Where do the ellipses come from?

Blair Hall
b.hall@irl.cri.nz

MSL is New Zealand's national metrology institute operating within Industrial Research under the authority of the Measurement Standards Act 1992.

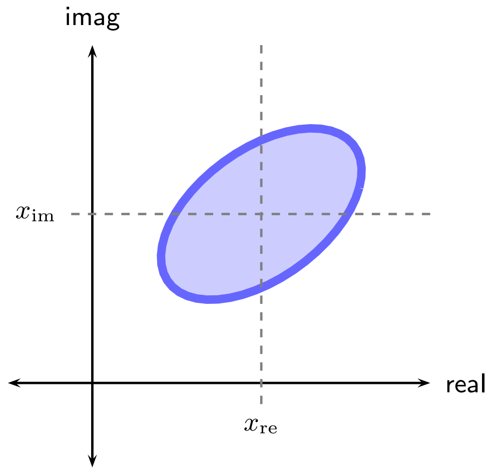
Talk given via internet to the 34^{th} ANAMET Meeting, October 21, 2010.
The meeting was held at Bushy House, Glazebrook Road, Teddington, Middlesex TW11 0EB, UK.

Introduction

An ellipse is the conventional shape of an uncertainty region associated with an estimate of a complex quantity.

Try to answer the following questions:

- how does this shape arise?
- are other shapes possible?
- what does it mean?



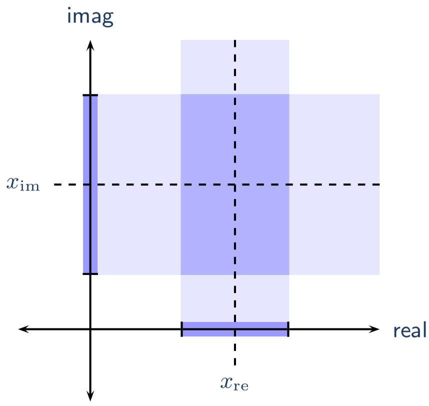
This talk will give an informal explanation of how a ‘conventional’ uncertainty region would be constructed to achieve a particular coverage probability.

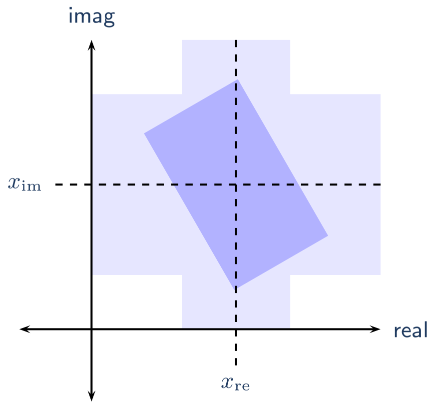
The term conventional is used to distinguish between ‘conventional’ (‘frequentist’) and ‘Bayesian’ viewpoints in statistics. The Bayesian viewpoint is different from the ideas presented here and will not be discussed.

Two uncertainty components need a region

Each component of a complex quantity will have measurement uncertainty and measurements may be correlated.

$$\boldsymbol{x} = x_{\mathrm{re}} + \mathrm{j}x_{\mathrm{im}}$$



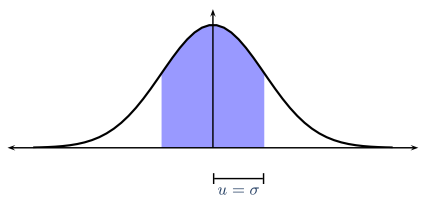


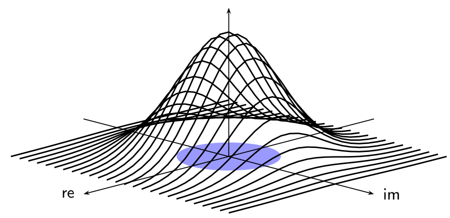
If one pictures a pair of uncertainty intervals in the complex plane, you get intersecting bands. If the intervals are constructed with coverage probability $p\%$, then the band intersection will have a lower coverage probability.

