Similarly, it can be shown that the Hessian $\boldsymbol{H}(\boldsymbol{\beta})$ has typical element

$$H_{ij}(\boldsymbol{\beta}) = -\frac{2}{n}\sum_{t=1}^{n}\Big(\big(y_t - x_t(\boldsymbol{\beta})\big)\frac{\partial X_{ti}(\boldsymbol{\beta})}{\partial \beta_j} - X_{ti}(\boldsymbol{\beta})X_{tj}(\boldsymbol{\beta})\Big). \tag{6.46}$$

When this expression is evaluated at $\boldsymbol{\beta}_0$, it is asymptotically equivalent to

$$\frac{2}{n}\sum_{t=1}^{n} X_{ti}(\boldsymbol{\beta}_0)X_{tj}(\boldsymbol{\beta}_0). \tag{6.47}$$

The reason for this asymptotic equivalence is that, since $y_t = x_t(\boldsymbol{\beta}_0) + u_t$, the first term inside the large parentheses in (6.46) becomes

$$-\frac{2}{n}\sum_{t=1}^{n}\frac{\partial X_{ti}(\boldsymbol{\beta})}{\partial \beta_j}\,u_t. \tag{6.48}$$

Because $x_t(\boldsymbol{\beta})$ and all its first- and second-order derivatives belong to Ω_t, the expectation of each term in (6.48) is 0. Therefore, by a law of large numbers, expression (6.48) tends to 0 as $n \to \infty$.

Gauss-Newton Methods

The above results make it clear that a natural choice for $\boldsymbol{D}(\boldsymbol{\beta})$ in a quasi-Newton minimization algorithm based on (6.43) is

$$\boldsymbol{D}(\boldsymbol{\beta}) = 2n^{-1}\boldsymbol{X}^{\top}(\boldsymbol{\beta})\boldsymbol{X}(\boldsymbol{\beta}). \tag{6.49}$$

By construction, this $\boldsymbol{D}(\boldsymbol{\beta})$ is positive definite whenever $\boldsymbol{X}(\boldsymbol{\beta})$ has full rank. Substituting equations (6.49) and (6.45) into equation (6.43) yields

$$\begin{aligned}\boldsymbol{\beta}_{(j+1)} &= \boldsymbol{\beta}_{(j)} + \alpha_{(j)}\big(2n^{-1}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{X}_{(j)}\big)^{-1}\big(2n^{-1}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{y} - \boldsymbol{x}_{(j)})\big)\\ &= \boldsymbol{\beta}_{(j)} + \alpha_{(j)}\big(\boldsymbol{X}_{(j)}^{\top}\boldsymbol{X}_{(j)}\big)^{-1}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{y} - \boldsymbol{x}_{(j)}).\end{aligned} \tag{6.50}$$

The classic **Gauss-Newton method** would set $\alpha_{(j)} = 1$, so that

$$\boldsymbol{\beta}_{(j+1)} = \boldsymbol{\beta}_{(j)} + \big(\boldsymbol{X}_{(j)}^{\top}\boldsymbol{X}_{(j)}\big)^{-1}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{y} - \boldsymbol{x}_{(j)}), \tag{6.51}$$

but it is generally better to use a good one-dimensional search routine to choose α optimally at each iteration. This modified type of Gauss-Newton procedure often works quite well in practice.

The second term on the right-hand side of (6.51) can most easily be computed by means of an **artificial regression** called the **Gauss-Newton regression**, or **GNR**. This artificial regression can be expressed as follows:

$$\boldsymbol{y} - \boldsymbol{x}(\boldsymbol{\beta}) = \boldsymbol{X}(\boldsymbol{\beta})\boldsymbol{b} + \text{residuals}. \tag{6.52}$$