measurements at station 3

S2 shifted to fit polar measurement to S1.
S3 fitted by same procedure as for S2.
Further points by intersection.

OR-LOC-1.WMF
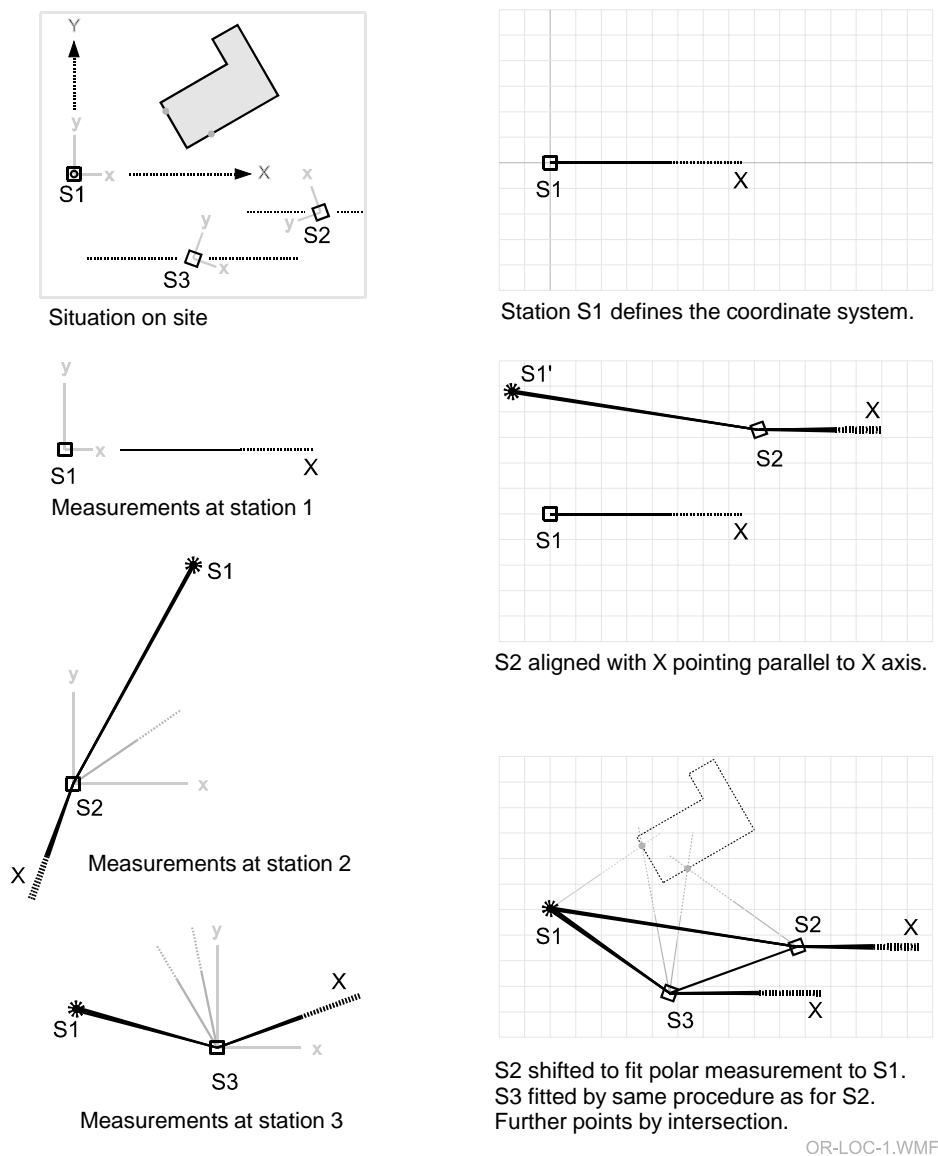
All stations are assumed to be approximately levelled.

One "key" station (station 1) defines the origin and local axes and measures the local X axis.

Further stations measure the X direction and the approximate polar location of the key station, i.e. by angular pointing (h,v) and manual estimation of distance (d).

None of the axial pointings or polar estimations are accurate and so none of these measurements are used in the optimized orientation.

Modifications to local orientation

It is not necessary for the polar measurements always to refer back to the key station. Any previously oriented station can be used.

Further stations need not be levelled. However in this case they must identify two of the local axes (X,Y or X,Z or Y,Z), not just one. This provides enough information to calculate the approximate tilt.

7.2 Controlled orientation methods

7.2.1 Orientation to control from a relative orientation

This technique assumes that some of the relative orientation methods have first created an oriented network. This complete network is then oriented to the control points by a 3D transformation.

7.2.2 ECDS "object" orientation

Simple object orientation
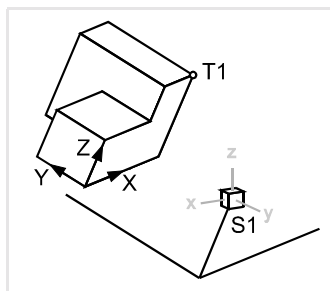
Situation on site
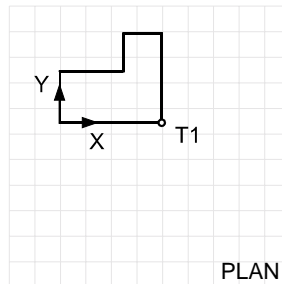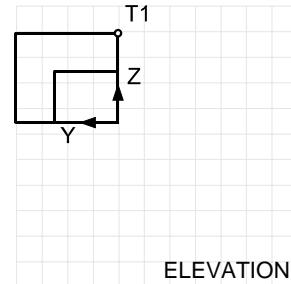
Control point in object system
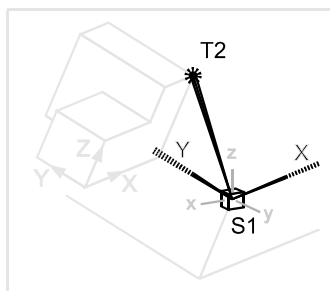
Site orientation measurements
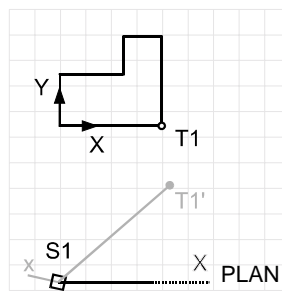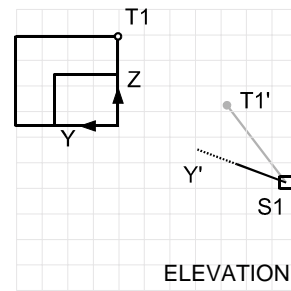
S1 aligned with X pointing parallel to X axis.
Y pointing not yet aligned.
Polar location of T1 from S1 not yet adjusted.
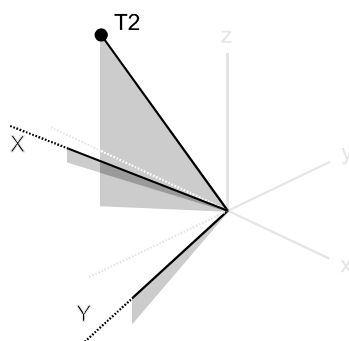
Detail of measurements
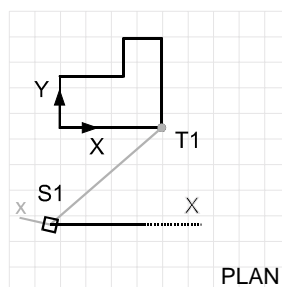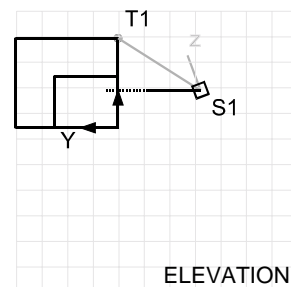
Rotate S1 to align Y pointing with Y axis.
Shift S1 to match polar location of T1.

OR-OBJ-1.WMF

No assumption is made about stations being levelled. They may or may not be levelled. The basic assumption is that the coordinate system of the control points may be tilted with respect to the local coordinate system of the station.

Stations are therefore approximately oriented to the object's control points by two measurement components.

Firstly a station approximately measures the directions of two of the object axes (X,Y or X,Z or Y,Z). This enables it to be given the correct angular tilt with respect to the object.

Secondly a station locates a control point on the object by approximate polar location, i.e. by angular pointing (h,v) and manual estimation of distance (d). This enables the station to be given the approximately correct position with respect to the object.

None of the axial pointings or polar estimations are accurate and so none of these measurements are used in the optimized orientation.

Modifications to object orientation

If a station's local z axis is parallel to the object's Z axis, only one object axis need be measured, X or Y. This situation occurs when the control points defining the object exist in a levelled system and the theodolite at the station is levelled.

Positioning a station by polar location can be done using any existing point which has object coordinates. The point need not be a control point.

8. Optimized orientation: the bundle adjustment

8.1 Bundle adjustment in brief

A bundle adjustment is the name given to a general least squares analysis package which takes the *bundles of pointings* from theodolites Total Stations or cameras to identified targets and processes them to create an optimal set of target coordinates which is a best fit to the angular and distance measurements made.

The adjustment simultaneously provides best estimates of the instrument or camera *orientation parameters*. These 6 parameters specify an instrument's 3D position and angular attitude (e.g. roll, pitch and yaw as in an aircraft).

Theodolite systems generally use the method to find orientation parameters. Further target coordinates are then found by intersection, which requires these parameters.

Photogrammetric systems often use the method to find the 3D coordinates of **all** targeted points on the object. In addition, a photogrammetric bundle adjustment often models the internal geometry of the camera, which frequently improves measurement quality. This is known as *self-calibration* because it does not require any external reference information as would be the case with conventional calibration techniques.

8.2 Mathematical details in brief

The equations used by a bundle adjustment are *non-linear*. In practical terms this means that it is not possible to compute instrument and target locations in a single computing step. In one step it is only possible to improve on an estimate of these positions by making suitable *adjustments* to the estimates. A technique of *iteration* is therefore required which is simply the process of repeating the calculation of improvements until no further improvement can be made.

Clearly the whole iteration sequence must be somehow initiated and other approximate methods are required which generate *initial values* or *trial values* of all the elements being computed. These approximate methods can include manual estimation of values, the use of design information from blueprints or CAD models or measurement techniques such as space resection.

To improve measurement quality and determine potential sources of error, it is common to use more measurements that the minimum which are theoretically necessary to generate the required data. When an excess of information is available, the solution is known as *overdetermined*. Not surprisingly, too little information results in an *underdetermined* set of equations which cannot be solved. Overdetermined data sets can find faults such as a mis-pointed theodolite and improve measurement quality by helping to "average out" the effects of random error.

Finally it is interesting to note that a bundle adjustment exhibits a distinctive structure in its mathematical solution which is particularly noticeable when applied to photogrammetry. Here there would typically be many more target points than camera locations. When the camera/target measurements are sorted according to target name rather than camera location (which is how they are first generated) then the matrices used to process the data contain large sections with zero values. Further re-organization of these matrices is then possible and this enables the solution to be compressed into a much smaller storage area on the computer. Despite the large amounts of storage which modern computers offer, compression algorithms are still very much in evidence and this feature can still be usefully employed, even if only to provide a fast solution for orientation parameters.

8.3 Bundle adjustment (and least squares)

The relative positions of theodolites can be found by direct methods such as the collimation technique. Very often however some approximate knowledge of the relative positions of targets and instruments is required for a particular method to work. This *initial estimate* is found by some simple procedure based, for example, on approximate pointings as specified by the ECDS method of relative orientation. The standard mathematical technique of *iteration* then takes the initial estimate and continually improves it until no further improvement is possible.

The mechanism used for improving estimates is the method of *least-squares*. Measurements can never be exactly correct, which is an unavoidable fact of life. Systematic sources of error caused, for example, when the line of sight is not exactly perpendicular to the transit (trunnion) axis, can be identified and largely eliminated by software compensation. However small random effects, such as a short-term temperature change which causes a small refraction error in a pointing, cannot be dealt with in this way. Fortunately we can reduce the effects of random errors by

averaging repeated measurements or using more information than is strictly needed.

The method of least squares creates a mathematical model of a situation and derives equivalent exact mathematical measurements from it. These values are compared with the actual measurements and the model altered in succeeding *iterations* until a best-fit between modelled measurements and actual measurements is obtained. The decision on the best-fit is reached by examining the sum of the squares of the differences and altering the model until this sum is a minimum. The model is then assumed to be the best description of the actual measurement situation. Typically the model will be defined directly in terms of the coordinates and rotational parameters needed for an orientation solution.

What then is the bundle adjustment? Simply the name given to the general least squares analysis package which, in this case, takes the *bundles of pointings* from each theodolite and processes them to create a set of coordinates which is a best fit to the angular and distance measurements made.

Although one objective of the bundle adjustment is to make proper use of excess measurements it will also function if only an absolute minimum of data is provided. Furthermore, although it may often be associated with non-linear solutions which require initial estimates it can also provide optimal results for simpler procedures such as orientation by collimation. (Although collimation may provide a direct answer it can still be further optimized. Indeed it must be further optimized if more than the minimum of measurements are made.)

In fact, the bundle adjustment can deal with many different measurement configurations and produce an optimal least squares result for each. In this respect it may make use of its own internal parameters which can be set to force certain conditions. For example, it is obvious that if optimized coordinates of orientation targets can be generated in the setup phase, a few critical object points could be thrown into the adjustment as well. Some of these object positions may be very accurately known, so accurate in fact that the user may wish to keep them at their design values rather than generating a model which may give them slightly different values. This is not a problem for the bundle adjustment which can, with the correct parameter settings, provide the user with a more appropriate model.

Positions held at their design coordinates are known as *control points.* They provide a way of forcing measurements into a particular coordinate system or even controlling the quality of a triangulation configuration if the situation on-site makes it difficult to optimize the measurement geometry.

8.4 Principal features of Axyz bundle adjustment

8.4.1 General

Maximum number of stations = 99
Stations can be occupied by theodolites, Total Stations or laser trackers.
Number of targets is unlimited.
Number of scale bars is unlimited.

Optionally choose between a balanced station network (free net adjustment) or 7 arbitrarily fixed parameters (origin located at lowest numbered station).

8.4.2 Measurements in general

Mixed polar and angular measurements are permitted.

Unlimited distance and angle measurements at any one station.
Reciprocal instrument pointings permitted (collimation measurements between theodolites and Total Stations)
Direct adjustment of angle and distance measurements, i.e. no conversion to other formats.
Scale bar lengths included in the adjustment as measured distances.

8.4.3 Points located by tracker

Only pointings to stationary targets are used.

Measurements made to moving reflector positions are excluded.

Note
Fixed points can be located indirectly by the circle and sphere methods. These are not included in the adjustment. There are no measurements associated with the target circle and sphere centres.

8.4.4 Hidden point devices

Measurements to offset targets on hidden point devices (device points) are not included in the adjustment.

Derived hidden points are also not included in the adjustment. There are no measurements associated with these points.

8.4.5 Scale

Measured scale bars may optionally be removed from the adjustment. If removed, the measurements to the scale bar targets are also removed, i.e. the scale bar targets are not treated as ordinary target points in this case.

Scale can be provided by:
- Inclusion of scale bar measurements
- Inclusion of distance measurements from stations to targets
- Inclusion of control points (targets with known coordinates)

If scale is not included a solution can still be calculated by assigning a default separation of 10 cm between the first two processed stations.

8.4.6 Weighting of measurements

All measurements have an individual weighting but this cannot be individually adjusted. Weights depend on default standard deviation values defined for all horizontal angles, all vertical angles and all distances at a particular measuring station. Variations can only be made for individual stations.

Multiple station/target pointings, including pointings in two faces, are always averaged or reduced to a single value. The bundle adjustment therefore only sees one weighted representative pointing between a station and a target.

A single representative measurement derived from multiple pointings has the same weight as a single pointing.

8.4.7 Weighting of targets

Target coordinates can be treated as known values by weighting. Weighting is achieved by assigning suitable standard deviations to the coordinates. These weighted coordinates are normally called *control*

coordinates. Typically weighting is high so that the corresponding coordinate values effectively remain fixed in the solution.

Control coordinates can be *fully weighted* or *partially weighted* and the weightings can be individually set and adjusted. Fully weighted means all three X,Y,Z values are weighted. Partially weighted means one or more individual coordinate values is treated as unknown and not weighted.

8.4.8 Fixing stations

Station parameters can optionally be held fixed. This treats the parameters as known values by weighting with a very small variance.

This enables new stations to be added to a network without changing the parameters of existing stations. It also enables the adjustment to be used as a single point solution for locating further target points by fixing existing station parameters.

8.4.9 Referencing to gravity

There is no requirement for any instrument to be levelled. However, a network can be referenced to gravity in a number of ways.
- Use control points with coordinates in a levelled coordinate system
- Precisely level one or more theodolites or Total Stations
- Make tilt sensor measurements at one or more laser trackers

Theodolites and Total Stations are referenced to gravity by making their standing axes exactly vertical (within practical tolerances).

Trackers are levelled by measuring the amount of residual tilt away from the vertical using a Leica NIVEL tilt sensor.

Levelled instruments can be optionally treated as non-levelled.

8.4.10 Theodolite modelling

Theodolites and Total Stations cannot be manufactured according to design. For example the telescope axis may not be exactly perpendicular to the transit axis. The bundle adjustment offers an option to calculate a theodolite model, also called theodolite indexing.

Modelling does not apply to individual instruments but to individual stations. Each station can be independently modelled but there is no common model if the same instrument is used at several stations.

8.5 Operation of Axyz control points

Control points have already been introduced. See "Control points" on page 102. Briefly, control points have the following features:

- They force the orientation results into the coordinate system of the control
- Excess control data influences the shape of a network and can therefore "control" weak network geometry to improve coordinate accuracy
- They can supply scale to a triangulation network

Typically all 3 coordinate elements of a control point are known but partially known control points are permitted. To define a coordinate system when scale is independently measured, control points must provide at least 6 coordinate elements. For example:

- One point supplies all 3 elements (reference X,Y,Z values)
- One point supplies 2 elements (reference X,Y values)
- One point supplies 1 element (reference Z value)

If control also defines scale, 7 coordinate elements are the minimum.

Normally at least 3 fully known control points, supplying 9 coordinate elements, are available. This is already an excess of control or *over-determined* situation which will influence the network shape. Since control can influence network shape <u>the control coordinates must be consistent to a high accuracy</u>, where this accuracy is better than **Axyz** can deliver. If this is not the case, the control will distort and degrade the inherently better results which **Axyz** can provide.