If there is correlation between the estimates of the real and imaginary components, then it is easy to see why the uncertainty region should be turned.

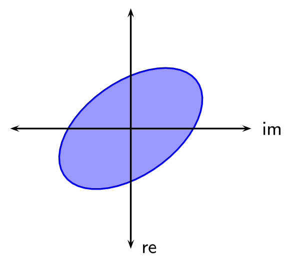
Measurement errors

For a single component (real or imaginary), assume a Gaussian measurement error; for a pair, a bivariate Gaussian error.





The contour of constant error probability in the complex plane is an ellipse



It is measurement errors that give rise to measurement uncertainty.

The usual assumption in real-valued measurements is that an approximately Gaussian error causes a result to deviate from the measurand.

In complex-valued problems, the assumption will usually be that an approximately bivariate Gaussian error causes a result to differ from the measurand. That is, the real and imaginary components of the measurement result are each subject to a Gaussian error and there may be some correlation between these errors.

The justification is based on the central limit theorem and on the fact that we are estimating a mean value (of the associated error distribution).

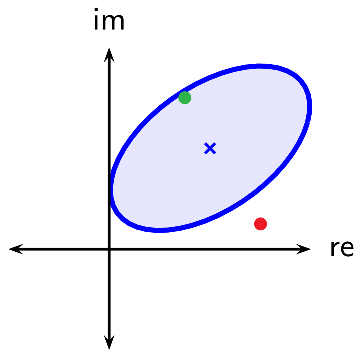
In the rest of the talk we will assume that the error distribution associated with a measurement is known (i.e., the case of infinite degrees-of-freedom). The more general case of finite degrees-of-freedom is easily introduced, but will not be discussed here.

Known measurand, unknown measurement result

If the measurand is known, knowing the error distribution tells us how observations are distributed.

For example, we could draw a contour around 95% of the *probability* density associated with measurement error.

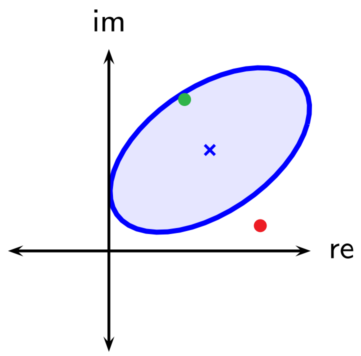
In that case, there would be a 95% chance (before measurement) that an observation would lie in the region.



Thinking in terms of errors, we cannot predict ahead of time what a particular observation will be. However, over may observations, we can predict the proportion that will fall within a given region around the measurand.

Possible measurand, known measurement result

An alternative viewpoint is to say that if an event is observed inside the region, it cannot reasonably be *rejected* as an observation of a given measurand.



However, there are other measurands that would not be rejected either ...

Alternatively, if consider that the value of a measurand is not known (but that the error distribution is). Given one observation, and our knowledge about the error distribution, it is sometimes possible to reject the hypothesis that the observation belongs to a particular measurand. That is, we may conclude that it is too unlikely the error needed to create such an observation of a particular measurand would have occurred by chance.

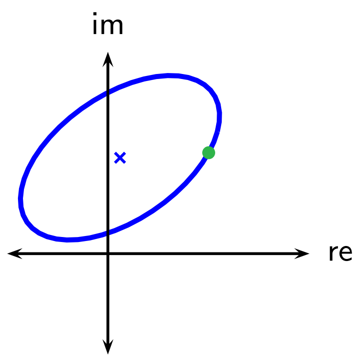
The green dot in the figure lies inside a $p\%$ contour for the error around a particular (candidate) measurand. We would not *reject* this observation as having come from that measurand to a level of confidence of $p\%$. Nevertheless, another measurand nearby could also have given rise to this observation, but we cannot *reject* this one on the basis of what we know.

The red dot, on the other hand, would be rejected at a level of confidence of $p\%$.

