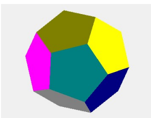
Zero-points for functions of several variables Newton-Rapson's method

Minimum for functions of several variables, Using the method of steepest descent

This is an article from my home page: www.olewitthansen.dk



Ole Witt-Hansen 2018

Contents

1. Formulation of the problem

Let there be given a vector function: $\underline{\underline{f}}(\underline{\underline{x}})\quad R^n \to R^n$.

We wish to solve the system of the *n* non-linear equations: $\underline{\underline{f}}(\underline{\underline{x}}) = \underline{\underline{b}} \quad \Leftrightarrow \quad \underline{\underline{f}}(\underline{\underline{x}}) - \underline{\underline{b}} = \underline{\underline{0}}$

If written out:

$$\begin{matrix} f_1(x_1,x_2,x_3,.....x_n) = b_1 & & f_1(x_1,x_2,x_3,.....x_n) - b_1 = 0 \\ f_2(x_1,x_2,x_3,.....x_n) = b_2 & & f_2(x_1,x_2,x_3,.....x_n) - b_2 = 0 \\ f_3(x_1,x_2,x_3,.....x_n) = b_3 & \Leftrightarrow & f_3(x_1,x_2,x_3,.....x_n) - b_3 = 0 \\ \ldots & & \ldots \\ f_n(x_1,x_2,x_3,.....x_n) = b_n & & f_n(x_1,x_2,x_3,.....x_n) - b_n = 0 \end{matrix}$$

Solving a system of equations $\underline{\underline{f}}(\underline{\underline{x}}) = \underline{\underline{b}}$ may always be reduced to finding the zero points of $\underline{\underline{f}}(\underline{\underline{x}}) - \underline{\underline{b}}$.

2.Newton Rapson's method

Finding numerically the zero points for a real function of one variable is done for example by Newton-Rapson's method. We shall assume that we start at the function value at x_0, belonging to an interval of the function which leads monotonically to the zero point, a condition, which is necessary for the method to work.

We therefore assume that, $f(x_0)$ is the first approximation to $f(x) = 0$, and we then seek a better approximation: $x_1 = x_0 + \delta_0$. To the first order in δ_0, we have:

$$f(x_1) = 0 \quad \Leftrightarrow \quad f(x_0 + \delta_0) = f(x_0) + f'(x_0)\delta_0 = 0 \quad \Rightarrow \quad \delta_0 = -\frac{f(x_0)}{f'(x_0)}$$

This iteration procedure is repeated with x_1:

$$f(x_2) = 0 \quad \Leftrightarrow \quad f(x_1 + \delta_1) = f(x_1) + f'(x_1)\delta_1 = 0 \quad \Rightarrow \quad \delta_1 = -\frac{f(x_1)}{f'(x_1)}$$

Then we determine $x_2 = x_1 + \delta_1$ and so on, until required accuracy is obtained.

3. Generalisation to vector functions of several variables..

It is almost trivial to generalize Newton-Rapson's method to vector function of several variables, apart from the fact that you can no longer do the calculations without a computer.

We settle for writing only one of the *n* equations, starting from: $f_i(x_1, x_2, x_3, ...x_n) = 0$

$$f_i(x_1 + \delta_1, x_2 + \delta_2, x_3 + \delta_3, ... , x_n + \delta_n) = 0 \quad \Leftrightarrow \quad f_i(x_1, x_2, x_3, ...x_n) + \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\delta_j = 0$$

The increments $\delta_1, \delta_2, \delta_3, \ldots \delta_n$ can then for each step be calculated from the solution to the linear system of equations below:

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & . & . & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & . & . & \frac{\partial f_2}{\partial x_n} \\ . & . & . & . & . \\ . & . & . & . & . \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & . & . & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ . \\ . \\ \delta_n \end{pmatrix} = - \begin{pmatrix} f_1(x_1, x_2, \dots x_n) \\ f_2(x_1, x_2, \dots x_n) \\ \\ \\ f_n(x_1, x_2, \dots x_n) \end{pmatrix}$$

The new trial point can then be computed as $x_i + \delta_i$, and where after the iteration is continued. In the (Danish) computer program: "Matematisk analyse", to be found on my home page, this method is used.

1. Formulation of the problem

To determine a local minimum for a function (of several variables) is a well known numerical problem, and there exists several numerical methods.
Here is presented a method, which as a starting point requires knowledge of the partial derivatives of the function. If the analytic expressions do not exists, the partial derivatives may readily be calculated numerically e.g. by Aitken's formula. In all cases the method requires numerical calculations. Let there be given a real differentiable function of *n* variables:

$$y = f(x_1, x_2, x_3, .., x_n)$$

The extrema (max, min, actually the zero points for the differential quotients) for the function are given by solving the equations:

$$\frac{\partial f}{\partial x_1} = 0 \ , \ \frac{\partial f}{\partial x_2} = 0 \ , \ \frac{\partial f}{\partial x_3} = 0 \ , ... \frac{\partial f}{\partial x_n} = 0$$

In general it is not possible to solve these *n* equations analytically, having *n* unknowns.