Control points are implemented by treating them as another type of measurement with a high weight. Use of weights allows them to "float" slightly from their nominal positions. This is a practical approach which is physically justified since they must themselves be measured by some other device such as a CMM. However it does mean that a bundle adjustment will compute small residuals for the coordinates. The technique used in **Axyz** does not permit control points to be absolutely fixed but by assigning them very small standard deviations, i.e. very high weights, they can for practical purposes be considered fixed.

Axyz does not alter control values but users should be aware that statistics involving residuals can indicate problems with control data.

Users have the option to check control points by temporarily treating them as normal points, thereby removing their "controlling" influence from the bundle adjustment.

8.5.1 Use of weight flags

To create control coordinates which are effectively fixed, known with a small degree of uncertainty, or unknown, ***Axyz*** uses weight flags. These can be defined and edited in the Data Manager.

FIXED

This flag assigns a high weight to the coordinate. Its modelled value should change very little, producing a very small residual which for practical purposes is zero.

Internal value $= 10^{-6}$ (10^{-12} for variances).

NOT FIXED

This flag assigns a very low weight. Its value is effectively unknown and large changes in the modelled value are possible.

WEIGHTED

This flag assigns a specific weight to the coordinate. It allows for small differences in the quality of control coordinates.

8.5.2 Operation of ECDS control with bundle adjustment

ECDS control points are used both to control the ECDS bundle adjustment and to act as reference values for 3D transformations.

As in ***Axyz***, ECDS also permits the use of partial control points and operates with weight flags.

ECDS weight flags

The ECDS concept does not allow for individual weights of control points. In effect the control values are either known and not allowed to change, or they are unknown. All uncertainties in coordinate values are assumed to be

associated with the measurements. There is therefore no ECDS flag corresponding to WEIGHTED in *Axyz*.

The actual ECDS flags are:

FIX (fixed)
The coordinate is given a very high weight and the modelled value changes very little from this. Effectively zero residuals are produced.

UNK (unknown)
The coordinate is assumed to be unknown and therefore has zero weight.

APX (approximate)
A coordinate flagged as approximate is also unknown but has been provided with a reasonable estimate of actual value. This is useful for calculating approximations, but otherwise the coordinate is treated as unknown.

8.6 Levelling constraints

Instruments need not be levelled in order to make successful and accurate measurements. Video cameras are good examples of instruments which are not designed to be levelled. Many laser trackers have only a simple bubble level and cannot be precisely levelled.

Theodolites and Total Stations are designed to be precisely levelled although it may not always be convenient to do this. For example, on a floating oil rig under construction you might choose to operate a Total Station with the compensator switched OFF.

Mathematically any instrument which is not precisely levelled must be treated by the bundle adjustment as non-levelled in order to preserve accuracy. However in practice most instruments will always be approximately levelled because they are designed to be used in an upright attitude. For initial orientation approximations it is sufficiently accurate to treat these as levelled, so there is no contradiction in treating instruments as levelled in the initial orientation phase and as non-levelled in the final optimization.

code .. run .. rec *see MTM/STM* As an option users can also temporarily remove the constraint to keep an instrument levelled in the bundle adjustment. This might be done to track down a source of error.

There are distinct advantages to levelling.
- When instruments are levelled you have an automatic check and correction of small tilts which might develop during measurement.
- Levelling also forces an additional constraint which can improve measurement quality.

If levelled stations are in use, the bundle adjustment recognizes several cases.

8.6.1 Relative orientation (no control points) with levelled theodolites

Here instruments are held levelled by fixing the ω and φ rotations of instruments at zero using high weights. Weighting allows small residual tilts to develop, which corresponds to the actual situation.

8.6.2 Orientation to levelled control points with levelled theodolites

The X and Y values of the control points are on a horizontal plane and the Z values represent heights. The Z axis of the control system is therefore vertical and parallel to the local z axes of any levelled instrument.

CDM *see*

In this case a minimum of 2 control points is permitted. The user must explicitly indicate that the Z axis of the control system points vertically up by selecting this option in the Orientation Module. This identification must be made by the user since the system has no reliable way of determining this fact. Even the use of only 2 control points, which <u>might</u> indicate the fact, could be an error by the operator whose control system is actually tilted with respect to the vertical and who has forgotten to measure a third control point.

Instruments are again held levelled by fixing their ω and φ rotations at zero using high weights.

8.6.3 Orientation to non-levelled control points with levelled theodolites

Here the Z axis of the control coordinate system is not vertical. You cannot therefore force ω and φ rotations of local instrument axes to zero as this makes an instrument's vertical axis parallel to the Z axis of the control system when it should be parallel to the direction of gravity.

In this case the levelled constraint is implemented by forcing all levelled instruments in the network to have their primary (standing or vertical) axes

parallel. This is done by adding two rotational parameters to the solution which effectively represent the direction of gravity with respect to the control system. The parameters are Ω and Φ representing rotations about the X and Y axes of the control system. They therefore represent the tilt of the control system with respect to gravity. The ω and φ values of every levelled instrument should be equal to Ω and Φ respectively and any small difference simply represents a small levelling error.

What if the object moves?

Axyz can only keep parallel the axes of a single sub-group of theodolites. This works when an object does not move during measurement and the instruments are moved around it.

However if the instruments are levelled but fixed in position and an object with control points is moved instead, for example by mounting it on a turntable, then more than one group of parallel axes is created.

To correctly process this situation you would need to have an additional pair of Ω and Φ rotational parameters for each sub-group. This is not currently possible. To resolve this problem treat one sub-group as levelled and remove the levelled constraint from all other instruments.

8.6.4 Orienting laser trackers to gravity

One or more laser trackers in a network can be "levelled" by mounting a NIVEL tilt sensor and adjusting the instrument so that the NIVEL is within its operating range for all pointings. Precise levelling by setting the standing axis vertical is not required since the residual tilt offset is measured.

The instrument is then levelled in the Orientation Module by fixing the ω and φ rotations to values corresponding to the residual tilts. If non-levelled control is used then the additional Ω and Φ rotational parameters for the tilted network are superimposed on these.

8.7 Networks with and without scale

The bundle adjustment will function without any scale information. Normally scale information is provided in several ways:

- Measuring at least one scale bar
- Making at least one polar measurement with a Total Station or tracker
- Including at least 2 control points in a levelled coordinate system
- Including at least 3 control points in a non-levelled coordinate system
- A combination of the above

If scale measurements have not been made a warning is given. However the analysis continues by automatically setting the distance between the first two stations in the network to be equal to 10 cm. This prevents the solution from failing because of a missing mechanical degree of freedom. By choosing an impossibly short separation between the instruments the resulting coordinates are much too small to be realistic. This serves as an additional reminder to the user that scale information was not provided.

Missing scale information is likely to occur when a network contains theodolites only, since trackers and Total Stations are probably employed precisely because they also supply distance information. The situation would then arise if the operators forgot to measure a scale bar.

If possible the scale should be measured and the bundle adjustment repeated. This may not always be possible. For example, each end of a long scale bar might only be visible from different parts of a large network. If the theodolites have been moved the additional measurements cannot be made from the original stations.

8.8 Balanced station v. free net adjustment

8.8.1 In brief: the balanced station network

The Bundle Adjustment has an option to select a *balanced station network*. This has similarities with a *free net adjustment* and is designed to remove an inconsistency in the quality analysis which results when an arbitrary datum is specified. When the datum is arbitrary the coordinates of the origin are assigned zero error. Since errors are relative this does not affect the quality analysis of derived elements such as the distance between two points, it only affects the errors estimated for coordinates. The balanced network distributes these in a more even-handed way.

If the option is not chosen the adjustment will set an arbitrary datum in one of two ways:
1. The first station used to build the network will define the origin and axes of a relative coordinate system. This is a standard *relative orientation* procedure.
2. If control points are used, the network will be located in the coordinate system of the control. This is an *orientation to control.*

A relative orientation will produce a network of the same shape as the balanced network but the first station will define the base system origin and be assigned zero error.

If orientation to control is used, this will override any selection of a balanced network. If you do not want this to happen you must indicate that any measured control points are to be treated as normal unknown points.

> **Note**
> An orientation to control will produce a network of a slightly different shape than a relative orientation, if the control specifies more than the minimum 6 or 7 elements.

8.8.2 What is a free net adjustment?

It is necessary to fix a minimum number of parameters in order to achieve a solution for the optimized network. In simple terms you need to force some location to be the origin, i.e. you must define a datum and the task is sometimes called solving the *datum problem.* If this is not done successive iterations will keep adding small changes to station and target coordinates which will cause the whole network to drift continuously and the coordinates will never converge to a stable solution.

If control points are used the problem is automatically solved. The network is attached to the control points which are fixed in space so the drift is not possible. Note that *Axyz* will only accept control if there is at least the minimum necessary to tie the network down and stop the drift. For example, including just one control point would not work since the whole network could still "spin" around this fixed point.

In the general case where the instruments have not been levelled and no control points have been measured, you actually need to specify 6 elements which must remain fixed. These are 3 coordinates and 3 rotational elements which, in a conventional solution, are defined to have specific

values at one particular station. Normally the first station in a network is assigned 3 zero rotational values and the location (0,0,0) i.e. the first station defines the origin and axes of a relative coordinate system. Strictly speaking a 7th. element must be specified which is scale, but distance measurements for scaling purposes are normally made. See "Networks with and without scale" on page 132.

This situation is known as a solution with *minimum constraints.* If further constraints are added, such as the condition that at least one station is levelled, fewer elements need to be arbitrarily specified.

The quality analysis problem

In the minimally constrained solution, any station can be the origin, not just the first one processed. No matter which is chosen the procedure still calculates the same shape for the network even though the actual spatial locations are different.

The analysis also provides quality information in the form of variance estimates for station parameters and coordinates of orientation targets. This is useful information which is also required in subsequent procedures such as locating additional target points by the single point solution.

The problem is that a statistical evaluation of <u>coordinate</u> errors depends on the choice of origin. Statistically the origin is assigned zero error and all other stations and points have errors with respect to this. If you change the origin then these numbers change. This does not affect error estimates of derived geometric elements, such as distances between points. Although the numbers at either end of the line may change, their combined effect on the length is the same.

However since much of the user's work relates directly to coordinates it would be a convenience to have some form of analysis which estimates coordinate errors independently of the chosen datum. There is an alternative mathematical approach called the *free net adjustment* which lets you do this.

Simple free net mathematics

The equations which define the network and relate the unknown parameters to the known measurements are processed by matrix analysis. The equations can be reduced to the following format:

$$\mathbf{N\,x} = \mathbf{t}$$

x is a vector or list of the unknown elements such as all the station and orientation target coordinates in the network.
t incorporates the known measurement data
N is known as the *normal matrix* and contains numbers relating to the current shape of the network.

If this was simple algebra you would divide **t** by **N** to get **x.** For example, if apples cost 10 Eurodollars each and you spend 50 Eurodollars, how many apples did you buy? There are an unknown x apples so:

$10\,x = 50$ therefore $x = 50/10$, i.e. 5 apples.

In matrix algebra you cannot simply divide by **N** which is an array of numbers, not a single value. Instead you calculate its *inverse*, written $\mathbf{N^{-1}}$.

$$\mathbf{x} = \mathbf{N^{-1}\,t}$$

Mostly it is a standard procedure to calculate the inverse of a square matrix such as **N.** When you multiply the original matrix by its inverse the result is a *unit matrix*, which has a similar function to number "1" in normal algebra.

Unfortunately, if the datum is not fixed in a bundle adjustment you cannot directly invert the normal matrix. If you attempt this you cause a division by zero at some point.

As an alternative to fixing the datum you can use a generalized form of the inverse which avoids the division by zero. This involves imposing another condition, in an analogous way to fixing a station. This condition states that the *trace* of the inverted matrix ($\mathbf{N^{-1}}$) must be a minimum. The trace is the sum of the elements on the diagonal.

This rather abstract condition has a practical physical effect. It holds the centre of gravity of all the coordinates (stations + targets) fixed at the

initial value during the process of iteration towards an optimal solution. In other words, in each iteration the solution moves the stations and targets to more optimal locations, but in such a way that their centre of gravity does not change. The solution cannot therefore continually drift and it will converge to a final optimized arrangement.

Although this form of solution is described as "free" something else has actually been fixed, in this case a centre of gravity. However, with this technique the final shape of the network and object points is the same as before but the quality estimates are assigned in a more even-handed way. No one station or point is given preferential treatment by being assigned zero error.

This effect on the quality analysis might be suspected from the mathematical condition of a minimum trace. When the weighting scheme is based on standard deviations, the inverse of the normal equations ($\mathbf{N^{-1}}$) is also the covariance matrix of the parameters which you are calculating. The diagonal elements of this matrix are the variances of the parameters and the square root of each diagonal element is the estimated standard deviation of the corresponding parameter. The condition of minimum trace simply means that the free net adjustment arranges matters so that the sum of the station and target variances is as low as possible. If you compare the results of a fixed and free net adjustment you will see that the error estimates have generally lower figures in the latter case.

In a free net adjustment the centre of gravity will depend on the chosen starting values and this may influence the actual statistical values by a small amount.

The description above is somewhat simplified. It is additionally necessary to prevent the system from "spinning" continuously so the method also holds the average rotation elements of all the stations fixed at their initial values.

Finally, you may want to know where the origin of the final coordinate system is located. In a conventional solution the origin is located at one of the instruments. In the free net adjustment it is very likely that one instrument initially starts out at the position (0,0,0), i.e. it initially defines the origin. However because it is the centre of gravity which is held fixed, the instrument initially at this location will move away during subsequent iterations. In the final network there will be no instrument located exactly at the origin although one may be close to it.

8.8.3 Balanced station network

In ***Axyz*** a *balanced station network* has been implemented instead of a free net adjustment in order to accommodate some differences in concept.

A free net adjustment is typically used by a photogrammetrist whose camera stations and object points are often all processed together at the same time. When instruments such as theodolites and video cameras are used it is very common to first create a measurement network and then to add more points. The majority of object points are typically not processed in the ***Axyz*** bundle adjustment and the balanced station network therefore concentrates on the stations.

When theodolites and Total Stations are used they are normally levelled and the balanced station network also incorporates this additional constraint.

In practice the balanced station network adopts the philosophy of the free net adjustment by ensuring that no preference is given to any individual station. It achieves this effect by taking the practical consequence of a free net adjustment and directly applying it. This means that it normally fixes the centre of gravity of the initial coordinates of all <u>stations</u> as well as the average value of the initial rotational parameters of the stations.

The method has the following features:

Scale

- If any distance measurements are made these will define an absolute scale for the network.
- If no distance measurements are made a pseudo scale is introduced by fixing the distance between the first two instruments to 10 cm

Since scale is always included by one of the above methods only a maximum of 6 further elements must be fixed in the network.

No constraints

If there are no levelled instruments and no measured control points then:
- The centre of gravity of the initial station positions is held fixed
- The average value of the initial ω, φ and κ rotations for each station is held fixed

Levelled stations

If the only constraint is that one or more stations is levelled and control points have not been measured, then the direction of gravity must define the Z axis of the final coordinate system. This means that all ω and φ rotations are fixed at zero and only κ rotations (bearings) remain free. In this case:

- The centre of gravity of the initial station positions is held fixed
- The average initial κ value of every station is held fixed

Existence of measured control points

If the acceptable minimum number of control points has been measured then the control system defines the final coordinate system. The result is identical to a normal orientation to control.

A balanced station network results in a relative coordinate system except, therefore, where control points override this to create a controlled coordinate system.

8.8.4 How much does the origin drift?

Users have reported that compared to a standard relative orientation with base origin and axes defined by the local axes of a station, the final location of the base axes might move from this as little as 200µm or as much as 2.5" (over 60mm).

8.8.5 Why choose a balanced station network?

A balanced station network results in a less biased view of the variance estimates of coordinates. In fact, it tends to produce smaller values for the variances, i.e. smaller values for the estimated standard errors. In contrast, an unconstrained relative orientation produces larger <u>relative</u> errors, but these are based on the fact that the station defining the origin has no error at all!

Both methods produce the same quality estimates of derived features such as lengths between points, radii, etc. compared to an unconstrained relative orientation. Note, however, that quality estimates will be different if levelling constraints can influence the relative orientation.

Usually the balanced would be chosen to see the effect of removing excess control information so that the network depends purely on the quality of the instrument and scale measurements.

To avoid the presence of control points from overriding the option for a balanced station network, control points must be treated as normal points. This is an additional option available to the user.

8.9 Bad measurements: Blunder detection

Blunders are genuine mistakes in measurement which will distort results and should be removed from the analysis process.

8.9.1 In brief: blunder detection

A simple form of blunder detection is implemented which detects a single bad pointing.

At the end of a bundle adjustment the algorithm calculates the pointing error of each normal or scale bar point used in the solution. If the largest error exceeds the tolerance value set in the "Warnings" section of the instrument module, the pointing is flagged as a blunder.

There is no automatic removal of blunders. Instead it is up to the user to respond to a blunder, for example by repeating the orientation without the offending measurement. Remember that there may be multiple blunders present which remain undetected and that a least squares procedure cannot guarantee that a bad residual means a bad measurement.

8.9.2 Blunder detection in Axyz

A blunder is a totally incorrect measurement or mistake such as a sighting to a wrong target or electronic corruption of data. In the **Axyz** bundle adjustment a very simple procedure is implemented to warn of blunders.

Firstly, pointings are assigned a tolerance level, see:
- STM/MTM Setup menu/Theodolite Warnings
  Coordinate tol. (distance) OR Coordinate tol. (angle)
- LTM Setup menu/Tracker Warnings
  Error tolerance for coordinate (distance)

When the bundle adjustment is complete the points are intersected to analyse the pointing error. Depending on specification by the instrument module, angle or distance offsets are evaluated.

The largest offset which exceeds the specified tolerance is reported as a blunder and the following message displayed:

"Possible blunder on point [wp]/[id]"

It is up to the user to respond to the flagged value, for example by eliminating the associated pointing or the target point itself from the solution and then repeating the computation. A bad residual might identify a bad target and not necessarily a bad pointing. If there were only two pointings into a target it may be the other one which is in error, although if either pointing were rejected it would no longer be possible to locate the target anyway. Users should remember that least squares is not an infallible method of detecting blunders and occasionally it is discovered that the true problem lies somewhere else. Some experimentation may be required and an adjustment may have to be run several times with different eliminations each time in order to track down the error.