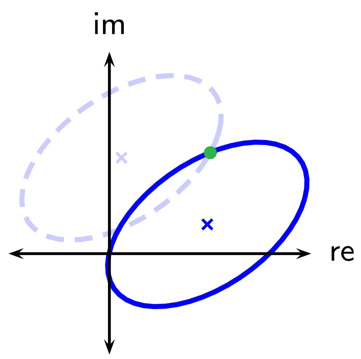
Unknown measurand, known measurement result

Usually the situation is more difficult: we have an observation, but do not know the measurand. At best, we may know the type and shape of the measurement error distribution, but not its location.

Our task is to isolate the measurand.



This measurand is a possibility. It is compatible with the observation, but any further away and we would reject it.



This measurand is also a possibility.

Clearly, for a particular observation there are many different measurands that would not be rejected at some level of confidence.

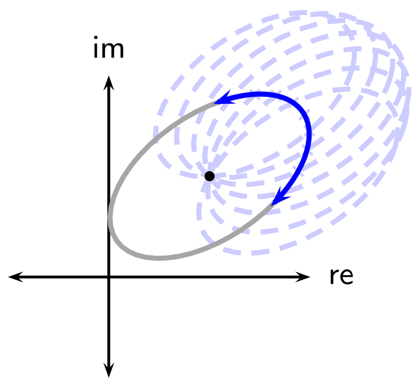
The most distant measurands from the observation that are still acceptable are those for which the observation lies on the perimeter of the error probability density contour.

Constructing the uncertainty region

To form an uncertainty region we consider *any* value that could be attributed to the measurand (with a reasonable level of confidence).

We *do not reject* a measurand at the center of any error ellipse that contains the observed value.

The region obtained this way is also an ellipse (provided the measurement is linear, i.e., that errors to not depend on the value of the measurand)

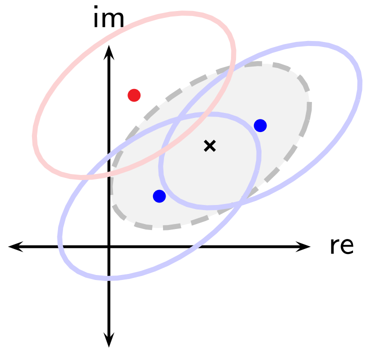


The construction of an uncertainty region determines the locus of all acceptable measurands. It can be envisaged as a process that rejects all situations that could not reasonably have given rise to the observation.

Note that there is no way to attribute more 'weight' (likelihood) to any particular candidate measurand compared to another.

What does the region mean?

- The uncertainty region is centered on the observed value, so for every new measurement it changes – the uncertainty region is random
- A particular region may not contain the measurand
- In the long run, the proportion of regions covering the measurand tends to the coverage probability.



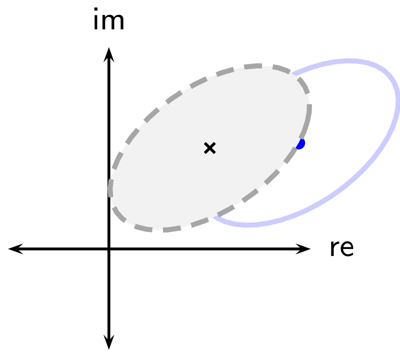
As a consequence of the way in which the uncertainty region has been constructed, we can better understand what it actually means.

Every observation is subject to error, which is unknown. So uncertainty regions are a random. Each uncertainty statement is different. Some of them are not going to 'good': they will fail to include the measurand.

Because we have constructed the uncertainty region with a particular level of confidence, we expect that the proportion of 'successful' regions over many different applications will be equal (or at least close to) that level of confidence.

Some misconceptions

- A region has p% probability of containing the measurand (in fact it either contains it or it doesn't)
- The measurand is close to the centre of the region (there is no reason to prefer any value in the region)
- Most of the distribution of error is inside the region (not if the measurand is near the perimeter!)



