4. Non-linear least squares

4.1 INTRODUCTION

The methods of chapter 3 have been designed to be effective on all sufficiently smooth objective functions. However it is sometimes possible to devise methods that are more efficient should the objective function have a special form. By far the most important of these forms encountered in practice is a sum of the squares of other non-linear functions:

\[
F(x) = \sum_{i=1}^{m} f_i^2(x) \quad (4.1.1)
\]

The minimization of functions of this kind is called **non-linear least squares**. Note immediately that the objective function can never take a negative value in these problems. It is convenient to gather the functions \( f_i \) together in vector form

\[
f(x) = \begin{bmatrix}
f_1 \\
f_2 \\
\vdots \\
f_m
\end{bmatrix}
\quad (4.1.2)
\]

when we can write

\[
F(x) = f^T(x) f(x) \quad (4.1.3)
\]

It is the aim of this chapter to discuss some of the most effective methods currently available for minimizing functions of the form (4.1.3). But first let us see some ways in which such functions arise in practice.

4.1.1 Non-linear regression

A problem commonly encountered is to fit a mathematical function to some experimental data by varying parameters in the function. A specific functional form \( E(x, z) \) will have been selected on theoretical or empirical grounds, the independent variables being the elements of \( z \). These are set experimentally at \( z_i \) and some
property \( y_i \) of the system being studied is measured. The process is then repeated for \( i = 1(1)m \). \( x \) is a vector of \( n \) parameters which are to be adjusted until the best fit of \( E \) to the data is found. These then are the variables of our optimization problem. When \( E \) is non-linear in \( x \), this type of problem is called \textit{non-linear regression}. Generally \( m > n \) and the system is then said to be \textit{overdetermined}; the case \( m < n \) characterises an \textit{underdetermined} system.

It is necessary to define more precisely what we mean by the best fit. For most overdetermined systems it will not be possible to find \( x \) such that \( E \) passes through all of the data points. The difference between \( y_i \) and the value \( E(x, z_i) \) predicted by the function is called a \textit{residual} \( r_i \). Thus

\[
 r_i(x) = E(x, z_i) - y_i \quad i = 1(1)m
\]

The \textit{least squares best fit} is obtained by minimizing the function \( R(x) \), which is the sum of the squares of the residuals

\[
 R(x) = \sum_{i=1}^{m} r_i^2(x)
\]

with respect to \( x \). Squares are taken to avoid cancellation between residuals of opposite sign and this is therefore a non-linear least squares problem. Note that \( F(x^*) \) is zero only if a perfect fit of \( E \) to the data can be found.

A closely related problem is \textit{weighted least squares}. Here the function to be minimized is of the form

\[
 R(x) = \sum_{i=1}^{m} w_i r_i^2(x)
\]

where the scalars \( w_i \) are positive \textit{weighting factors}. In this way a varying importance can be attached to the data values.

4.1.2 Simultaneous non-linear equations

Another common problem is to find the solution of a system of \( n \) non-linear equations

\[
f_1(x) = 0 \\
\vdots \\
f_n(x) = 0
\]

in \( n \) unknowns \( x \). In vector notation equations (4.1.7) assume the more compact form

\[
f(x) = 0
\]