8.10 Normal and alternate rotational parameters

Axyz uses three standard rotation parameters to define an instrument's angular attitude in space. Unfortunately their standard order of application ($\Omega * \Phi * K$) cannot be used to cover all cases in a bundle adjustment. If Φ is a multiple of $\pi/2$ then an infinite number of different combinations of Ω and K is possible in order to get the same final result. Since the adjustment works by making small changes to each parameter in every iteration, it would never converge to a solution in this case. Each change to Ω or K would be compensated by a corresponding change in the other parameter and the solution would chase itself in circles.

The situation can be avoided by using the same 3 rotation parameters but applied in a different order ($\Phi * \Omega * K$). This case, of course, fails when Ω is a multiple of $\pi/2$ for exactly the same reason as the standard sequence fails. Together, however, both patterns cover all cases. (3 rotational parameters have quite different numerical values when they are used in different sequences to define a particular rotation matrix.)

The adjustment examines the starting values of the rotational parameters for each instrument and if Φ is within $\pm\pi/8$ of a multiple of $\pi/2$ then the alternate sequence is used to define the associated equations instead of the standard sequence.

When the bundle adjustment finishes it always stores the final rotational parameters in the standard sequence. If the alternate sequence has been

used for calculation the full rotation matrix is first created and then the standard sequence derived from it. In the case that Φ is a multiple of $\pi/2$ there is again the problem than an infinite number of combinations of Ω and K are possible. However in this case you can arbitrarily set Ω = zero and thereby derive a particular value for K, since any one of the infinite number of combinations produces the same result. The problem that existed in the bundle adjustment does not exist here because the final rotation matrix is known and a single, on-off choice can be made.

8.11 Summary of error analysis in bundle adjustment

8.11.1 Target residuals

For each target, the pointings into the target generate the following data on residuals:
- Individual EDM residuals, if relevant
- Individual angle residuals
- Maximum angle residual (reported as maximum angle pointing error)
- Maximum pointing error in distance units

8.11.2 Control point residuals

Individual x, y, z residuals in base coordinate system of any control coordinates, if relevant.

8.11.3 Network statistics

- Individual scale bar residuals
- Maximum angle residual in network
- Blunder detection:
  Pointing with largest error exceeding 3 times set tolerance

8.11.4 RMS values

If RMS values have been requested:
- RMS of full offset distances using perpendicular or spatial offsets, depending on whether measurements are directions or polar. Known as total RMS value.
- RMS axial values of full offset residuals. Each residual is expressed in its axial components in the base coordinate system. The RMS is calculated for each axial set, i.e. RMSx, RMSy, RMSz
- RMS of scale bar residuals

8.11.5 Variance factor

Always calculated:
- Individual variance factors for each point in the network
- The variance factor for network, $\hat{\sigma}_0{}^2$.

8.11.6 Analytical quality estimates (Error propagation)

Input

- Preliminary variance factor $\sigma_0{}^2 = 1.0$
- Standard deviations of angles
  Estimates by user. Defined and edited in station setup.
- Standard deviations of EDM measurements
  Estimates by user. Defined and edited in station setup.
- Standard deviations of scale bar lengths
  Estimates by user. Defined in DM.
- Measurements are assumed to be uncorrelated.

Output

- For each station a 3x3 covariance matrix of location and a 3x3 covariance matrix of rotation.
- No covariances recorded between stations
- No covariances recorded between a station's location parameters and its rotation parameters
- 3x3 covariance matrix for each target
- No covariances recorded between targets

The covariance matrices are calculated using the preliminary variance factor $\sigma_0{}^2$.

8.11.7 Further use of analytical quality estimates

The analytical quality estimates are further used as follows:
- Station covariance data is used in the single point solution
- Target covariance data is used in shape fits

9. Locating points by the single point solution

9.1 Introduction to single point solution

When the relative positions of theodolites in a triangulation network are known, i.e. after orientation, target coordinates can be found by the *intersection* of pointings from all theodolite positions which sight the target. Since pointings do not generally intersect at a single 3D location, and may not truly intersect at all, a least squares solution is used to find a single optimal location for the target which best fits the pointings.
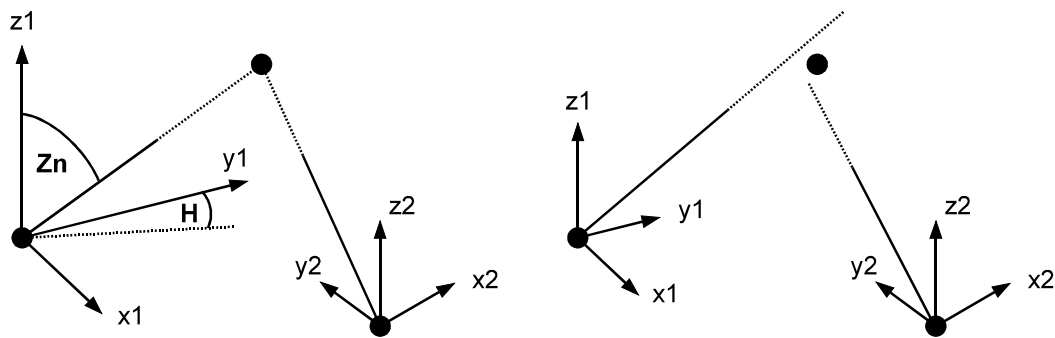
Perfect triangulation   Triangulation in practice

SGLPT-c.WMF

Targets can also be located from a single Total Station by *polar location*. Where several Total Stations are involved in a complex network, and a target is located from more than one Total Station, least squares optimization is again employed to find a single optimal target location.
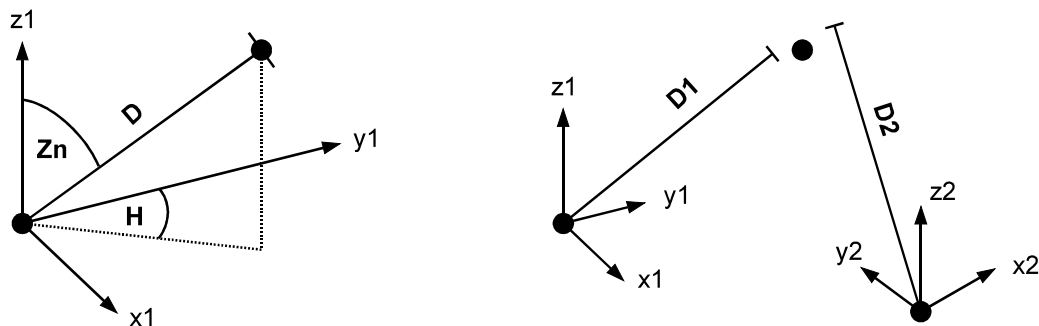
Simple polar location   Multiple polar location

SGLPT-d.WMF

Axyz uses a single method to process both angular pointings only and angles combined with distances in order to locate points in 3D space. This technique is known as the *single point solution.* If measurements involve angles only it is equivalent to an optimized *intersection.* If the measurements come from a single Total Station it creates coordinates without residuals, equivalent to a *polar location.*

9.1.1 Measurements used in Single Point Solution

The following measurements are used:
- Normal measurements
- Stationary target measurement by the tracker
- Measurements to device points on hidden point devices

The following features have no associated measurement and are not therefore processed:
- Fixed points located by the tracker using circle or sphere methods
- Calculated hidden points

9.2 Computation in brief

The single point solution uses the same equations as the bundle adjustment. However, since the station parameters (position and angular orientation of particular instruments) have already been computed by a previous bundle adjustment, they are no longer unknown values. Only the target position is a true unknown.

The bundle adjustment permits both stations and targets to be either known or unknown. The choice in any particular case is decided by the weights assigned to the station parameters or target locations. Very accurately known targets, i.e. control points, can be given very high weights so that their values are fixed in the solution. The produces a controlled orientation. Conversely, stations themselves can be fixed by assigning high weights to their parameters. This is sometimes done in the bundle adjustment itself, when the operator wants to calculate the parameters for a new station in the network but also wants existing stations to keep exactly the same parameters as they currently have.

The same mechanism is used in the single point solution to fix the station parameters. The only real source of uncertainty originates in the pointings and produces the measurement residuals.

9.2.1 Summary of steps in solution

The instrument orientation parameters (position and rotation) are known from the bundle adjustment.

Approximate target coordinates are calculated as for the bundle adjustment.

Optimized target coordinates are based on the same co-linearity equations used in the bundle adjustment.

Instrument parameters are held fixed by assigning each of them a very small variance value of $1/10^{16}$. A small variance generates a high weight.

The individual pointings are weighted according to the preliminary estimates of standard error for horizontal angles, vertical angles and distances.

The least squares solution minimizes the following sum:
(weighted residuals of pointing)^2 + (weighted residuals of inst. params)^2

Since the instrument parameters are effectively fixed, their residuals are effectively zero, so the minimized value is effectively:
(weighted residuals of pointing)^2

9.3 A close-up of the action

Single point solution with pointing residuals

SGLPT.WMF

The diagram outlines the elements involved in locating a target using measurements from a Total Station and a theodolite, but full details are only provided at the Total Station. The example has been deliberately chosen to indicate that both angle and distance measurements are processed. Briefly, in the diagram:
- Black heavy lines indicate what you start with in terms of known data and measurements.
- Dark grey fine lines show what is modelled.
- Light grey indicates the difference between known and modelled data, i.e. the residuals.

The black instrument positions are produced by the bundle adjustment in the base coordinate system indicated by the reference axes XYZ. Local instrument axes are indicated by uvw.

The measured values at the Total Station are horizontal angle H, zenith angle Zn (the vertical circle reading), and distance D. There are corresponding angle measurements at the theodolite (not explicitly shown).

When the optimized target location is found, the difference between measured and modelled pointing produces the standard pointing residuals indicated by δH, δZn, δD.

The best fit minimizes a sum based on the squares of these residuals from <u>each</u> instrument involved in locating the target. These are not summed directly since they must be weighted to allow for mixed angle and distance values and potentially different measurement qualities. It is the appropriately weighted sum which the least squares solution minimizes.

9.4 Quality analysis of single point solution

There are two aspects to the quality analysis.

1. Any particular calculation produces gaps between the pointings (except for a single polar measurement). These are described by the residuals.

2. Assuming that the pointings typically have a particular standard error, the standard error in the target coordinates can be derived purely analytically using the technique of error propagation.

The analytical procedure is necessary because there are not generally enough pointings to generate reliable statistical estimates from the residuals themselves. Most users like to see the residuals in some form to find out how well they "hit the target", particularly since a bad residual may well indicate a mispointing. However a good estimate of the variance or standard error in the target's position can only be found by propagating good estimates of the quality of the measurements. The measurement quality is something the user has already defined and is based on accumulated experience and results.

9.5 Offset residuals and pointing error

The residuals are a measure of the lack of intersection of a particular solution, in contrast to a purely analytical error propagation.

As an alternative to the residuals of pointing, i.e. the measurement residuals directly produced by the single point solution, offset residuals can also be calculated for the optimized target location.
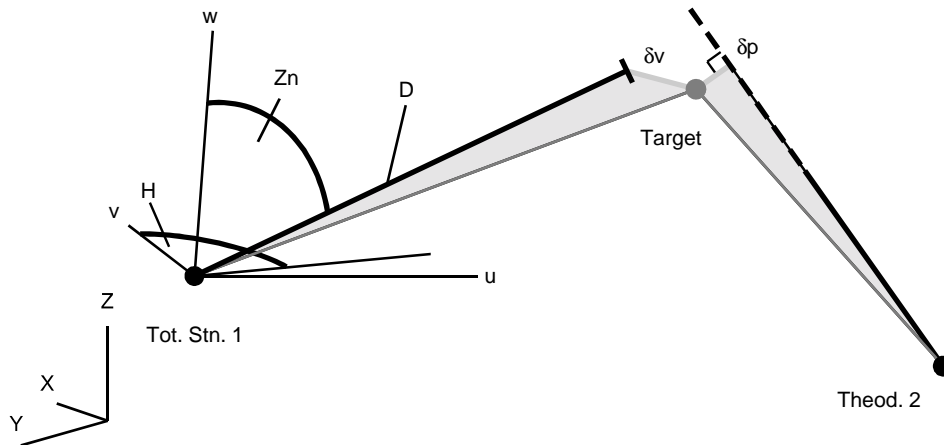
Single point solution with offset residuals

SGLPT-b.WMF

Two slightly different types of offset residual are produced, as illustrated by δv and δp.

δv applies to full polar measurements and is the spatial (vector) offset between the optimal target position and the position calculated from the station using the actual measurements.

δp applies to theodolite pointings and is the perpendicular offset from the optimal target position to the actual theodolite pointing from the station.

Each residual subtends a corresponding spatial angle at the instrument, as indicated by the light grey triangle.

9.5.1 Pointing error (linear and angular)

Offset residuals are evaluated to produce a *pointing error* for the single point solution. This error is displayed numerically and graphically in an MTM measurement window.

The pointing error is a single measure of how multiple rays into a single target fail to meet exactly at the optimized target location. The value is relevant to multiple pointings but not to single polar measurements. The

laser tracker module, which currently only supports a single connected instrument, does not therefore show this value in its measurement windows (***Axyz*** ver. 1.2).
The pointing error is given as either a linear or angular value.

From the optimized point location, the perpendicular offsets to angle pointings or the vector offsets to polar pointings are calculated. The <u>largest</u> of these offsets is the *<u>linear pointing error</u>*.

For each perpendicular and vector offset there is a spatial angle subtended at the corresponding instrument. The <u>largest</u> of these angles is the *<u>angular pointing error</u>*.

9.6 Summary of error analysis in single point solution

9.6.1 Pointing error

One of the following is shown:
- Pointing error in angular units (maximum spatial angle residual)
- Pointing error in linear units (maximum linear offset)

9.6.2 Target residuals (off line)

The off-line Single Point Solution shows the following
- Individual angle residuals
- Individual EDM residuals, if relevant

9.6.3 RMS values

If requested:
- RMS axial values of full offset residuals. Each residual is expressed in its axial components in the base coordinate system. The RMS is calculated for each axial set, i.e. RMSx, RMSy, RMSz.
- Total RMS value. This is the RMS value of the lengths of the full offset residuals

RMS values of angle and EDM residuals are not calculated and therefore not available.

9.6.4 Variance factor

If requested
- The calculated variance factor $\hat{\sigma}_0{}^2$.

Note
There are often not many pointings into a target and the minimum of two would not be unusual. The statistical information is therefore limited in such cases and a value of the variance factor which is very different from 1.0 may not have much meaning.

9.6.5 Analytical quality estimates (error propagation)

Input

- Preliminary variance factor = 1.0
- Standard deviations of angles
  Estimated by user. Defined and edited in station setup.
- Variances for station parameters = $1/10^6$ (effectively fixes the values)

Output

- 3x3 covariance matrix for target, stored in job file

The covariance matrix is calculated using the **preliminary** variance factor $\sigma_0{}^2 = 1.0$

The square root of the diagonal elements of this matrix are the estimated standard deviations of the individual coordinates in the base system. These are displayed in the measurement window.

10. Shapes

10.1 Introduction

This section deals with the creation of standard shapes by form fitting to measured 3D points. ***Axyz*** stores coordinate systems as shapes but these are separately discussed under "Coordinates and coordinate systems" on page 55. In this section a "shape" applies to standard geometric forms such as circles, cylinders and planes.

For further analyses involving points and shapes, such as the intersection of planes or calculation of perpendiculars to the surfaces of shapes, see also:

- "Intersecting shapes" on page 179
- "Bisectors" on page 188
- "Perpendiculars" on page 194
- "Parallels" on page 202
- "Evaluating points" on page 203
- "Creating points" on page 209

10.2 Creating shapes

Shapes are created in several ways by:

- Form fitting
- Intersection of existing shapes
- Geometric constructions

In form fitting a shape is fitted to a set of points using a least-squares technique. Points have 3 sources in this case:

- Measured
- Calculated
- Shape origins

Most points are measured. Calculated points may be derived from the intersection of shapes, for example the intersection of a line with a plane. A calculated point may also be a shape origin, such as the centre of a circle, which has been separately stored as a point. However, by directly naming a shape in a list of points to be fitted, the shape <u>origin</u> is indirectly implied.

Shapes created by form fitting may be directly intersected to create additional shapes. For example, intersection of two planes to create a line.

Points, lines and planes may also be used to construct additional shapes. These constructional features are:
- Perpendiculars from points to shapes and between shapes which create new lines
- Parallels which create lines and planes parallel to existing lines and planes
- Bisectors between points and shape components

10.3 Form fitting

10.3.1 Form fitting in brief

The **Axyz** shape fitting package includes routines to generate the following 7 *standard shapes* from a suitable set of measured points.

- Line (3D)
- Plane
- Sphere
- Circle (3D)
- Cylinder
- Cone
- Paraboloid

The standard shapes are created using a least squares procedure to find the best fitting shape for each set of points.

10.3.2 Form fitting mechanism

The best fit equations work in terms of coordinate offsets between the point to which the surface is fitted and the corresponding modelled point on the shape's surface. The result minimizes the weighted sum of squares of these offsets.

With this technique the software can either treat all fitted points equally (unit weighting) or can take into account variations in measurement quality between points and/or allow for variable quality in different directions at an individual point.

Generally speaking, triangulation and polar methods generate point data whose quality varies in different directions and therefore it may be useful to take account of this. Alternatively unit weighting is a common technique used by other software packages and is sometimes convenient to use regardless of any variation in point quality.

1) Fitting in progress   2) Final fit, unit weighting   3) Final fit, weight by s.d.
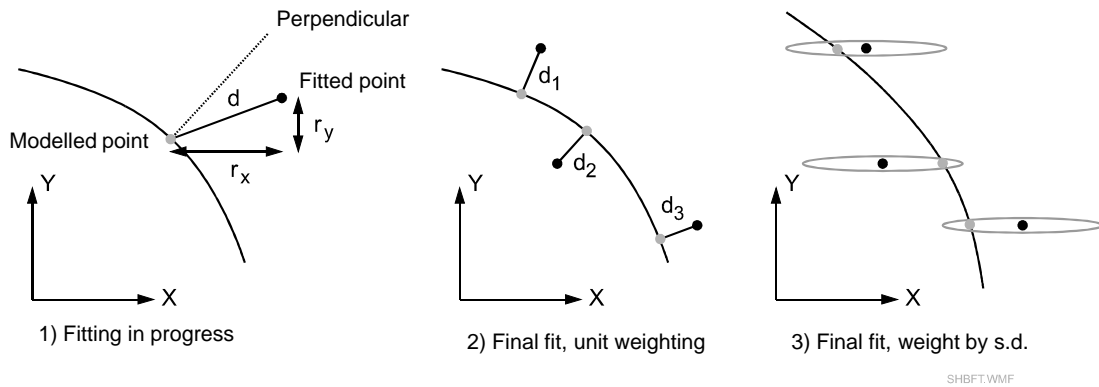
The diagram shows the situation in a simple 2D case and the following comments are readily extended to the 3D case. For each fitted point there is a corresponding modelled point lying on the surface of the fitted shape. The least squares analysis evaluates the residuals of each fitted point r_x and r_y in the direction of the base system axes.

