it must be equal to minus the information matrix for all sample sizes and not just asymptotically. Thus we may replace $n\mathcal{H}$ by $-\boldsymbol{I}$:

$$\ell(\boldsymbol{\theta}) = \hat{\ell} - \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \boldsymbol{I}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}). \tag{13.07}$$

Evaluating this expression at $\tilde{\boldsymbol{\theta}}$ and substituting into the definition of the LR statistic, (13.06), we see that, when the loglikelihood function is quadratic, the LR statistic can be rewritten as

$$LR = (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^\top \boldsymbol{I}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}). \tag{13.08}$$

Now consider the LM statistic. From (13.07), it is easy to see that the gradient $\tilde{\boldsymbol{g}}$ is just $-\boldsymbol{I}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$. Then, from (13.04), it follows that LM is equal to $(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^\top \boldsymbol{I}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$, which is simply expression (13.08). Thus we see that the LM and LR statistics are numerically equal when the loglikelihood function is quadratic.

Proving that these two statistics are equal to the Wald statistic in this special case is a little bit harder. We begin by making another assumption, one which, as we will see later, does not in fact entail any loss of generality. It is that the restrictions associated with the null hypothesis take the form

$$\boldsymbol{\theta}_2 = \boldsymbol{0}. \tag{13.09}$$

Here we have partitioned the parameter vector as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1 \,\vdots\, \boldsymbol{\theta}_2]$, with $\boldsymbol{\theta}_2$ an r–vector and $\boldsymbol{\theta}_1$ therefore a $(k-r)$–vector. We can also partition the information matrix so as to conform to this partition of $\boldsymbol{\theta}$:

$$\boldsymbol{I} = \begin{bmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix}.$$

With $\boldsymbol{\theta}$ and $\boldsymbol{I}$ partitioned in this way, expression (13.07) for the loglikelihood function becomes

$$\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \hat{\ell} - \frac{1}{2}\begin{bmatrix} \boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1 \\ \boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2 \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{12} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1 \\ \boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2 \end{bmatrix}. \tag{13.10}$$

At the restricted MLE, $(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{0})$, the first-order condition for a restricted maximum must be satisfied. By differentiating (13.10) with respect to $\boldsymbol{\theta}_1$ and evaluating the result at $\boldsymbol{\theta}_2 = \boldsymbol{0}$, we find that this first-order condition is

$$\boldsymbol{0} = D_1\ell(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{0}) = -\big(\boldsymbol{I}_{11}(\tilde{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_1) - \boldsymbol{I}_{12}\hat{\boldsymbol{\theta}}_2\big).$$

From this it follows that

$$\boldsymbol{I}_{11}(\tilde{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_1) = \boldsymbol{I}_{12}\hat{\boldsymbol{\theta}}_2. \tag{13.11}$$