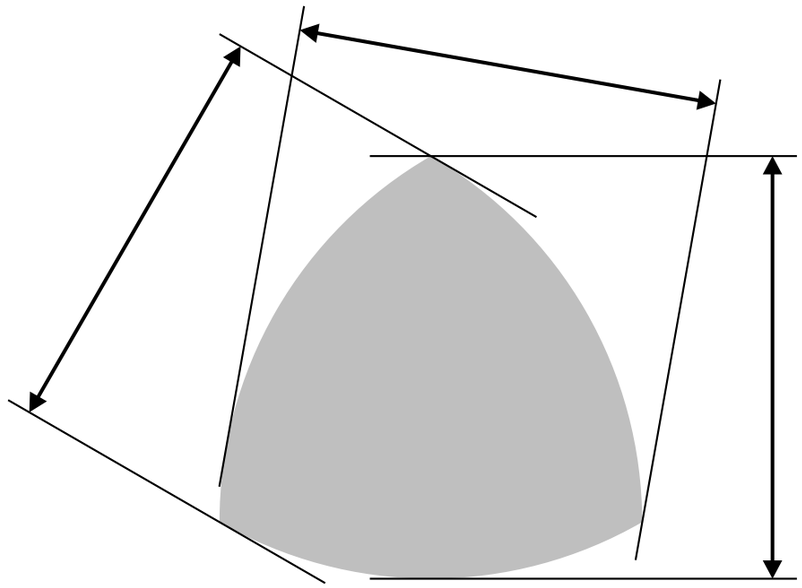
Integral equations and inversion problems

Inverse problems

Craig, Brown: *Inverse problems in astronomy*, Adam Hilger 1986.

Example: we observe an asteroid at opposition. Regardless of its rotation its brightness does not vary with time. Thus we can assume that the projected area remains constant? Does the asteroid have rotational symmetry?

Not necessarily; it could be e.g. a Reuleaux triangle:



Similar problems are very common in astronomy, since the observational geometry cannot be changed.

Direct method: Construct a model, for the observed phenomenon e.g. by simulating it on a computer. Calculate what kind of observations are obtained when the parameters of the model are given certain values. Change the parameters until the calculated values correspond to the observations.

Problem: is the solution unique?

Another problem: stability of the solution. Often we observe an integrated quantity that is not strongly affected even by big changes of the parameter values. Even small errors in observations may lead to very different values for the parameters.

Inverse method: The parameters describing the object are computed from the observations. Often difficult mathematically.

Observations often involve integration of the real quantity, which will destroy information. (Even if we know that $x + y = 1$, we still have no idea about the numbers x and y).

Integration is a smoothing operation, and the result is insensitive to small deviations in the quantity being integrated. The inverse operation requires some kind of differentiation, which will enhance noise. Therefore small observational errors may lead to large errors in the model derived from the observations.

Integral equations

Integral equations contain an unknown function and its integral, e.g.

$$f(x) = \int f(t)\sin(x+t)\mathrm{d}t.$$

The known function in the integral (here $\sin(x+t)$ is the *kernel* of the equation.

Sometimes the equation can be converted to a differential equation.

Often the measured quantity g and the property to be determined f are related by an equation

$$g(x) = \int_a^b \mathcal{K}(x,t) f(t)\, dt.$$

This is *Fredholm's integral equation* of the first kind. The equation of the second is

$$f(x) = g(x) + \lambda \int_a^b \mathcal{K}(x,t) f(t)\, dt.$$

In these the limits of integration are fixed. If the integration range depends on x, we have *Volterra's integral equations* of the first and second kind.

If λ in the equation of the second kind is small, the functions f and g are almost identical, and the integral term describes a small perturbation. The equation can then be solved by iteration. We can start by assuming e.g. $f_0 = g$.

$$f_{k+1}(x) = g(x) + \lambda \int \mathcal{K}(x,t) f_k(t)\, dt.$$

This method does not work if λ is big. Then the contribution of the original function is small, and the smoothing effect of the integral will dominate. The equations approaches an equation of the first kind.

Fredholm's equations of the first kind are often hard to solve. In a discrete case such equations can be written in a matrix form:

$$\mathbf{g} = \mathbf{H}\mathbf{f}.$$

For example, in image processing $\mathbf{H}$ is essentially a function describing the shape of the image of a point source (PSF, point spread function). If the signal is passed unaffected, $\mathbf{H}$ is an identity matrix. If the signal spreads out a little, $\mathbf{H}$ is a band matrix, with nonzero elements in the diagonal and close to it. The more the signal is spread out, the more nonzero elements there are in $\mathbf{H}$ and the smaller the diagonal elements become.

When the signal has completely spread out, all elements of the matrix $\mathbf{H}$ are roughly of the same magnitude. Thus the matrix is nearly singular, and inverting it is a very unstable operation. This corresponds also to the physical interpretation of the situation: signals are so badly spread out and mixed that the original image is hardly visible.