If the fitted points are treated equally and given unit weights, then for N measured points, the software minimizes the quantity:

$$(r1_x{}^2 + r1_y{}^2) + (r2_x{}^2 + r2_y{}^2) + ... (rN_x{}^2 + rN_y{}^2)$$

This is equivalent to minimizing:

$$d1^2 + d2^2 + ... dN^2$$

This quantity is clearly a minimum when the offsets are aligned to the local perpendicular, hence this procedure effectively minimizes the sum of squared perpendicular residuals.

Alternatively the fitting procedure can take account of a variation in quality of the fitted points. This technique uses the covariance matrix, which also defines the standard deviation (s.d.) of coordinates at a point, to create suitable weights. (Any correlations between points are ignored.)

The diagram illustrates this with a situation in which points are much less accurately defined in the X direction than in the Y direction. The ellipses

indicate this variable quality. In this case the results of the best fit will tend to produce much larger X residuals than Y residuals. This makes better use of the best values but no longer produces a result which minimizes the sum of squared perpendicular residuals.

10.4 Shape parameters

10.4.1 Shape parameters in brief

All standard shapes are defined by the parameters of a local origin, a set of 3 local reference axes and possibly one other parameter of form. If 3 local axes are more than the minimum necessary, the excess axes are generated by default. Most 3D shapes require at least one main axis which is designated the local z axis.

Shape	Origin	Local z axis	Form parameter
Line	Any point on line	Along the line	None
Plane	Any point on plane	Normal to plane	None
Sphere	Centre	Z axis of base system	Radius
Circle	Centre	Normal to plane of circle	Radius
Cylinder	Any point on axis	Along the axis	Radius
Cone	Apex	Axis of symmetry	Apex angle
Paraboloid	Vertex	Axis of symmetry	Focal distance

The parameters of the axes can be specified in two ways
- **Unit vector:**
    Components of local z axis unit vector in the currently active coordinate system.
- **Rotation angles:**
    Omega, phi, kappa rotation angles of the rotation matrix for the local axes with respect to the currently active coordinate system.

10.4.2 Shape parameters in detail

All shapes are defined with a local frame of reference (local origin and local axes) and, in most cases, one other parameter. For example, a 3D circle has a local origin at its centre, local axes defined by its orientation in space and a radius.

The parameters of the local reference frame are not always unique. For example, the local origin of a line can be any point along the line and one is chosen for convenience, usually close to the first measured point.

Local axes are also not necessarily unique. Each shape has a full set of 3 axes but some of these can be set arbitrarily. For instance, a line requires only one axis, the direction of the line, to define its orientation in 3D space but two other orthogonal axes are generated perpendicular to this line and through the selected local origin. There are an infinite number of pairs of axes lying in the plane normal to the line's direction which could be selected. In the case of a line the additional axes depend on the current orientation of the active coordinate system and for other shapes the first measured point is typically used to define another local reference axis.

Of the 3 local reference axes, one is always considered the main axis and is designated as the local z axis. The local z axis in all cases is:

- Line (3D)   - the direction of the line
- Plane   - the normal to the plane
- Sphere   - arbitrary, depends on Z axis of base system
- Circle (3D)   - the normal to the plane of the circle
- Cylinder   - the direction of the cylinder axis
- Cone   - the axis of rotational symmetry
- Paraboloid   - the axis of rotational symmetry

When displaying shape parameters the user can choose between unit vector components (direction cosines) of the z axis in the current coordinate system or angle parameters defining the rotation matrix between the local coordinate system and the current coordinate system. In each case there are 3 values. Angle parameters are standard omega, phi and kappa rotations.

Since every shape has a local frame of reference associated with it, these reference frames can be used as alternative coordinate systems. For example, all measured data on a large engine block could be transformed into the coordinate system defined by one particular cylinder bore.

10.4.3 Inside and outside

It is convenient to identify the positive side (outside) and negative side (inside) of a shape's surface. For example, when point offsets are calculated a positive offset will imply a point on the outside and vice versa. It is also important for identifying the direction of a corrective offset due to reflector dimensions and attached targets.

Plane, circle
The positive side is the side of the positive local z axis.

Sphere
A point on the positive side is further from the centre than the radius. From the positive side the surface looks convex.

Cylinder
The negative side is on the same side as the axis.
From the negative side the surface looks concave. From the positive side it looks convex.

Cone
The negative side is on the same side as the axis.
From the negative side the surface looks concave. From the positive side it looks convex.

Paraboloid
The negative side is on the same side as the focus.
From the negative side the surface looks concave. From the positive side it looks convex.

10.5 Setup points for form fitting

Since the least squares analysis is non-linear in every case (even for a line in space!), the shape must be approximately defined before a best fit can be calculated. The shape parameters may either be manually estimated or *setup points* can be used. Setup points represent minimum information and in some cases a simplified definition of the shape to be fitted. Much simpler algorithms, involving direct mathematical solutions, can be applied to these setup points in order to generate the approximate parameter values.

Setup points can be specified in two ways.
1. The first points on the selected list of points to be fitted will be used by default.
2. The user can specify exactly which points in the list of points to be fitted are to be used.

In either case, the setup points may have to conform to a certain geometry. For example, the cylinder fitting routine requires 3 setup points which should lie approximately on a circular section of the cylinder. The parameters of this circle are then easily calculated and used as approximate values for the parameters of the cylinder.

The number of setup points may be less than the minimum number of points required for the actual shape fit.

Shape	Minimum points	Setup points	Geometry of setup points
Line (3D)	2	2	Two points, ideally with a wide separation along the line.
Plane	3	3	Three points forming a triangle (i.e. not collinear) and ideally widely separated.
Sphere	4	4	4 points not all on a circle and ideally widely separated.
Circle (3D)	3	3	Three points on an arc of the circle and ideally widely separated.
Cylinder	5	3	Three points on an arc of a circular section of the cylinder and ideally widely separated.
Cone	6	6	Three points on one arc of a circular section of the cone and ideally widely separated. Another three points on an arc of a different circular section of the cone and ideally widely separated.
Paraboloid	6	5	Three points on an arc of a circular section of the paraboloid and ideally widely separated. A 4th. and 5th. point at different positions along the axis and not in the same plane as the circular section.

The diagrams for the individual shape fits (see later) indicate the setup points and their geometric arrangement.

Note

As explained in "Shape parameters" on page 153, the parameters defining a shape are not unique. Different setup points will normally produce closely similar shapes but some of the parameters describing them may be very different. Users can see this effect by switching the view of parameters between *unit vector* and *rotation angles*. Apart from a possible change of sign, unit vector components of the local z axis should remain the same when setup points are altered. By comparison, rotation angles may show a change only in the rotation about the z axis but equally well may show changes in all 3 rotational elements, particularly if the order of the setup points changes between clockwise and anticlockwise.

10.6 Manually estimating and fixing shape parameters

The *Axyz* shape fitting routines give you the option to manually estimate parameters and optionally fix them by assigning high weights to the estimates. A typical use for fixing parameters is to display residual offsets from a design shape rather than generating a best fitting shape. For example, the user can force a circle's radius to be its design value rather than computing the best fit value. Another use is in cases where an origin is not unique and you may want to force it to be at a particular location.

The parameters are classified into groups:
* Origin coordinates (3)
* Rotation angles (3)
* Size or form parameter (1)

When estimating parameters it is not necessary to estimate all 7 items but if you estimate any parameter within a group the others within the group must also be estimated.

Estimated positional and shape parameters relate to the coordinate system active at the time. It is important to remember this since the values change depending on which coordinate system is currently in use.

The following diagram summarizes the effects in the case of a 2D line fit.
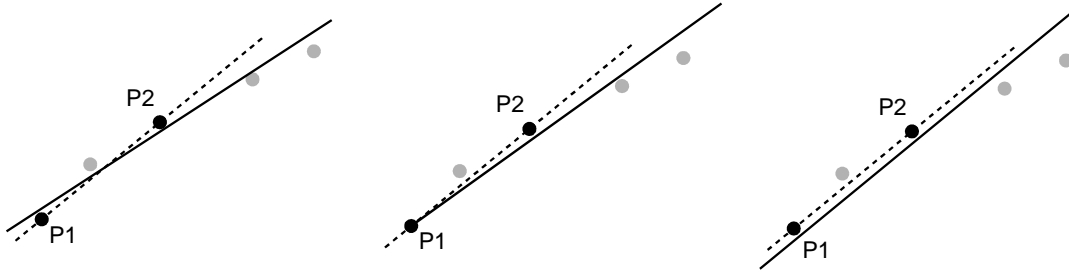
Initial origin at P1	Black best fitting line	Black best fitting line with
Initial direction to P2	with origin fixed at P1	direction fixed at initial value

Black line is best fit

SHBFTLNB.WMF

Notes

In every case the direction of the local x axis is arbitrary. This orientation is defined by the κ (kappa) rotational parameter which is a rotation about the local z axis. The κ value can be fixed at very different values and the routines will still work.

If parameters are fixed at certain values these must be very close to the values which would be calculated if they were not fixed. Fixing a value should be simply a way of preventing the algorithm from letting the value drift slightly. If the fixed value is too far away bad results will be obtained, i.e. the residuals will be large.

10.7 Steps in the fitting procedure

Approximate shape parameters are obtained either from setup points or are entered by hand.

Internally the measured data is in the base coordinate system where calculations are actually made. Approximate parameter data entered by hand is transformed from the currently active system into the base system before use. This transformation procedure also converts any weighted parameters to an equivalent weight in the base system.

In every case the local z axis is the main shape axis. For example, when fitting a cylinder to 5 or more points the cylinder axis is the local z axis. This z axis is critical to the fitting procedure.

It is a mathematical convenience if the local z axis of the shape to be fitted is almost parallel to the Z axis of the base system. Since this is not generally the case the data is "pre-rotated" so that the initial local z axis is parallel to the base Z axis.

The best fit analysis is then actually applied to this pre-rotated version of the data and when it is complete the pre-rotation is reversed. A 2D example in the case of a line fit outlines the method.

Steps in a best fit procedure
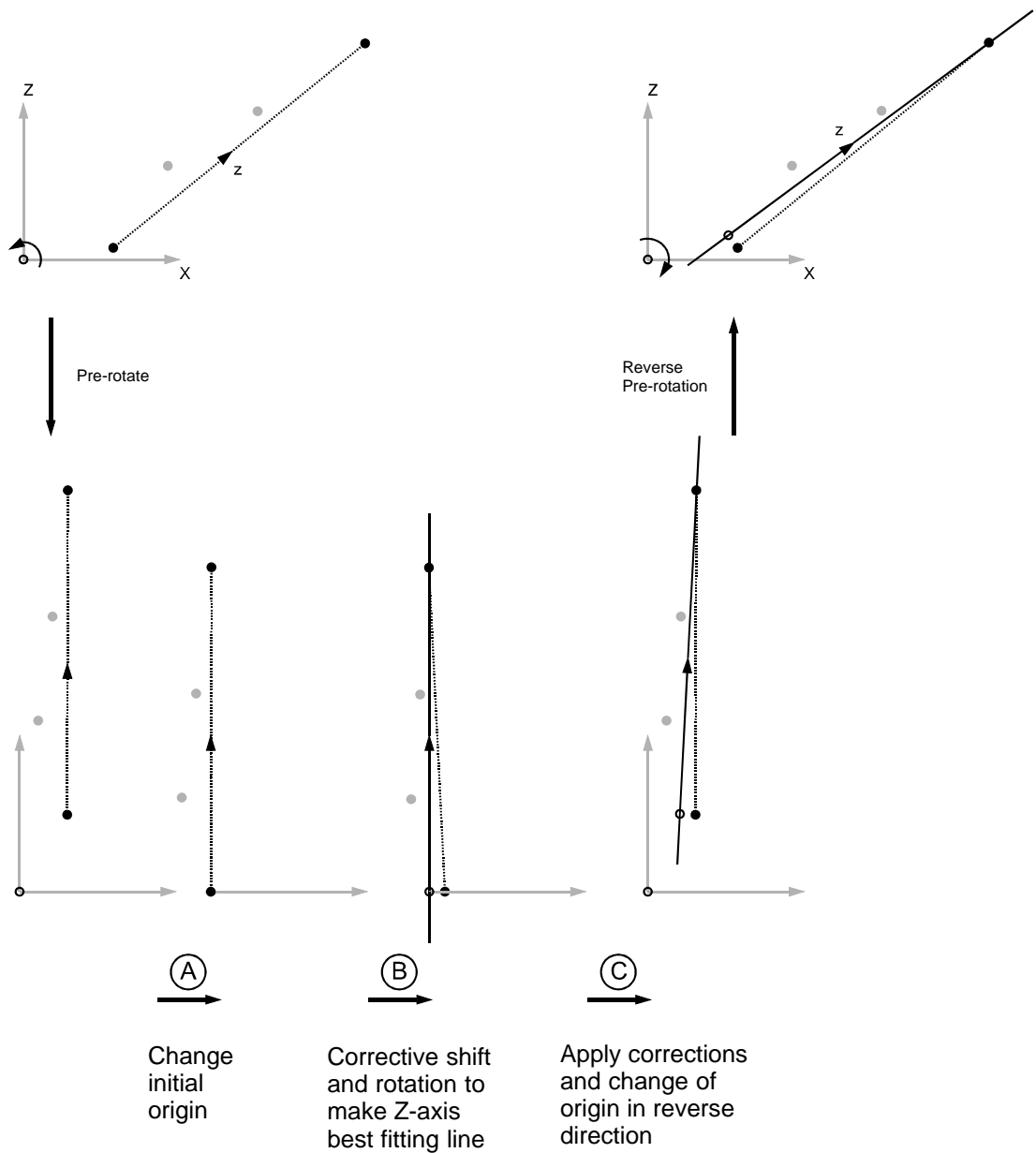
SHBFTLNa.WMF

In the top right the first two measured points have been chosen as the setup points. They create an approximate line with local origin at P1 and local z axis positive towards P2. The points are then pre-rotated to make the local z axis parallel to the Z axis of the base system. At this point the best fit procedure can be started.

The steps in the best fit are indicated by (A), (B) and (C).

A) Shift the local origin onto the base origin.

B) Apply small shifts and rotations by an iterative procedure until the Z axis is the best fitting line to the data.

C) Reverse the total shift and rotation computed in stage (B) and the origin shift made in stage (A). Include a copy of the Z axis so that the best fitting line moves with the data.

Finally the pre-rotation is reversed to bring the data and the best fitting line back into the base system.

10.8 Which method of weighting?

When fitting shapes to measured points, it can be assumed either that the x,y and z coordinates of the points have the same measurement quality (unit weighting) or that the quality depends on their calculated variances (weighting by variance). Since different parameters are minimized, results are not the same in both cases, i.e the fitted shape parameters will depend on the selected weighting method.

Where measurements are better than the shape, and imperfections in the object are the main source of deviation, then unit weighting is probably best.

Where a very accurate shape is measured and measurement errors are the main source of deviations, then weighting by variance is probably best.

The following sections explore these comments in more detail.

### 10.8.1 Diagram: Measurements better than shape
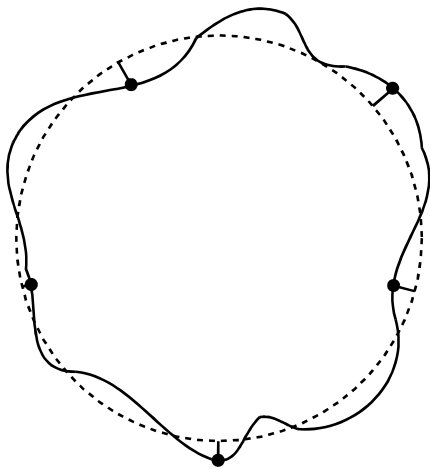
SHAPEWT1.WMF

10.8.2 Measurements better than shape

Ideally measurements are so good that they detect the imperfect edge or surface on an object to which a perfect shape should be fitted. In this case the principal deviations from the shape are due to the imperfections in the object, not the measurement, and it would make little difference if unit weights or weighting by variance were used.
Here it is simpler to apply unit weighting to the measured coordinates and thereby minimize the perpendicular offsets of the measured positions from the fitted shape.

It may be that the imperfect edge or surface shows systematic effects. For example, a fitted plane may show a bulge on one side. This may indicate some damage which the user is looking for or that another shape, such as a sphere, would make a better fit. Either way the user has a pointer to the problem.

10.8.3 Diagram: Shape better than measurements

SHAPEWT2.WMF
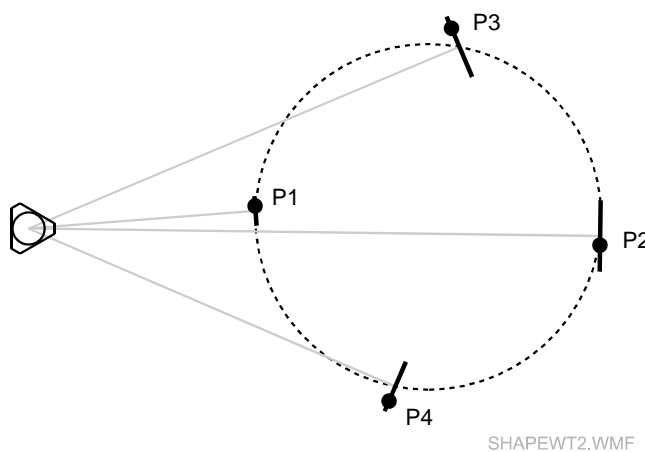
10.8.4 Shape better than measurements

Here it is the measurements which are the main cause of deviations from the fitted shape's surface. In this case it may make more sense to maximize use of the best measurements, i.e. take variable measurement quality into account and use weighting by variance.