We write: $\overline{x} = (x_1, x_2, x_3 ... x_n)$ and correspondingly for the other variables.
We take the starting point at $\overline{x_0}$ and move in the direction of $\overline{a}$, where $\overline{a} = (a_1, a_2, a_3, \dots, a_n)$ is a unit vector. We then look at the real function $y = f(t)$, of a single real variable *t*:

$$f(t) = f(\overline{x_0} + t\overline{a}) = f(x_1 + ta_1, x_2 + ta_2, x_3 + ta_3, \dots, x_n + ta_n)$$

2. Method of steepest descend. Extremum with a side condition

The task is now to determine $\overline{a}$, such that we get the steepest descent, and we do so by finding the extremum for $f'(t)$ from $\overline{x_0}$, that is, for $t = 0$. So we find the differential quotient of $f(t)$ for $t = 0$.

$$f'(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} a_i = F(\overline{a})$$

$F(\overline{a})$ can then be considered as a function of the *n* variables $(a_1, a_2, a_3, \dots, a_n)$, and the task is then to determine the extremum for this function.

It could appear as the same problem that we started out with, but the crucial difference is that $F(\overline{a})$ is a linear function of $\overline{a} = (a_1, a_2, a_3, \dots, a_n)$.
Since $\overline{a}$ is supposed to be a unit vector, we may determine the extremum under the *side condition* $|\overline{a}| = 1$ or $\sum_{i=1}^{n} a_i^{\ 2} = 1$.

We put: $G(\overline{a}) = \sum_{i=1}^{n} a_i^{\ 2}$, and according to the theory of extremum with a side condition, this can be done by seeking extremum for the function: $F(\overline{a}) + \lambda G(\overline{a})$, where λ is a so called Lagrange multiplier, which has to be determined from the border conditions of the problem.

This is accomplished by differentiating the function:

$$F(\overline{a}) + \lambda G(\overline{a}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} a_i + \lambda \sum_{i=1}^{n} a_i^{\ 2}$$

$$\frac{\partial F}{\partial a_i} + \lambda \frac{\partial G}{\partial a_i} = 0 \quad \Leftrightarrow \quad \frac{\partial f}{\partial x_i} + 2\lambda a_i = 0$$

From which follows: $a_i = -\frac{1}{2\lambda} \frac{\partial f}{\partial x_i}$

λ is then determined by the normality condition: $\sum_{i=1}^{n} a_i^{\ 2} = 1$, which gives:

$$4\lambda^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \quad \Leftrightarrow \quad \lambda = \tfrac{1}{2} \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2}$$

3. Numerical methods

We have thus found the steepest descent, along the vector $(a_1, a_2, a_3, \ldots, a_n)$ but we still do not know, how far we should move in the direction of *a*. We could take a small step in the direction of $(a_1, a_2, a_3, \ldots, a_n)$, and repeat the calculation, but this is not really usable.

Instead we might assume that the function is approximately parabolic in the vicinity of the minimum, and then determine the function value in the position of the bottom point.
If we have found the minimum, we are finished, but otherwise we repeat the procedure with the bottom point as the new starting point.

To find the equation for the parabola, we need to know the first and second derivative of *f*(*0*)
A quadratic polynomial, passing through the point (t_0, $f(t_0)$), and has the same 1. and. 2. derivative as the function at $t = 0$, can be written:

$$y = \tfrac{1}{2} f''(0) t^2 + f'(0) t + f(0)$$

The first derivative of a function in 0 can, to the second order in *h,* be calculated numerically as:

$$f'(0) = \frac{f(\frac{h}{2}) - f(-\frac{h}{2})}{h}$$

It then follows:

$$f'(-\tfrac{h}{2}) = \frac{f(0) - f(-h)}{h} \qquad f'(\tfrac{h}{2}) = \frac{f(h) - f(0)}{h} \qquad f''(0) = \frac{f'(\tfrac{h}{2}) - f'(-\tfrac{h}{2})}{h}$$

Notice that we only need to know the function value in the three points; *f*(-*h*), *f*(0) and *f*(*h*).

$f(0) = f(\bar{x}_0)$ $\quad f'(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} a_i$, where the partial derivatives should be computed according to the formula.

$$\left.\frac{\partial f}{\partial x_i}\right|_{t=0} = \frac{f(x_1, x_2, \ldots x_i + \tfrac{h}{2}, \ldots x_n) - f(x_1, x_2, \ldots x_i - \tfrac{h}{2}, \ldots x_n)}{h}$$

Once the first and the second derivatives are calculated, we may insert in the formula

$$y = \tfrac{1}{2} f''(0) t^2 + f'(0) t + f(0)$$

Differentiating we find: $y' = f''(0)t + f'(0)$. The bottom point is at $y' = 0$, from which it follows that the bottom point is at $t = -\frac{f'(0)}{f''(0)}$

If $f''(0)$ is not negative, as is required for a minimum, then we are either far away from a min, or the function may have a max (highly unlikely).

The considerations above are the work processed, when I worked still as a student in 1970.
It turned out, in most cases, to be very efficient, compared to the traditional standard programs.

The method is used in the (Danish) program from 1996. “Matematisk analyse”, to be found on my home page. www.olewitthansen.dk