ACTUARIAL RESEARCH CLEARING HOUSE
1985 VOL. 1

NON-LINEAR PARAMETER ESTIMATION

E.S. ROSENBLOOM

Department of Actuarial and Management Sciences
University of Manitoba
Winnipeg, Manitoba
R3T 2N2

ABSTRACT. This note points out that the non-linear parameter estimation technique presented in the Part 5 Study Note is in fact the Gauss - Newton method for the non-linear least-square problem.

In a Society of Actuaries 1984 Study Note for Part 5, Richard L. London [3,6.3.4] presents an algorithm for the non-linear parameter estimates of A, B and c in Makeham's mortality formula

$$\mu_x = A + Bc^x. \qquad (1)$$

However, this algorithm is simply a particular example of a well known method in optimization called the Gauss-Newton method.

The Gauss-Newton method is used in solving the non-linear least-square problem. The least-square problem concerns the unconstrained minimization of a function which is the sum of squares of non-linear functions. That is

$$\text{minimize} \quad f(\underline{x}) = \tfrac{1}{2} \sum_{i=1}^{m} [f_i(\underline{x})]^2. \tag{2}$$

$$\underline{x} = (x_1 \; \ldots\ldots \; x_n)^T$$

(The ½ is included in order to avoid a factor of 2 in the derivatives).

The least-square problem is often used in non-linear parameter estimation. Each individual function $f_i(x)$ is defined as $y_i - \phi(\underline{x}, t_i)$ where $\phi(\underline{x}, t)$ represents a desired model function with an independent variable t and parameters $\underline{x} = (x_1, x_2 \ldots x_n)^T$ to be estimated. The y_i's are data points which may be subject to experimental error.

A classic method for solving the least-square problem is the Newton-Raphson method for finding a root of the gradient of $f(\underline{x})$. Newton's method employs both the gradient $\underline{g}(\underline{x})$ and the Hessian $G(\underline{x})$. Given an estimate $\underline{x}^k$ of the optimal solution, Newton's method obtains a new estimate, $\underline{x}^{k+1}$, by solving the system of linear equations

$$G(\underline{x}^k)\ \underline{d}^k = -\underline{g}(\underline{x}^k) \tag{3}$$

for $\underline{d}^k$ and setting

$$\underline{x}^{k+1} = \underline{x}^k + \underline{d}^k. \tag{4}$$

In the particular case where $f(\underline{x})$ is given by (2), the gradient $g(\underline{x})$ can be written as

$$g(\underline{x}) = (J(\underline{x}))^T\ \underline{F}(\underline{x}) \tag{5}$$

where $J(\underline{x})$ is the $m \times n$ Jacobian matrix and $F(\underline{x}) = (f_1(\underline{x}), f_2(\underline{x}), \ldots f_m(\underline{x}))^T$. The Hessian $G(\underline{x})$ can be written as

$$G(\underline{x}) = J(\underline{x})^T J(\underline{x}) + \sum_{i=1}^{m} f_i(\underline{x}) \; G_i(\underline{x}) \tag{6}$$

where $G_i(\underline{x})$ is the Hessian of $f_i(\underline{x})$.

The major disadvantage of Newton's method is the enormous amount of calculation required in the evaluation of all the second derivatives in $\sum_{i=1}^{m} f_i(\underline{x}^k) \; G_i(\underline{x}^k)$. However, if the second derivative term $\sum_{i=1}^{m} f_i(\underline{x}^k) \; G_i(\underline{x}^k)$ is dropped, the Newton direction can be approximated by the solution of the system of equations

$$J(\underline{x}^k)^T \; J(\underline{x}^k) \; \underline{p}^k = -J(\underline{x}^k)^T \; \underline{F}(\underline{x}) \tag{7}$$

This solution $\underline{p}^k$ is called the Gauss-Newton direction. The new estimate of the optimal solution is obtained by setting

$$\underline{x}^{k+1} = \underline{x}^k + \underline{p}^k. \tag{8}$$

The algorithm is called the Gauss-Newton method and in the particular example of parameter estimation in Makeham's mortality formula it reduces to the algorithm derived in [3] with $\underline{x} = (A, B, c)^T$ and $f_i(\underline{x}) = \sqrt{2} \; w_i^{\frac{1}{2}}(y_i - A - Bc^{t_i})$.

Under certain conditions the Gauss-Newton method can achieve a quadratic rate of convergence despite the fact that no second derivative information is used. However, difficulties can arise if the Jacobian $(J(\underline{x}^k))$ is not of full column rank or if the dropped second derivative information $\sum_{i=1}^{m} f_i(\underline{x})\, G_i(\underline{x})$ is not negligible.

A complete discussion of the Gauss-Newton and other least-squared methods can be found in many books for example [1] and [2].

REFERENCES

[1] J.E. Dennis Jr., "Non-Linear Least Squares and Equations", in *The State of the Art in Numerical Analysis* (D.A.H.Jacobs, ed.) pp. 269-312, Academic Press, London, 1977.

[2] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, New York, 1981.

[3] R.L. London, *Part 5 Study Note 54-01-84, Graduation: The Revision of Estimates*, Society of Actuaries, 1984.