The diagram shows a simplified situation in which a laser tracker measures 4 points on a circle. Interferometric distance measurements are much more accurate than the angular measurements so that the region of measurement uncertainty is on a line perpendicular to the true pointing. True pointings are shown by grey lines, uncertainty regions by short black bars and the measured locations by black circles. (The error bars get longer in

proportion to distance from the tracker, since it is the assumed constant angle tolerance which determines their length.)

P1 and P2 are almost on line with the tracker and the centre of the circle. Their separation depends almost exclusively on the high quality distance measurement and effectively provides a very good determination of the circle's diameter.

Weighting by variance should therefore produce an accurate diameter. Small residuals would then be expected at P1 and P2 since all points on the corresponding bars lie close to the circle. Significant residuals would be expected at P3 and P4.

In the example, P3 and P4 lie outside the fitted circle. If unit weights are used in this situation then the fitted circle could expand. This would reduce residuals at P3 and P4 and increase them at P1 and P2. This may be an acceptable result if residuals are then generally smaller and the diameter is not of prime importance.

If weighting by variance gives a very similar result to unit weighting, then measurement quality does not significantly vary. This is an ideal situation which means that measurement quality has been optimized throughout the measurement space. It may therefore be worthwhile, if conditions permit, to create a measurement network which largely achieves this condition.

10.9 Summary of error analysis for shapes

10.9.1 Offset residuals

Offsets are displayed as individual coordinate residuals in the shape's coordinate system, as well as the length of the equivalent 3D offset vector.

The offsets represent the difference between a fitted point and its corresponding modelled point on the surface of the shape. If unit weighting has been selected, the fitted point is perpendicularly from the modelled point on the surface.

Points fitted to 3D circles have their residuals expressed as coordinate offsets but instead of an additional total offset, two other components are used instead:
- Radial, in the plane of the circle
- Perpendicular to the plane of the circle

Note
You expect residuals to have small values. If the displayed coordinate type is cylindrical or spherical a very small residual offset may well show large values for the angular components of the offset. These components simply indicate the direction of the offset and say nothing about the physical size. Directions can have any value between 0° and 360°.

10.9.2 RMS values

Shown only if unit weighting has been selected. The RMS value of the vector length of the residuals is calculated.

10.9.3 Variance factor

Shown only if weighting is by covariance matrices.
If the redundancy is zero in this case, the displayed variance factor is zero .

10.9.4 Analytical quality estimates of shape parameters

Input

Depending on chosen option:
Either 3x3 covariance matrices for points, calculated from bundle adjustment
Or unit weights

Output

If weighting by variance:
Error estimates of shape parameters are calculated by standard error propagation using a value of 1.0 for the variance factor (i.e. the preliminary value).

If unit weighting used:
Error estimates of shape parameters are calculated by standard error propagation using the calculated value of the variance factor. If there is zero redundancy, then a calculation is made using an assumed **preliminary** variance factor of 1. (The same procedure as used for error propagation when weighting by variance.) However, this means $1m^2$, which further means that you are propagating point errors with an assumed standard deviation of 1m. See "Error propagation with unit weighting" on page 52.

Note

When fitting a shape there are two sources of error:

- Errors due to measurements
- Errors due to an imperfectly shaped object

Error propagation can only assume errors in the measurements although the final residuals are due to both sources of error and both will influence the final shape parameters.

10.10 Line (3D)

10.10.1 Geometrical conditions

Minimum 2 points.

The 2 setup points should be well separated in order to ensure a reasonable estimation of line direction.

10.10.2 Initial coordinate system and parameters

SHLINE.WMF
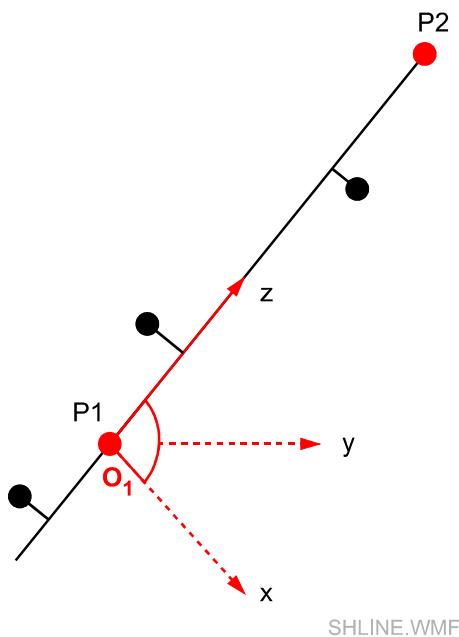
Setup points P1,P2

First setup point, P1, defines the local origin.

Direction from P1 to second setup point P2 defines positive direction of local z axis.

Local x,y axes are arbitrary and depend on the X,Y axes of base system

10.10.3 Manual definition of parameters

The user can manually estimate:
1. The position of the origin in coordinates of the currently active system
2. The direction of the line by defining ω, φ rotations from the currently active system

If you want to fix rotations to force the line into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the line can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$		Line forced **perpendicular** to x axis of base system (Line is parallel to yz plane of base system)
2	$\omega = 0$ φ not fixed		Line forced **perpendicular** to y axis of base system (Line is parallel to xz plane of base system)
3	EITHER ω not fixed $\varphi = \pi/2$	OR $\omega = \pi/2$ φ not fixed	Line forced **perpendicular** to z axis of base system (Line is parallel to xy plane of base system)
4	$\omega = 0$ $\varphi = \pi/2$		Line forced **parallel** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$		Line forced **parallel** to y axis of base system
6	$\omega = 0$ $\varphi = 0$		Line forced **parallel** to z axis of base system

10.11 Plane

10.11.1 Geometrical conditions

Minimum 3 points.
The 3 setup points must define a triangle and not lie close to a straight line.

10.11.2 Initial coordinate system and parameters

SHPLANE.WMF
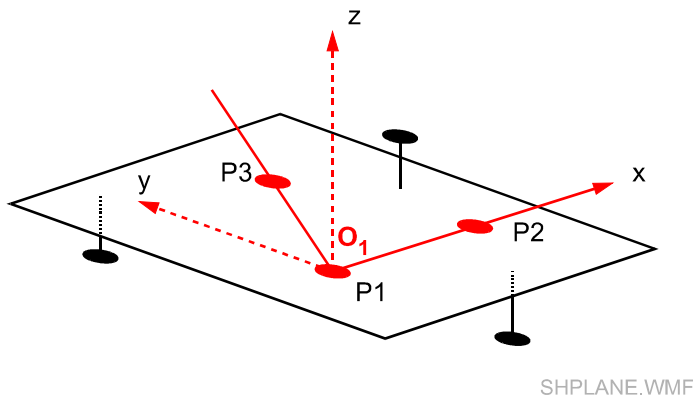
Setup points P1, P2, P3

First setup point, P1, defines the origin.

Direction from P1 to second setup point P2 defines positive direction of local x axis.

The local xy plane contains P1, P2, P3

Local y axis is perpendicular to x and positive from P1 towards P3

Local z axis is computed to create a right-handed set

10.11.3 Manual definition of parameters

The user can manually estimate:
1. The position of the origin in coordinates of the currently active system
2. ω, φ, κ rotations from the currently active system

If you want to fix rotations to force the plane into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the plane can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$ κ not fixed		Plane forced **parallel** to x axis of base system
2	$\omega = 0$ φ not fixed κ not fixed		Plane forced **parallel** to y axis of base system
3	EITHER ω not fixed $\varphi = \pi/2$ κ not fixed	OR $\omega = \pi/2$ φ not fixed κ not fixed	Plane forced **parallel** to z axis of base system
4	$\omega = 0$ $\varphi = \pi/2$ κ not fixed		Plane forced **perpendicular** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$ κ not fixed		Plane forced **perpendicular** to y axis of base system
6	$\omega = 0$ $\varphi = 0$ κ not fixed		Plane forced **perpendicular** to z axis of base system

10.12 Circle (3D)

10.12.1 Geometrical conditions

Minimum 3 points.
Method not suitable for short arcs (subtended angles less than, say, 10°).

10.12.2 Initial coordinate system and parameters
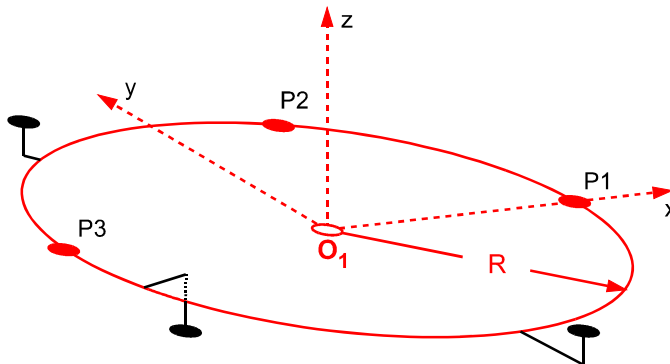
SHCIR.WMF

Setup points P1, P2, P3

The 3 setup points should be well separated in order to ensure a reasonable estimation of the circle parameters. The algorithm uses the first 3 points in the list of points to be fitted.

A radius and centre point are first derived from the setup points.

The direction from the centre to the first setup point P1 defines the positive direction of the local x axis.

The local xy plane contains the 3 setup points.

The local y axis is perpendicular to x and positive from the centre towards P2.

The local z axis creates a right-handed set with the local x and y axes.

10.12.3 Solution for best fit

For a circle fit in 3D the data points are simultaneously fitted to two surfaces:
1. A best fitting plane
2. A best fitting circle in the plane

The least squares solution actually minimizes:
  (sum of squared perpendicular offsets from plane)
 + (sum of squared perpendicular offsets from circle)

10.12.4 Manual definition of parameters

The user can manually estimate:
1. Radius of circle
2. The position of the origin in coordinates of the currently active system
3. ω, φ, κ rotations from the currently active system

If you want to fix rotations to force the plane of the circle into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the plane of the circle can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$ κ not fixed		Circle forced **parallel** to x axis of base system
2	$\omega = 0$ φ not fixed κ not fixed		Circle forced **parallel** to y axis of base system
3	EITHER ω not fixed $\varphi = \pi/2$ κ not fixed	OR $\omega = \pi/2$ φ not fixed κ not fixed	Circle forced **parallel** to z axis of base system
4	$\omega = 0$ $\varphi = \pi/2$ κ not fixed		Circle forced **perpendicular** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$ κ not fixed		Circle forced **perpendicular** to y axis of base system
6	$\omega = 0$ $\varphi = 0$ κ not fixed		Circle forced **perpendicular** to z axis of base system

10.13 Sphere

10.13.1 Geometrical conditions

Minimum 4 points, not all on a circle.
Method not suitable for small surface patches. Tests show that if the points subtend an arc less than 15° the solution will diverge or produce incorrect results.

10.13.2 Initial coordinate system and parameters
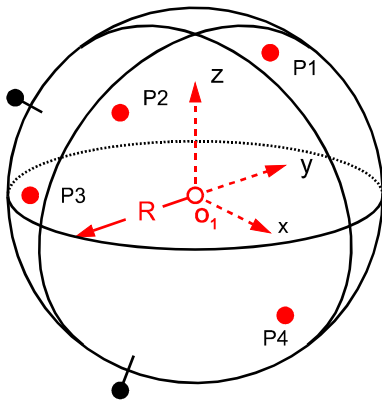
SHSPHRE.WMF

Setup points P1, P2, P3, P4

The setup points are used to calculate the centre and radius by a direct linear method.

Local axes are arbitrary and are set parallel to base axes X,Y,Z

10.13.3 Manual definition of parameters

The user can manually estimate and optionally fix
1. Radius of sphere
2. The position of the centre in coordinates of the currently active system

Axyz currently does not offer the option to specify the directions of the local axes.

10.14 Cylinder

10.14.1 Geometrical conditions

Minimum 5 points.

10.14.2 Initial coordinate system and parameters
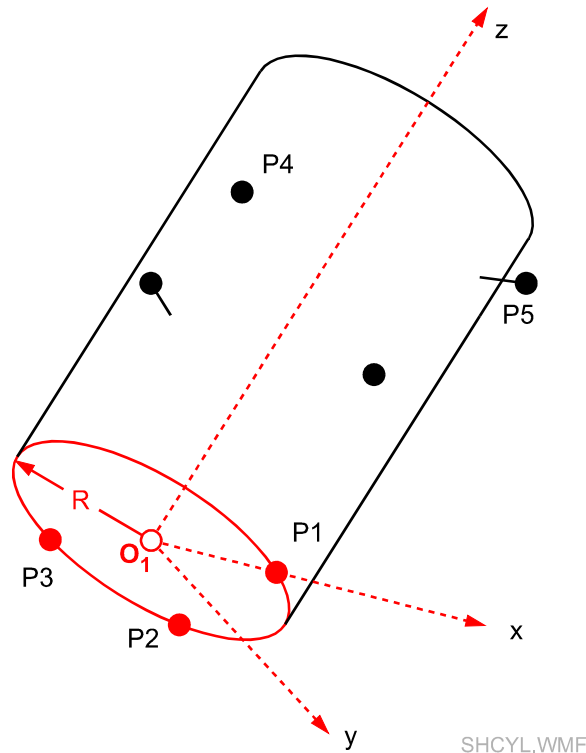
Setup points P1, P2, P3

The 3 setup points should be well separated and lie on a circular section of the cylinder.

A radius and centre point are first derived from the setup points. The circle centre is the local origin.

The cylinder axis passes through the centre and is perpendicular to the plane through the setup points.

The position of P1 along the cylinder axis defines the spatial position of the origin.

The radius of the circle defines the cylinder radius.

The direction from the origin to P1, perpendicular to the cylinder axis, defines the positive direction of the local x axis.

The local y axis is in the plane of the circle and positive towards P2.

The positive local z axis is along the cylinder axis and forms a right-handed set with the local x and y axes.

10.14.3 Manual definition of parameters

The user can manually estimate:
1. Radius of cylinder
2. The position of the origin in coordinates of the currently active system
3. ω, φ, κ rotations from the currently active system

If you want to fix rotations to force the axis into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the axis of the cylinder can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$ κ not fixed		Axis forced **perpendicular** to x axis of base system (Axis is parallel to yz plane of base system)
2	$\omega = 0$ φ not fixed κ not fixed		Axis forced **perpendicular** to y axis of base system (Axis is parallel to xz plane of base system)
3	EITHER ω not fixed $\varphi = \pi/2$ κ not fixed	OR $\omega = \pi/2$ φ not fixed κ not fixed	Axis forced **perpendicular** to z axis of base system (Axis is parallel to xy plane of base system)
4	$\omega = 0$ $\varphi = \pi/2$ κ not fixed		Axis forced **parallel** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$ κ not fixed		Axis forced **parallel** to y axis of base system
6	$\omega = 0$ $\varphi = 0$ κ not fixed		Axis forced **parallel** to z axis of base system

10.15 Cone

10.15.1 Geometrical conditions

Minimum 6 points.
Method is not suitable for apex angles close to 0° or 180°.
Points should be well distributed around the axis.

10.15.2 Initial coordinate system and parameters
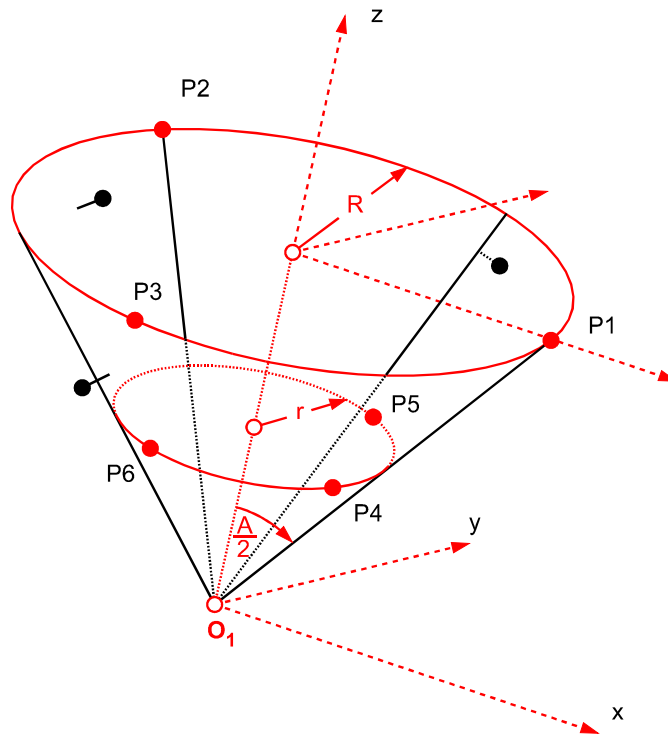
SHCONE.WMF

Setup points P1, P2, P3, P4, P5, P6

The first 3 setup points P1, P2, P3 should be well separated and lie close to a circular section of the cone.

The second 3 setup points P4, P5, P6 should be well separated and lie close to a different circular section of the cone.

Centre points and radii are first derived for the two circles.

The direction of the local z axis defines the axis of the cone and is positive from the small circle to the large circle.

Note 1.
When calculating the vector of the z axis, one of two methods is chosen.

- If the separation of the circles is less than the small radius r, the vector perpendicular to the plane of the large cone is used.
- If the separation of the circles is greater than the large radius R, the vector joining the two centres is used.

The apex of the cone defines the local origin.

The direction from the first circle centre to P1, perpendicular to the cone axis, defines the positive direction of the local x axis.

The local y axis forms a right-handed set with the local x and z axes.

The difference in circle radii and the separation of circle centres enables tan (A/2) to be calculated, from which the apex angle A can be derived.

Note 2.
The fitting procedure uses a temporary origin at the first circle centre but the apex of the cone, when determined, defines the origin of the cone. The location of the apex is easily found from the half angle A/2 and the centre and radius of either circle.

10.15.3 Manual definition of parameters

The user can manually estimate:
1. Slope angle (apex half angle) of cone
2. The position of the origin in coordinates of the currently active system
3. ω, φ, κ rotations from the currently active system

If you want to fix rotations to force the axis into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the axis of the cone can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$ κ not fixed		Axis forced **perpendicular** to x axis of base system (Axis is parallel to yz plane of base system)
2	$\omega = 0$ φ not fixed κ not fixed		Axis forced **perpendicular** to y axis of base system (Axis is parallel to xz plane of base system)
3	EITHER ω not fixed $\varphi = \pi/2$ κ not fixed	OR $\omega = \pi/2$ φ not fixed κ not fixed	Axis forced **perpendicular** to z axis of base system (Axis is parallel to xy plane of base system)
4	$\omega = 0$ $\varphi = \pi/2$ κ not fixed		Axis forced **parallel** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$ κ not fixed		Axis forced **parallel** to y axis of base system
6	$\omega = 0$ $\varphi = 0$ κ not fixed		Axis forced **parallel** to z axis of base system

10.16 Paraboloid

10.16.1 Geometrical conditions

Minimum 6 points.