As a consequence of the way in which the uncertainty region has been constructed, we can better understand what it actually means.

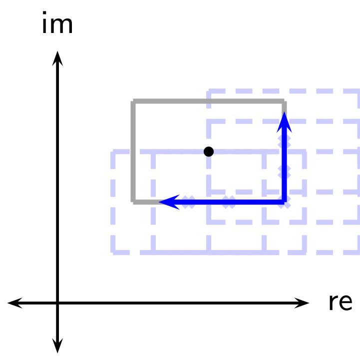
The measurand is a fixed quantity. It's just that we do not know what it is. The uncertainty region is also fixed, for a given observation. So, it is incorrect to talk about the probability of a measurand being inside an uncertainty region: the region either does or doesn't include the measurand (even though we never know which it is).

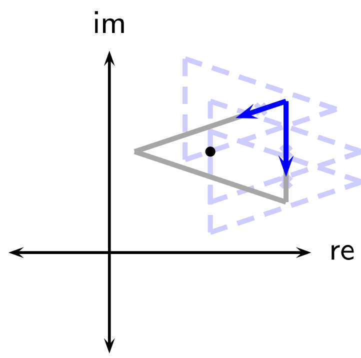
The uncertainty region was constructed by considering, without bias, all of the possible measurands that could have given rise to an observation. So there is no reason to believe that that particular observation is going to be close to the measurand.

Other shapes

If the shape of the error-contour is not an ellipse, then the 'correct' uncertainty region will not be an ellipse either.

If the shape of the error-contours do not have inversion symmetry, then the uncertainty region has the *inverted* contour shape.

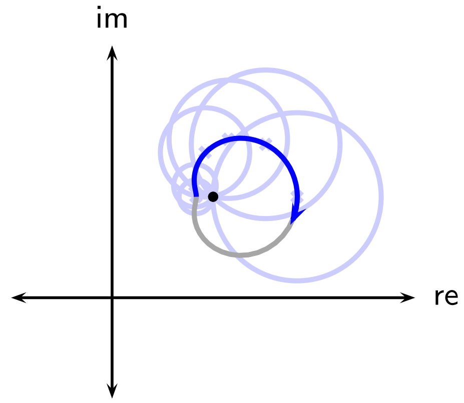




Although it is unlikely to be of practical interest, the same method of construction will generate different shaped uncertainty regions if the error distribution changes.

Other shapes - non-linearity

If a measurement function is non-linear, the error-contour may change size as a function of location in the complex plane. The ‘correct’ uncertainty region will not be an ellipse.



In this example, the error contours are circular. However, the radius (c.f., standard deviation) varies linearly with the real component.

Of more interest is the case when a measurement function is non-linear. In that case, the extent of the error region becomes a function of the coordinates in the complex plane.

This slide shows what happens when the radius of a circular error contour depends on the value of the real coordinate (increasing as the real component increases).

We see that each error contour is a circle, but that the circles get bigger to the right, the shape of the uncertainty region is complicated and is definitely not circular.

Remembering that uncertainty regions constructed in this way *must* achieve correct coverage probability, we can start to see why non-linearity poses real difficulty for uncertainty analysis. To perform a full uncertainty analysis of a non-linear problem (like, for example, the polar representation of uncertainty) we need to consider how the error distributions are transformed by the measurement equation for all the candidate measurands in different locations of the complex plane. The shape of the uncertainty region is going to be complicated and will not generally resemble the shape of the error contours used to construct it.

Conclusions

- An elliptical uncertainty region follows from the assumption that observations (results) are subject to a (bivariate) Gaussian error.
- (However) an uncertainty region is not a contour of the error distribution.
- An uncertainty region contains all those candidate measurands that cannot be rejected on the basis of information available about measurement errors.
- Different types of error distribution would change the shape of the uncertainty region (as would strong measurement function non-linearity).
- An uncertainty region will often contain the measurand, but not always. In the long-run, the coverage probability should match the success-rate of capturing the measurand.