Example:

$$\int_0^1 (x+y)f(y)\,dy = x.$$

If we substitute a trial solution $f(y) = ay + b$, the lefthand side will be

$$\left(\frac{1}{2}a + b\right)x + \frac{1}{3}a + \frac{1}{2}b.$$

Since this has to be identically x, we must take $a = -6$ and $b = 4$. Therefore the solution of the equation is

$$f(y) = 4 - 6y.$$

Let's try to solve the equation numerically. Using the trapezoidal rule the integral is

$$\frac{h}{2}\big[(x)f(0) + 2(x+h)f(h) + \cdots \\ + 2(x+(n-1)h)f((n-1)h) + (x+1)f(1)\big].$$

The expression contains $n+1$ unknowns, namely the values of the function f: $f(0)$, $f(h)$, …, $f(1)$. The equation can be written separately for different values of x to get enough equations to solve all these unknowns.

Let's take $h = 0.2$; then there are six unknown values of the function. We'll use the values 0, 0.2, ..., 1 for x.

$$\frac{0.2}{2}\big[0 + 0.4f(0.2) + 0.8f(0.4) + 1.2f(0.6) + 1.6f(0.8) + f(1)\big] = 0,$$

$$\dots$$

$$\frac{0.2}{2}\big[1 + 2.4f(0.2) + 2.8f(0.4) + 3.2f(0.6) + 3.6f(0.8) + 2f(1)\big] = 1$$

or in matrix form

$$\frac{0.2}{2}\begin{pmatrix} 0 & 0.4 & 0.8 & 1.2 & 1.6 & 1.0 \\ 0.2 & 0.8 & 1.2 & 1.6 & 2.0 & 1.2 \\ 0.4 & 1.2 & 1.6 & 2.0 & 2.4 & 1.4 \\ 0.6 & 1.6 & 2.0 & 2.4 & 2.8 & 1.6 \\ 0.8 & 2.0 & 2.4 & 2.8 & 3.2 & 1.8 \\ 1.0 & 2.4 & 2.8 & 3.2 & 3.6 & 2.0 \end{pmatrix} \begin{pmatrix} f(0) \\ f(0.2) \\ f(0.4) \\ f(0.6) \\ f(0.8) \\ f(1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0.2 \\ 0.4 \\ 0.6 \\ 0.8 \\ 1 \end{pmatrix}.$$

The coefficient matrix is nearly singular. The solution may have nothing to do with reality.

Regularisation

A singular matrix means a situation in which information has been lost. There is no way to recover the lost information, but we can try make the inversion of the matrix more stable.

The idea of regularisation is to increase the dominance of the diagonal elements.

Let's write the equation of the previous example as

$$\lambda f(x) + \int_0^1 (x+y) f(y)\, dy = x,$$

where the regularisation parameter λ is a positive constant. When this is discetised in

the same way as before, we get a set of euations

$$\begin{aligned}
&\lambda f(0) + \frac{0.2}{2}\big[0 + 0.4f(0.2) + 0.8f(0.4) + \\
&\qquad\qquad 1.2f(0.6) + 1.6f(0.8) + f(1)\big] = 0, \\
&\qquad \cdots \\
&\lambda f(1) + \frac{0.2}{2}\big[1 + 2.4f(0.2) + 2.8f(0.4) + \\
&\qquad\qquad 3.2f(0.6) + 3.6f(0.8) + 2f(1)\big] = 1.
\end{aligned}$$

The diagonal elements are now bigger depending on the parameter λ. The sensitivity to errors (condition number) is smaller, and the equations can be solved without problems.

If λ is large, the new equation differs considerably from the original one; also its solution is very different from the solution of the original equation.

Obviously there is some optimal value of the regularisation parameter λ, yielding a solution not too different from that of the original equation but leading to a problem that is not yet too badly disturbed by the singularity of the matrix.

Solutions of the example corresponding to different values of λ:

λ	cond($\mathbf{H}$)	$f(0)$	$f(1)$
0.05	43	9.1	-5.6
0.01	123	4.3	-2.1
0.005	242	4.0	-1.9
0.001	1200	3.8	-1.8

The best result is obtained when $\lambda = 0.005$. If the parameter is larger, the equation differs too much from the original one; if the parameter is smaller, the matrix begins to become almost singular.

A possible solution is to convert the original problem to an optimisation problem. The equation $\mathbf{Hf} = \mathbf{g}$ may not have a solution at all, or the solution may be very sensitive to observational errors. Yet we can always find such a solution f that will minimise the norm

$$\| \mathbf{Hf} - \mathbf{g} \| .$$

If the coefficient matrix is singular, the solution is not unambiguous. This problem can be avoided by regularisation and minimising the norm

$$\| \mathbf{Hf} - \mathbf{g} \|^2 + \lambda^2 \| \mathbf{f} \|^2 .$$

As a solution to this optimisation problem we'll get both the vector $\mathbf{f}$ and the parameter λ.