10.16.2 Initial coordinate system and parameters
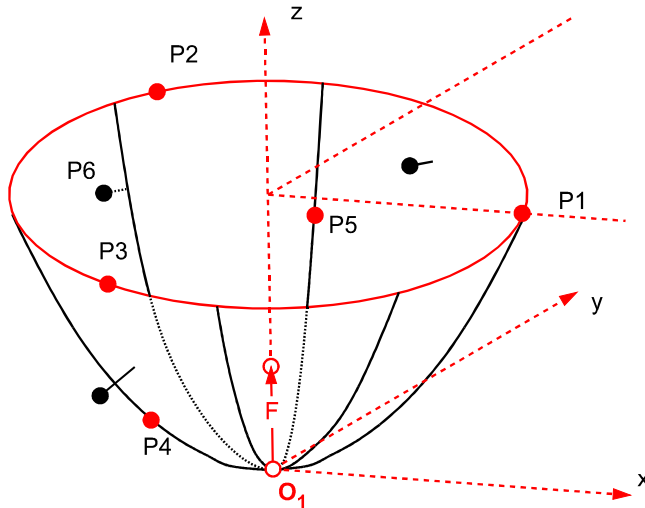
SHPARAB.WMF

Setup points P1, P2, P3, P4, P5

The first 3 setup points P1, P2, P3 should be well separated and lie close to a circular section of the paraboloid. A circle with corresponding centre is calculated for these points.

The axis of the paraboloid passes through the circle centre and is perpendicular to the plane of the circle.

A fourth point close to the vertex would be sufficient to create the approximate paraboloid. However on a physical object such as a parabolic radar dish, the vertex may not be readily accessible or even physically defined. Two additional points off the vertex are therefore used.

Points P4 and P5 should be at different axial heights and not in the same plane as P1, P2, P3. They are used to calculate the focus, focal point and vertex.

The vertex is the local origin.

The local z axis lies along the axis of the paraboloid. Its positive direction is from the origin towards the circle centre.

The direction from the origin, perpendicular to the axis of the paraboloid and towards P1, defines the positive direction of the local x axis.

The local y axis forms a right-handed set with the local x and z axes.

10.16.3 Manual definition of parameters

The user can manually estimate:
1. Focal length, F, of paraboloid
2. The position of the origin in coordinates of the currently active system
3. ω, φ, κ rotations from the currently active system

If you want to fix rotations to force the axis into a particular direction, this will currently only function for definitions made in the base system.

By fixing certain values of ω and φ with high weights (low standard deviations) the axis of the paraboloid can be forced to conform to standard geometrical configurations.

1	ω not fixed $\varphi = 0$ κ not fixed	Axis forced **perpendicular** to x axis of base system (Axis is parallel to yz plane of base system)
2	$\omega = 0$ φ not fixed κ not fixed	Axis forced **perpendicular** to y axis of base system (Axis is parallel to xz plane of base system)
3	EITHER $\quad$ OR ω not fixed $\quad$ $\omega = \pi/2$ $\varphi = \pi/2$ $\quad$ φ not fixed κ not fixed $\quad$ κ not fixed	Axis forced **perpendicular** to z axis of base system (Axis is parallel to xy plane of base system)
4	$\omega = 0$ $\varphi = \pi/2$ κ not fixed	Axis forced **parallel** to x axis of base system
5	$\omega = \pi/2$ $\varphi = 0$ κ not fixed	Axis forced **parallel** to y axis of base system
6	$\omega = 0$ $\varphi = 0$ κ not fixed	Axis forced **parallel** to z axis of base system

10.17 Vector

The main z axis associated with each of the standard shapes may be separately stored as a vector, which is stored internally as a further type of shape.

A vector may only be created by first creating one of the standard shapes by a form fitting routine. Shapes created by other means do not allow the user to create a vector.

Unlike the standard shapes a vector only records a single direction using unit vector components (direction cosines). An origin and local reference axes are not stored with the vector data so a vector cannot define a coordinate system.

11. Intersecting shapes

11.1 Introduction to intersections

CDM

*Analyse/
Intersection*

Intersecting elements can be calculated for the following combinations:
- Shape axis - shape axis (line - line)
- (z axes of any shape or coordinate system)
- Shape axis - surface (line - surface)
- (z axis of any shape with line, plane, circle, sphere, cylinder, cone, paraboloid. Results are points)
- Surface - surface
- (Plane intersected with plane, circle, sphere, cylinder)

Tolerances on the intersection of lines and parallelism of lines and planes can be defined in the CDM "Warnings" menu. These are used to produce warning messages and may prevent calculation of a result if exceeded by defined amounts.

11.2 Intersect: axis - axis (line - line)
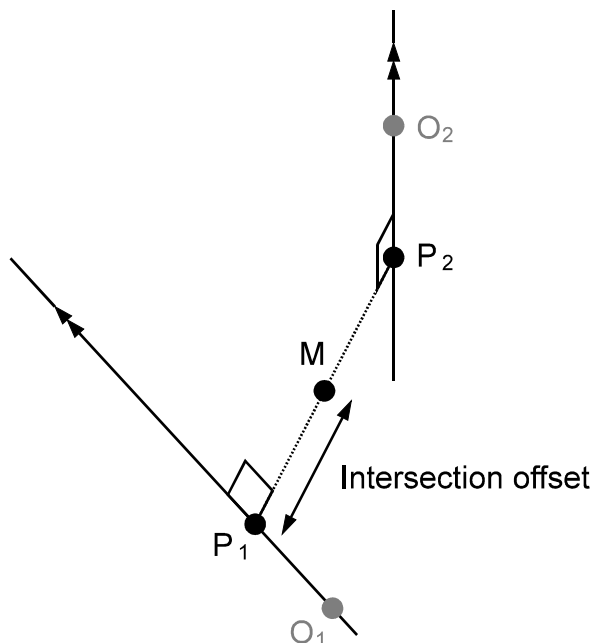
ISCTLNLN.WMF

The diagram shows lines 1 and 2 with origins at O_1 and O_2. The feet of the perpendicular between the lines are at P_1 and P_2 and the intersection point M is midway between them.

11.2.1 Results

Normal
Intersection point M
Intersection offset

Parallel lines (within angle tolerance)
Perpendicular connection starts by default at origin O_1.

Lack of intersection
Ideally lines or axes intersect at a single point. In practice, they never intersect exactly and there will always be a small gap between their closest points. The intersection point is chosen midway between these two points and the intersection offset is half their separation.

The intersection offset (half the line separation) is compared with the tolerance value to produce the following results:
- Offset less than tolerance value:
  Result calculated, no message
- Offset 1x - 2x tolerance:
  Result calculated, warning message "Intersection tolerance exceeded"
- Offset > 2x tolerance:
  Result <u>not</u> calculated, warning message "Tolerance exceeded"

If the lines are far apart and you still require a result, consider using the "Bisector" function. See "Bisector: Shape axis - shape axis" on page 190.

11.3 Intersect: axis - surface (line - surface)

11.3.1 Intersect: axis - PLANE (LINE - PLANE)
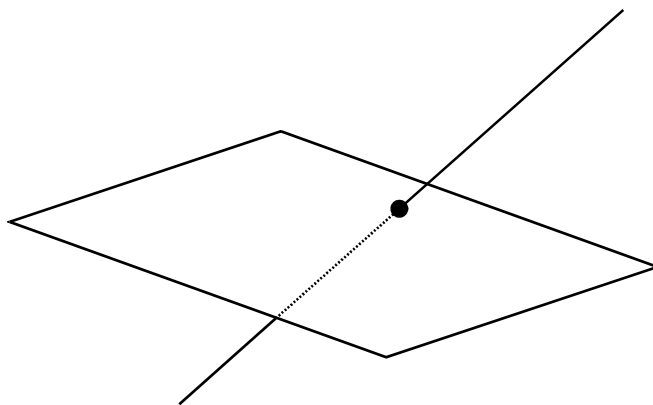
ISCTLNPL.WMF

Results

Normal
Intersection point

Line parallel to plane (within angle tolerance)
Warning message, no result.

> **Note**
> If the line is not parallel to the plane but forms a small angle with it, the intersection point may be very distant and have large coordinate values.

11.3.2 Intersect: axis - CIRCLE (LINE - CIRCLE)

As for any shape axis - PLANE
The plane of the 3D circle is used

11.3.3 Intersect: axis - SPHERE (LINE - SPHERE)
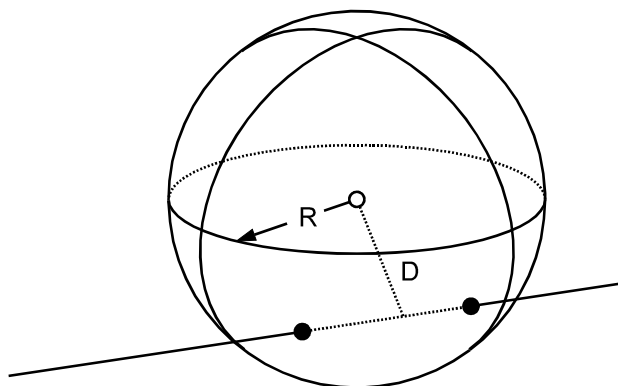
ISCTLNSP.WMF

Results

Normal
Two intersection points
Offset distance D

Axis tangential to sphere
One contact point, offset distance D = R

Axis outside sphere
No result

11.3.4 Intersect: axis - CYLINDER (LINE - CYLINDER)

ISCTLNCY.WMF
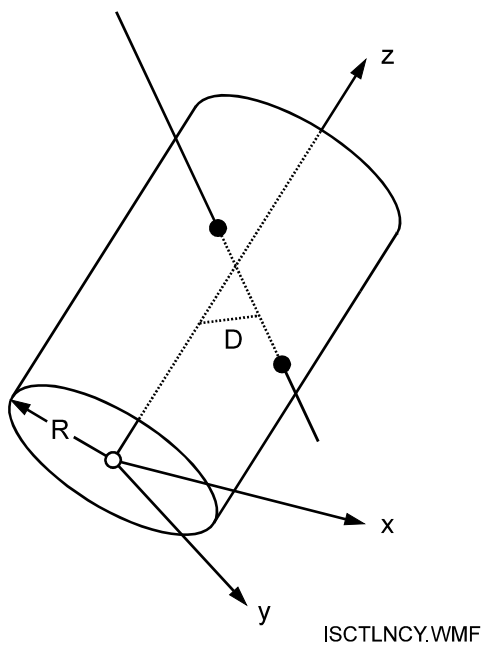
Results

Normal
Two intersection points

Axis tangential to cylinder
One contact point

Axis outside cylinder
No result

Axis on surface of cylinder (parallel to cylinder axis)
No result

11.3.5 Intersect: axis - CONE (LINE - CONE)
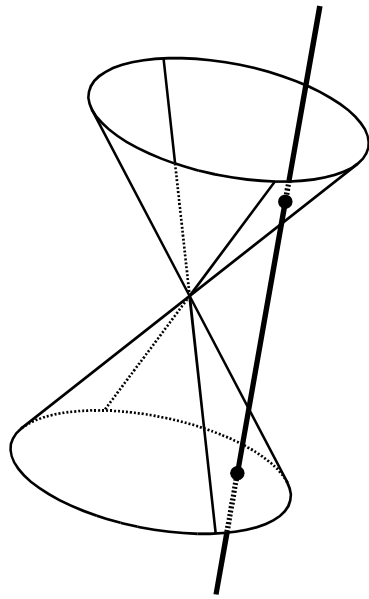
Intersection on both
sides of cone pair
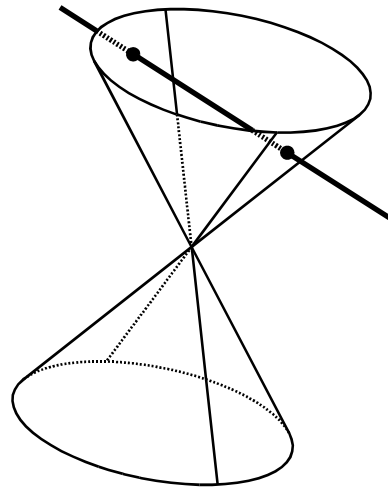
Intersection on one
side of cone pair   ISCTLNCO.WMF

Results

Normal
Two intersection points

Axis tangential to surface of cone
One contact point

Axis on surface of cone (generating line of cone)
No result.

Axis through apex of cone
One contact point identical with apex

Axis outside cone
No result

> **Note**
> The equation of a cone applies to either side of the apex and the
> intersection points may not therefore lie on one side only.

11.3.6 Intersect: axis - PARABOLOID (LINE - PARABOLOID)

ISCTLNPB.WMF
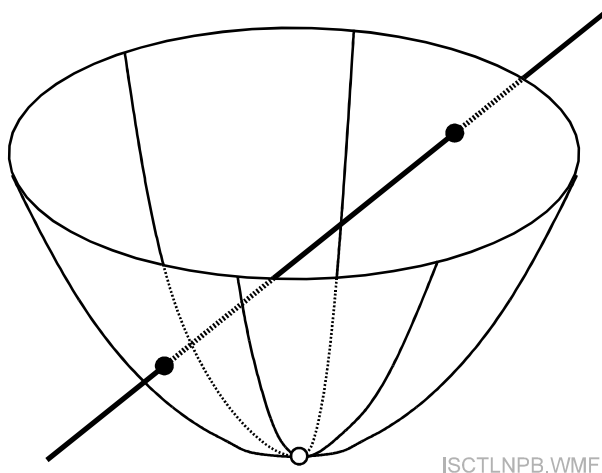
Results

Normal
Two intersection points

Axis tangential to surface of paraboloid
One contact point

Axis through vertex of paraboloid
One contact point identical with vertex

Axis outside paraboloid
No result

11.4 Intersect: surface - surface

11.4.1 Intersect: PLANE - PLANE

ISCTPLPL.WMF

Intersection angle
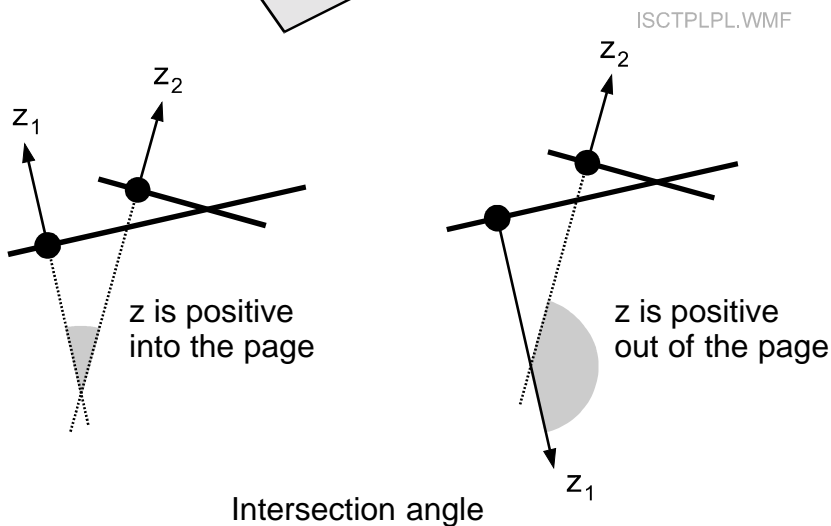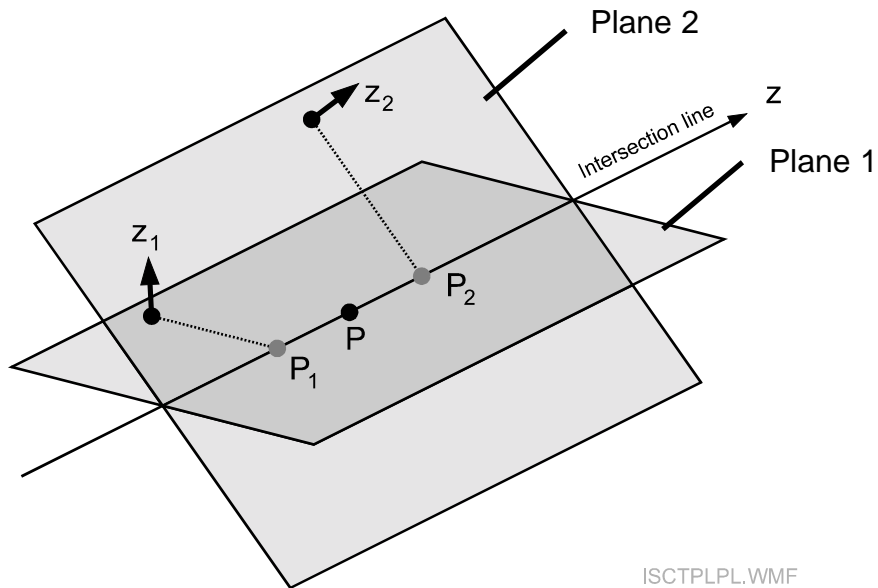
INTPLPLa.WMF

Results

Normal
Line of intersection
Intersection angle between normal vectors (acute or obtuse, depending on directions)

Planes parallel within angle tolerance
Warning message, no result.

Parameters of intersection line
The origin of the intersection line is at P, midway between the feet of the perpendiculars P1 and P2 from the plane origins to the intersection line.

The local z axis is along the line of intersection. To determine its positive direction, look along the line of intersection. Imagine the normal vector of the first selected plane rotated through the intersection angle towards the normal vector of the second selected plane. If the rotation is clockwise the positive direction is away from you and if anticlockwise it is towards you.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the bisecting line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

11.4.2 Intersect: PLANE - CIRCLE
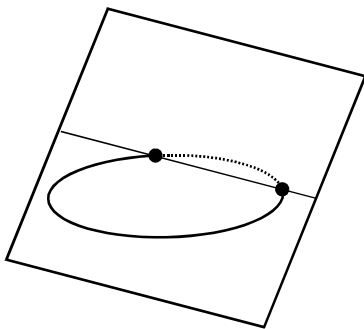
ISCTPLCR.WMF

Results

Normal
Two points of intersection

Plane tangential to circle
Tangent point

Plane outside circle
No result

11.4.3 Intersect: PLANE - SPHERE
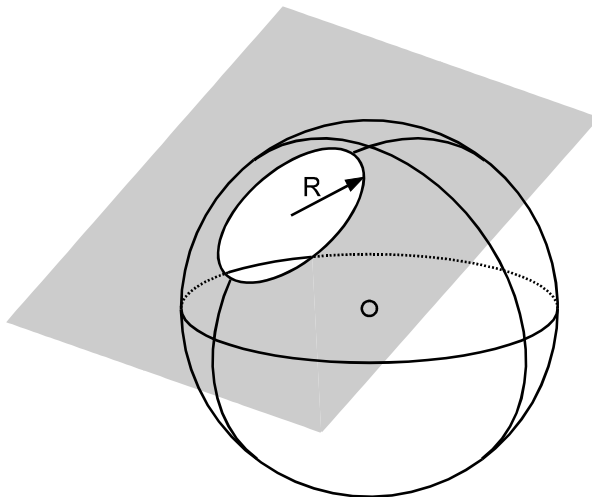
ISCTPLSP.WMF

Results

Normal
Circle of intersection
Radius of circle

Plane tangential to sphere
Tangent point

Plane outside sphere
No result

Orientation parameters of circle
The local origin is at the centre.
The local x,y,z axes are parallel to the local x,y,z axes of the intersecting plane.

12. Bisectors

CDM

Analyse/
Bisectors
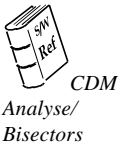
Bisecting elements are centre points, centre lines or centre planes and can be calculated for the following combinations:
- Point - point
- Point - shape axis (line)
  (z axis of any standard shape or coordinate system, result is a point)
- Point - plane
  (result is a point)
- Shape axis - shape axis (line - line)
  (z axes of any standard shape or coordinate system, result is a line)
- Shape axis - plane (line - plane)
  (z axis of any standard shape or coordinate system, result is a line)
- Plane - plane
  (result is a plane)

Some of these calculations generate points. See also section 16 Creating points.

Tolerances on the intersection of lines and parallelism of lines and planes can be defined in the CDM "Warnings" menu. These are used to produce warning messages and may prevent calculation of a result if exceeded by defined amounts.

12.1 Bisector: Point - point

This generates the following:
- A point mid way between the two given points.
- The distance from the mid point to either end point.

12.2 Bisector: Point - shape axis (line)

BISPTLN.WMF
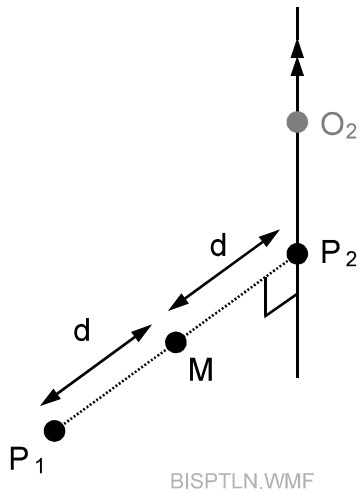
The bisector between an offset point P_1 and a line with origin at O_2. The line can be the shape axis or z axis of any standard shape or coordinate system.

Results

A point M which is midway between the offset point P_1 and the foot of the perpendicular P_2 from this point to the line.

The distance d between M and the offset point (or foot of perpendicular).

12.3 Bisector: Point – plane
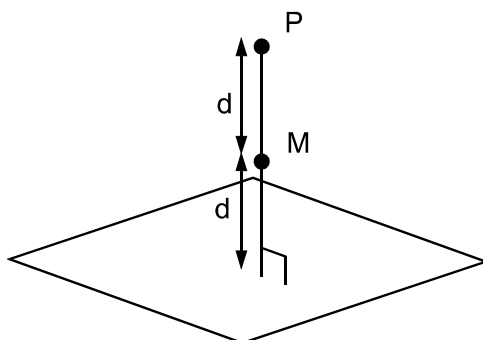
BISPTPL.WMF

Result

A point which is midway between the offset point P and the foot of the perpendicular from this point to the plane.

The distance d between the mid point and the offset point (or foot of perpendicular).

12.4 Bisector: Shape axis - shape axis
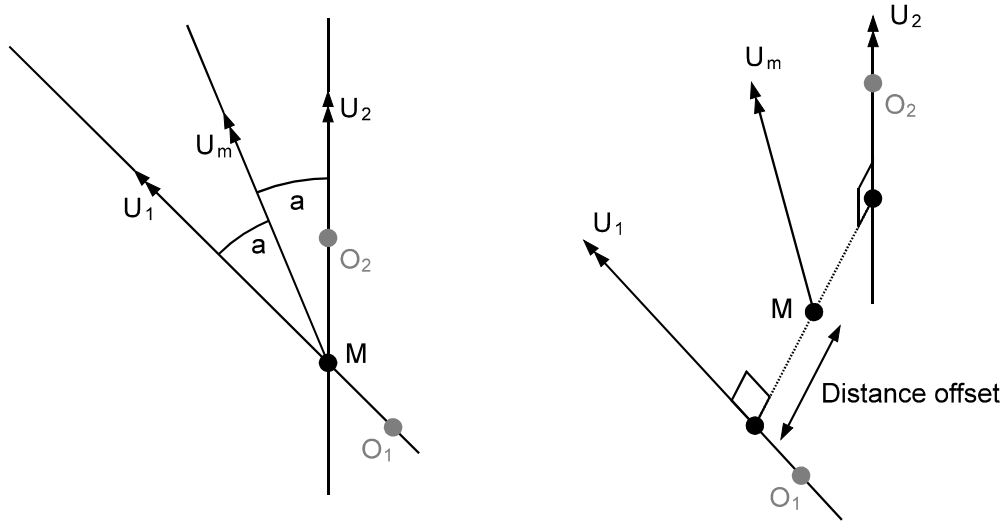
a) Bisector of axes: simple case   b) Bisector of axes: general case

BISLNLNa.WMF   BISLNLNb.WMF

This function creates the bisecting line between two lines or axes which may be the z axes of any standard shape or coordinate system.

Result

Each line or axis is defined by an origin point and a unit direction vector. The bisecting line has a unit direction vector whose direction is the average of the unit direction vectors of the two lines.

In a simple case both defining lines intersect at a single point M which is conveniently taken as the origin point for the bisecting line. The bisecting line makes equal angles with the defining lines and lies in the plane defined by the lines.

In practice the defining lines do not intersect. The origin point of the bisecting line is then the midpoint of the connecting perpendicular between them.

Parameters of line
Origin at M
Local z axis defined by positive direction of bisecting line (unit vector U_m)

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the bisecting line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

12.5 Bisector: Shape axis - plane

BISLNPL.WMF
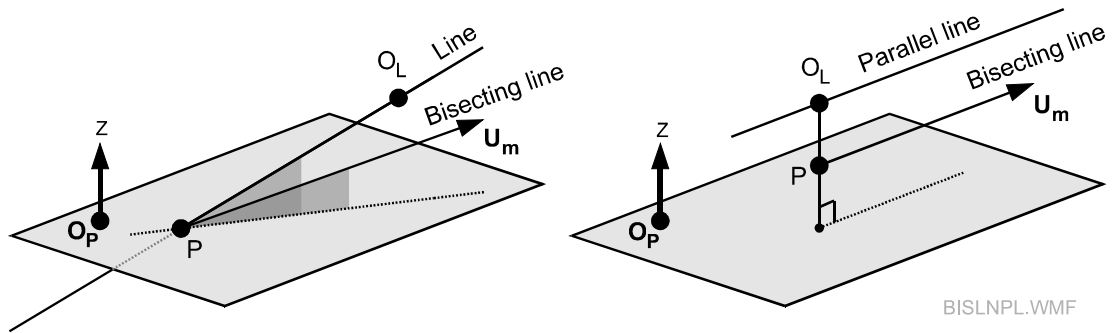
This function creates the bisecting line between a plane and a line or z axis of any standard shape or coordinate system.

Result

Normal
The chosen axis will intersect the plane at an oblique angle. The intersecting axis will have a projected component in the surface of the plane. The new line has its origin at the point of intersection P of line and plane surface and its direction is the bisecting vector between the specified line and its projected component in the specified plane.

Axis parallel or near parallel to plane (within tolerance)
In this case the origin of the new bisecting line is the mid point P of the perpendicular of the origin O_L of the selected line from the surface of the selected plane. The direction of the new line is calculated as in the normal case.

Axis perpendicular to plane
No result.

Parameters of bisecting line
Origin at P
Local z axis along positive direction of the bisecting line

Local x,y axes are derived from the local x,y axes of the plane. Local x,y,z axes of plane are rotated about x and y to point the local z axis along the bisecting line. Rotated x,y axes of plane then represent the local x,y axes of the line.

12.6 Bisector: Plane - plane

BISPLPL.WMF
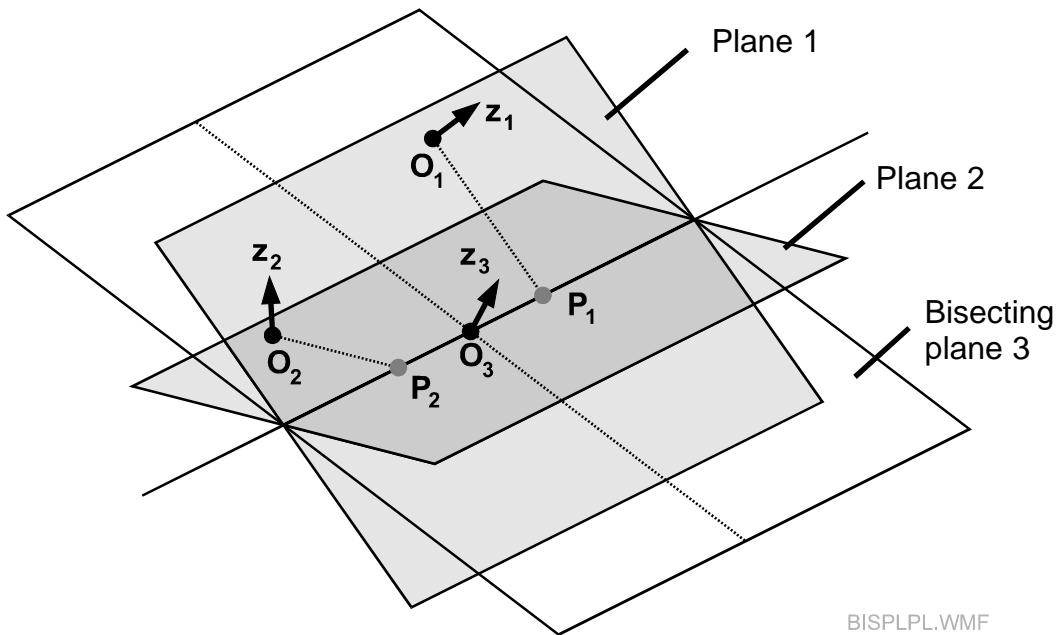
Result

Normal
The planes intersect in a line. The new plane passes through this line. The new origin is the mid point of the feet of the perpendiculars P_1, P_2 from the origins Z_1, Z_2 of the selected planes to the line of intersection. The axis of the new plane is directed along the bisecting vector between the axes of the two specified planes.

Parallel planes
No result.

Parameters of bisecting plane
Origin at P_3, mid-way between P_1 and P_2.
Local z axis is Z_3 which is never more than 45° away from Z_1.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the

bisecting line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13. Perpendiculars

CDM
Analyse/
Perpendicular
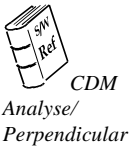
The calculations between the following combinations result in a line and the length of the perpendicular between the elements.

- Point - shape axis (line)
(Point to z axis of any standard shape or coordinate system)
- Shape axis - Shape axis (Line - line)
(z axes of any standard shape or coordinate system)
- Point - shape surface
  (From point to line, plane, sphere, cylinder, cone, paraboloid, but **not** circle)

Tolerances on the intersection of lines and parallelism of lines and planes can be defined in the CDM "Warnings" menu. These are used to produce warning messages and may prevent calculation of a result if exceeded by defined amounts.

13.1 Perpendicular: Point - shape axis (line)

PERPPTLN.WMF
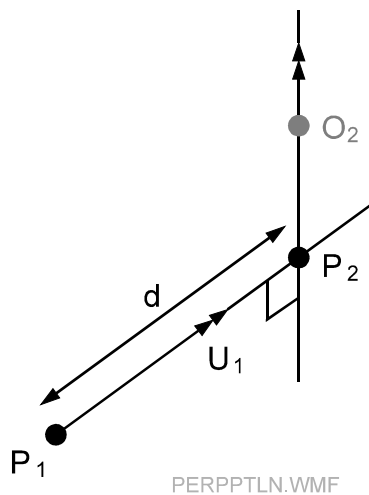
This function creates a perpendicular line from a point to a line which may be the z axis of any standard shape or coordinate system.

Results

Normal
Offset distance d, LINE of perpendicular

Point lies on axis
No result.

Parameters of perpendicular line
The origin is the offset point P_1
The local z axis points from P_1 towards the shape axis (U_1).

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13.2 Perpendicular: Point - shape surface

13.2.1 Perpendicular: Point – plane

PRPPTPL.WMF
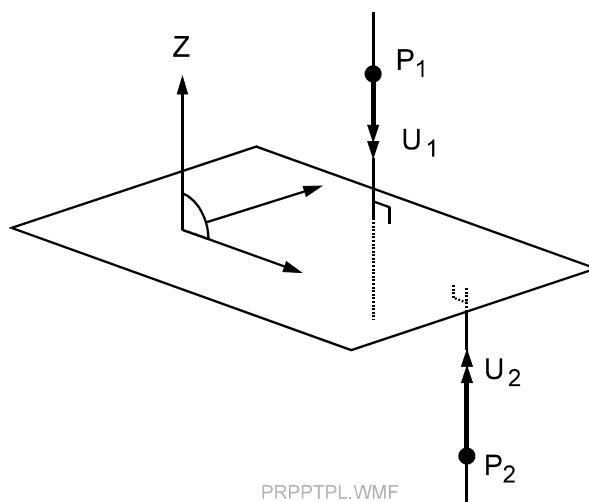
Results

Normal
Line calculated from offset point towards surface of plane.
The offset point is the origin of the line.

Offset point lies on plane
No perpendicular is calculated in this case.

Parameters of perpendicular line
Origin is at the offset point.
Local z axis is positive from the offset point towards the plane.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13.2.2 Perpendicular: Point - sphere
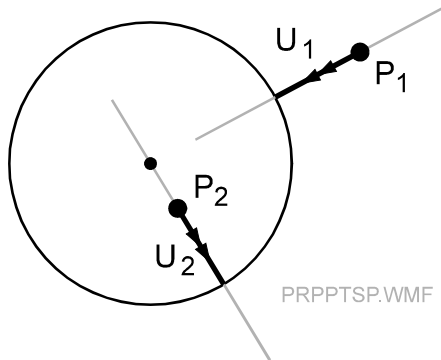
The perpendicular connects the centre of the sphere to the offset point.

Results

Normal
If the point lies outside the sphere, the line is positive towards the centre.
If the point lies inside the sphere, the line is positive away from the centre.

Offset point is at centre of sphere
No perpendicular is calculated in this case. (Infinite number of possibilities.)

Offset point lies on surface of sphere
The perpendicular is also calculated in this case.
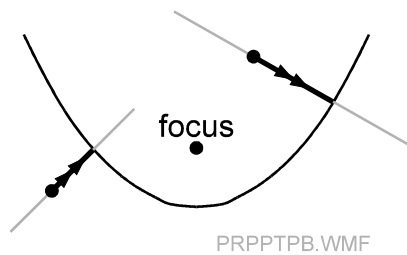The line is positive away from the centre.

Parameters of perpendicular line
The origin is the offset point.
The local z axis is positive as indicated above.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13.2.3 Perpendicular: Point - cylinder
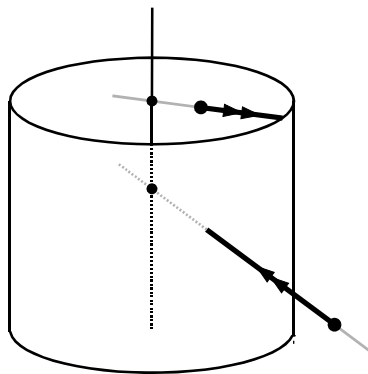
PRPPTCY.WMF

The calculated perpendicular is a line through the offset point and perpendicular to the cylinder axis.

Results

Normal
If the point lies outside the cylinder, the line is positive towards the axis.
If the point lies inside the cylinder, the line is positive away from the axis.

Offset point is on axis of cylinder
No perpendicular is calculated in this case. (Infinite number of possibilities.)

Offset point lies on surface of cylinder
No perpendicular is calculated in this case.

Parameters of perpendicular line
The origin is the offset point.
The local z axis is positive as indicated above.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13.2.4 Perpendicular: Point - cone
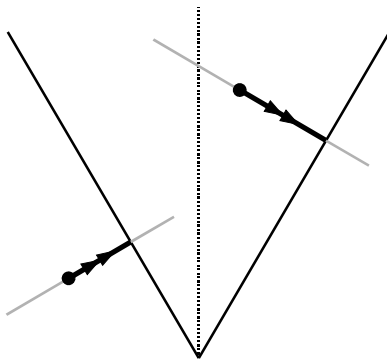
PRPPTC0.WMF

Results

Normal
The line is positive from the offset point towards the surface of the cone.
If the offset point is inside the cone, the nearest part of the surface is
chosen.

Offset point is on axis of cone
No perpendicular is calculated in this case. (Infinite number of
possibilities.)

Offset point lies on surface of cone
No perpendicular is calculated in this case.

Parameters of perpendicular line
The origin is the offset point.
The local z axis is positive as indicated above.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of
base system are rotated about X and Y to point the Z axis along the
perpendicular line. Rotated X,Y axes of base system then represent the
local x,y axes of the line.

13.2.5 Perpendicular: Point - paraboloid

PRPPTPB.WMF

Results

Normal
The line is positive from the offset point towards the surface of the paraboloid.
If the offset point is inside the paraboloid, the nearest part of the surface is chosen.

Offset point is on axis of paraboloid
There are an infinite number of possibilities in this case. One is chosen such that the perpendicular line lies in the local zy plane of the paraboloid.

Offset point lies on surface of paraboloid
No perpendicular is calculated in this case.

Parameters of perpendicular line
The origin is the offset point.
The local z axis is positive as indicated above.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

13.3 Perpendicular: Shape axis - shape axis (line - line)

Length of perpendicular

Perpendicular between parallel lines

PERPLNLN.WMF   PRPLNLNb.WMF
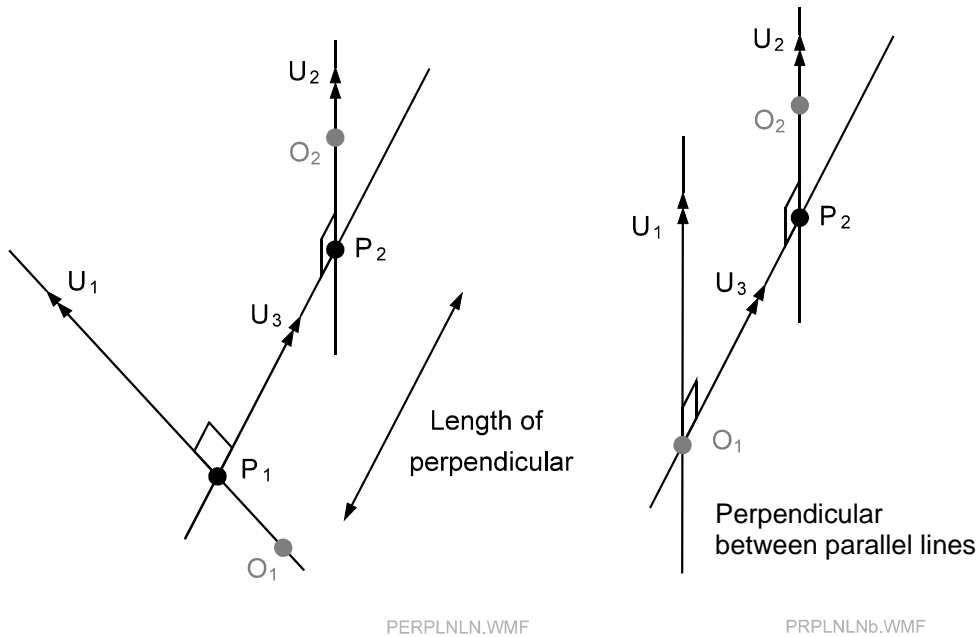
This function creates a perpendicular line between two lines which may be the z axis of any standard shape or coordinate system.

Results

Normal
Offset distance, LINE of perpendicular

Normal parameters of perpendicular line
The diagram shows lines 1 and 2 with origins at O_1 and O_2. The feet of the perpendicular between the lines are at P_1 and P_2 and the intersection point M is midway between them.

The origin of the perpendicular line is at P_1
Local z axis is positive from P_1 to P_2.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

Axes intersect at a point
No result.

Parallel lines
If lines 1 and 2 are parallel there are an infinite number of perpendicular lines between them. By default a perpendicular line is then chosen with origin at O_1 and direction vector towards line 2.

Parameters of perpendicular line between two parallel lines
Origin is at O_1, the origin of the first line.
Local z axis is positive from line 1 towards line 2.

Local x,y axes are derived from X,Y axes in base system. X,Y,Z axes of base system are rotated about X and Y to point the Z axis along the perpendicular line. Rotated X,Y axes of base system then represent the local x,y axes of the line.

14. Parallels

CDM

Analyse/
Parallel

These calculations create one of the following:

- A line through a specified point which is parallel to the z axis of any shape or coordinate system
- A plane through a specified point which is parallel to another plane or circle

15. Evaluating points

15.1 Introduction

This section deals with dimensions involving groups of points and vectors.

15.2 Two point analysis

CDM
Analyse/Two point
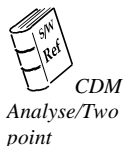
The two point analysis provides full information on the vector between two specified points. This involves:
- The distance between the points
- The components of the unit vector between the points

15.3 Distances between elements

CDM
Analyse/ Distance

Computed distances are calculated along perpendicular connections between the following elements.
- Point - Point
- Point - shape axis
  (z axis of any standard shape or coordinate system)
- Point - shape surface
  (plane, sphere, cylinder, cone, paraboloid, **not** circle)
- Shape axis - shape axis
  (z axes of any standard shape or coordinate system)
- Plane - plane
- Shape axis - plane
  (z axis of any standard shape or coordinate system)

Note

These functions calculate distances based on the length of perpendicular lines. *Axyz* also provides separate functions to calculate perpendicular lines. See "Perpendiculars" on page 194. If there are two equivalent functions, the full perpendicular computation will generate the same value as the distance computation, as well as the definition of the perpendicular line.

15.3.1 Distance: Point - Point

This calculation uses the standard formula for the distance between two points in space:

$$\text{vec1} = (x_1, y_1, z_1) \qquad\qquad \text{vec2} = (x_2, y_2, z_2)$$

$$\text{dist} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}$$

15.3.2 Distance: Point - shape axis (line)

DSTPTLN.WMF
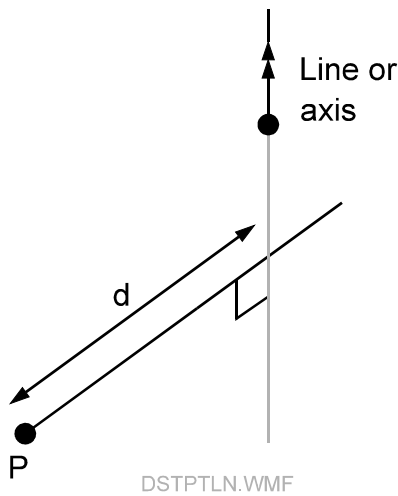
Result

The distance is the perpendicular offset of the point from the line or axis.
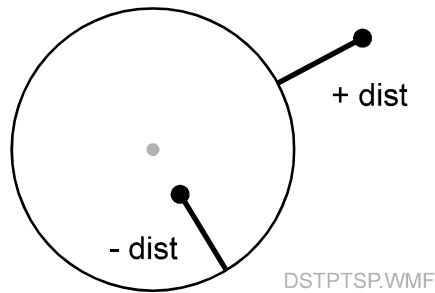
15.3.3 Distance: Point - shape surface

Distance: Point - plane

DSTPTPL.WMF
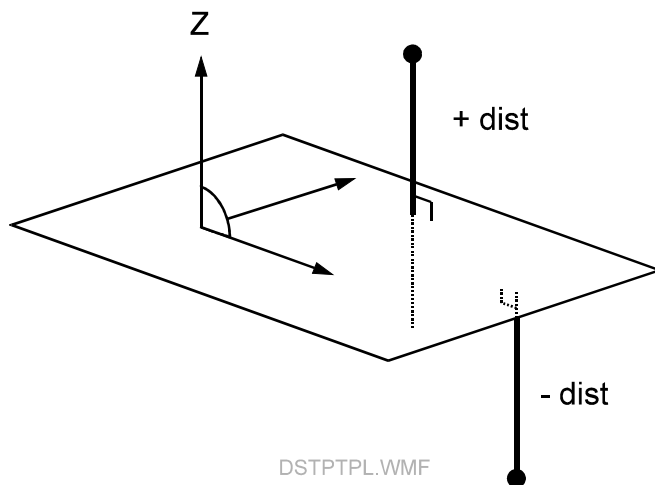
Result

The distance is the perpendicular offset of the point from the plane.

The positive direction of the local z axis is *above* the plane.

The distance is positive if the point lies above the plane and negative if it lies below the plane.
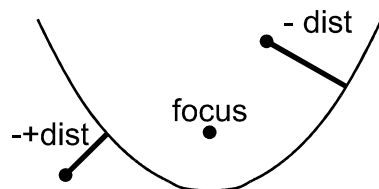
Distance: Point - sphere

DSTPTSP.WMF

Result

The distance is the perpendicular offset of the point from the surface of the sphere.

The distance is positive if the point lies outside the sphere and negative if it lies inside the sphere.

When a point lies inside the sphere the shortest perpendicular distance to the surface is calculated.

Distance: Point - cylinder

DSTPTCY.WMF
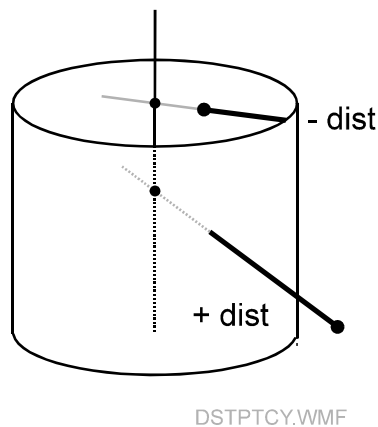
Result

The distance is the perpendicular offset of the point from the surface of the cylinder.

The distance is positive if the point lies outside the cylinder and negative if it lies inside the cylinder (on the same side as the axis).

When a point lies inside the cylinder the shortest perpendicular distance to the surface is calculated.

Distance: Point - cone
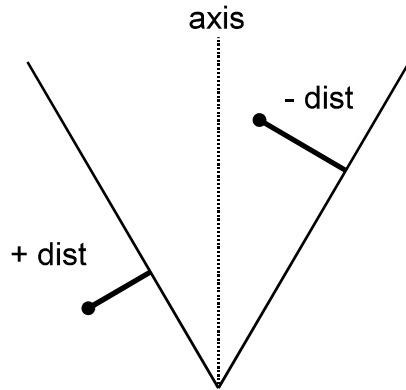
DSTPTC0.WMF

Result

The distance is the perpendicular offset of the point from the surface of the cone.

The distance is positive if the point lies outside the cone and negative if it lies inside the cone (on the same side as the axis).

When a point lies inside the cone the shortest perpendicular distance to the surface is calculated.

Distance: Point - paraboloid

DSTPTPB.WMF

Result

The distance is the perpendicular offset of the point from the surface of the paraboloid.

The distance is negative if the point lies on the same side of the paraboloid as the focus (inside the paraboloid) and positive if it lies on the other side (outside the paraboloid).

15.3.4 Shape axis - shape axis (line - line)
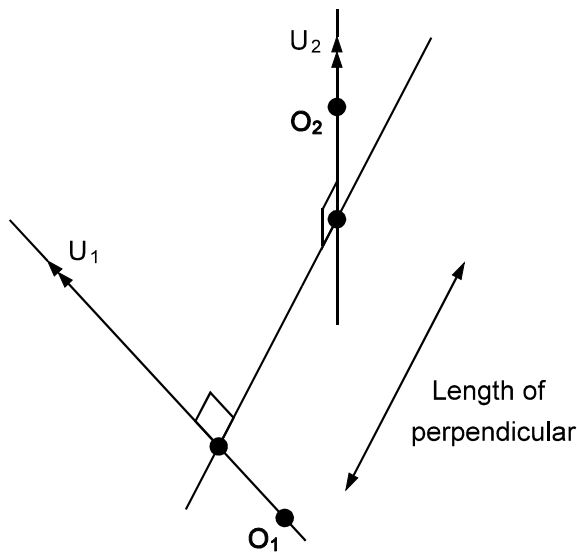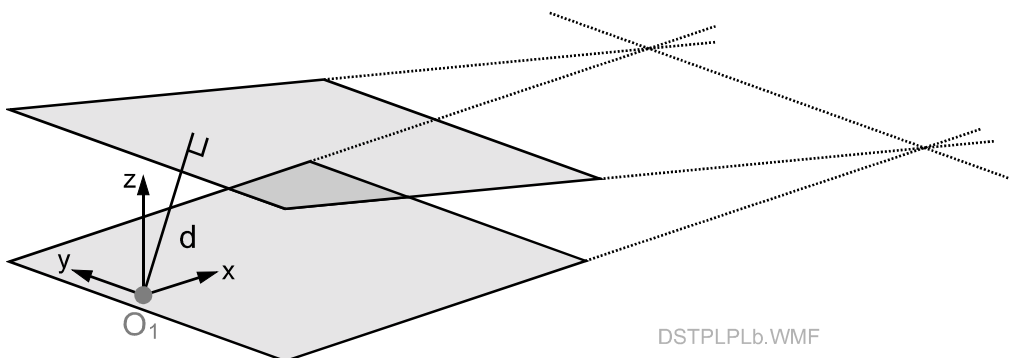

DSTLNLN.WMF

Result

The distance between 2 axes or lines is the length of the perpendicular line
between them.

15.3.5 Plane - plane
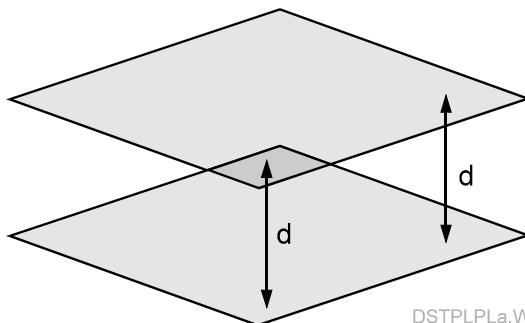
DSTPLPLa.WMF

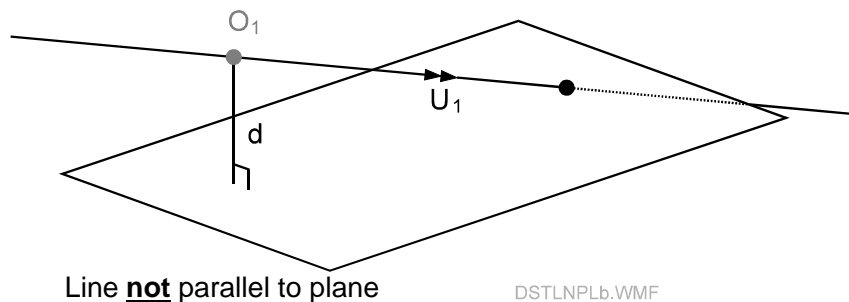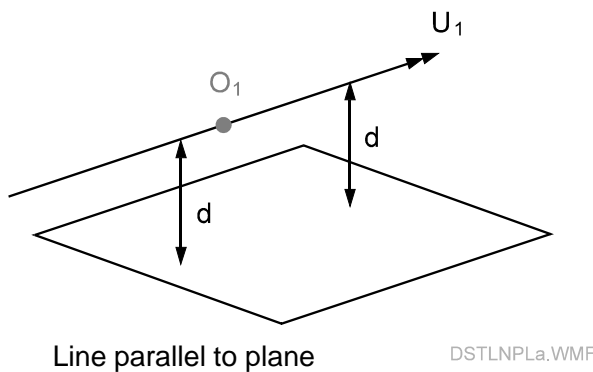Separation of parallel planes

DSTPLPLb.WMF

Separation of planes which are not parallel

Strictly speaking the separation of two planes is only meaningful when they are parallel. Otherwise they intersect and their separation depends on where the separation is measured.

In practice two nominally parallel planes will be measured with a slight angle between them and this angle must first be checked to see if it is within the currently set tolerance for parallelism. The angle between the normal vectors is the angle checked.

If this angle is within the tolerance the separation of the planes is calculated as the perpendicular distance of the origin of the first plane from the surface of the second plane.

15.3.6 Shape axis - plane (line - plane)

Line parallel to plane   DSTLNPLa.WMF

Line **not** parallel to plane   DSTLNPLb.WMF

As in the separation of two planes, the separation of an axis (line) and plane is only meaningful when they are parallel. (A line is parallel to a plane if it is at right angles to the normal vector to the plane.)

Again a check is made to see if the line and plane are parallel within the currently set tolerance for parallelism.
If this angle is within the tolerance the separation is calculated as the perpendicular distance of the origin point on the line from the surface of the plane.

16. Creating points

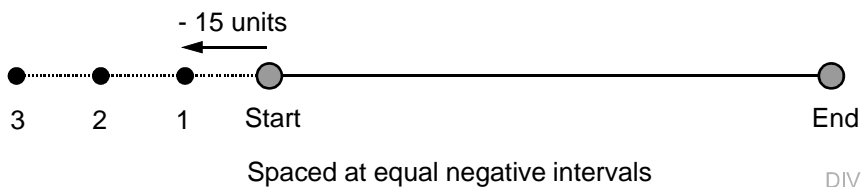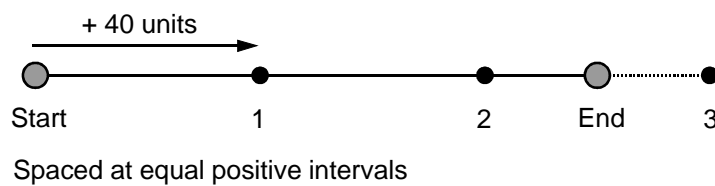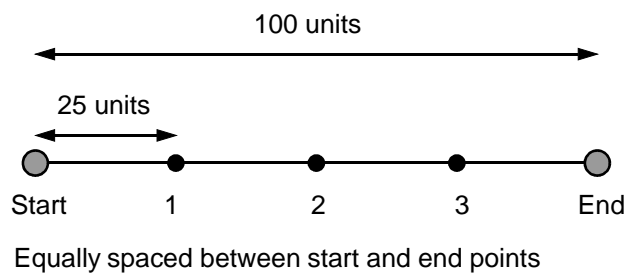## 16.1 Division of lines

CDM
Analyze/
Divide line

This function creates a set of equally spaced points along a line defined by two existing points. The number of new points is specified and the spacing method can be chosen in one of two ways:

1. The new points are equally spaced between the two existing points
2. The new points are equally spaced along the line at a defined interval

Any points or shape origins can be used to define the line. If only one point is requested between the end points the result is a mid point. In this case the same result is obtained using the bisecting function for "point - point". See section 12 Bisectors.

16.1.1 Examples: Division of line

100 units

25 units

Start   1   2   3   End

Equally spaced between start and end points

+ 40 units

Start   1   2   End   3

Spaced at equal positive intervals

- 15 units

3   2   1   Start   End

Spaced at equal negative intervals   DIV-LN.WMF

The diagram shows an example of a start and end point separated by 100 units. A number of new equally spaced points (3 in the example) can be defined in two different ways:

1. Between the start and end points so that they are separated by 25 units in the example.

2. At equal intervals along the line from start point to end point. The example shows 3 points with a positive interval of 40 units and a negative interval of 10 units.

17. Bibliography

17.1 References

1. Analysis and adjustment of survey measurements.
   E. M. Mikhail, G. Gracie, Van Nostrand Reinhold, ISBN 0-442-25369-9
   (Out of print.)

2. Statistical concepts and their application in photogrammetry and
   surveying.
   M. A. R. Cooper, P. A. Cross, Photogrammetric Record, 12(71):637-663
   (April 1988).

17.2 Further reading

An introduction to the algebra of matrices with some applications.
E. H. Thompson, published by Adam Hilger, London. (Out of print.)

18. External cross references.