The basic steps of the Wilkinson algorithm are as follows:

Algorithm 3.1

Step 1: Compute the residuals **RY**.

Step 2: Use the operator **S**, which Wilkinson calls a *sweep* (not to be confused with the sweep method of Section 11.2.2), to produce a vector of *apparent residuals* $\mathbf{RY} - \mathbf{Z}\hat{\gamma}_G$ $(= \mathbf{SRY})$.

Step 3: Applying the operator **R** once again, reanalyze the apparent residuals to produce the correct residuals **RSRY**.

If the columns of **Z** are perpendicular to the columns of **X**, then $\mathbf{RZ} = \mathbf{Z}$ and, by (3.34),

$$\begin{aligned} \mathbf{RSR} &= \mathbf{R}(\mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{RZ})^{-1}\mathbf{Z}')\mathbf{R} \\ &= \mathbf{R} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R} \\ &= \mathbf{SR}, \end{aligned}$$

so that step 3 is unnecessary. We see later (Section 3.9.3) that the procedure above can still be used when the design matrix **X** does not have full rank.

By setting **X** equal to the first k columns of **X**, and **Z** equal to the $(k+1)$th column $(k = 1, 2, \ldots, p-1)$, this algorithm can be used to fit the regression one column of **X** at a time. Such a stepwise procedure is appropriate in experimental design situations because the columns of **X** then correspond to different components of the model, such as the grand mean, main effects, block effects, and interactions, and some of the columns are usually orthogonal. Also, the elements of the design matrix **X** are 0 or 1, so that in many standard designs the sweep operator **S** amounts to a simple operation such as subtracting means, or a multiple of the means, from the residuals.

EXERCISES 3f

Thm 3.6(iii) + use $\hat{\gamma}_G = MZ^TRY$ by (i)

1. Prove that gives LHS $= \hat{\gamma}_G^T Z^T RY$

$$\mathbf{Y}'\mathbf{RY} - \mathbf{Y}'\mathbf{R}_G\mathbf{Y} = \sigma^2\hat{\gamma}_G'\left(\mathrm{Var}[\hat{\gamma}_G]\right)^{-1}\hat{\gamma}_G = \hat{\gamma}_G^T M^{-1} \hat{\gamma}_G$$

by (iv)

2. Prove that $\hat{\gamma}_G$ can be obtained by replacing **Y** by $\mathbf{Y} - \mathbf{Z}\gamma$ in $\mathbf{Y}'\mathbf{RY}$ and minimizing with respect to γ. Show further that the minimum value thus obtained is $\mathbf{Y}'\mathbf{R}_G\mathbf{Y}$. Matrix calculus... it's ~~good~~ bfr envelope thm.

3. If $\hat{\beta}_G = (\hat{\beta}_{G,j})$ and $\hat{\beta} = (\hat{\beta}_j)$, use Theorem 3.6(iv) to prove that

$$\mathrm{var}[\hat{\beta}_{G,j}] \geq \mathrm{var}[\hat{\beta}_j].$$

$(X^TX)^{-1} \preceq (X^TX)^{-1} + LML^T$

4. Given that $Y_1, Y_2, \ldots, Y_n$ are independently distributed as $N(\theta, \sigma^2)$, find the least squares estimate of θ.

   (a) Use Theorem 3.6 to find the least squares estimates and the residual sum of squares for the augmented model Start w/ $\hat{\theta} = \bar{Y}$

   $$Y_i = \theta + \gamma x_i + \varepsilon_i \qquad (i = 1, 2, \ldots, n),$$

   RSS $= s^2$

   where the ε_i are independently distributed as $N(0, \sigma^2)$. Then by #2

   (b) Verify the formulae for the least square estimates of θ and γ by differentiating the usual sum of squares.

   Yes, checks out. $(Y - \gamma z)^T R (\cdot\cdot)$

3.8 ESTIMATION WITH LINEAR RESTRICTIONS

As a prelude to hypothesis testing in Chapter 4, we now examine what happens to least squares estimation when there are some hypothesized constraints on the model. We lead into this by way of an example.

EXAMPLE 3.5 A surveyor measures each of the angles α, β, and γ and obtains unbiased measurements Y_1, Y_2, and Y_3 in radians, respectively. If the angles form a triangle, then $\alpha + \beta + \gamma = \pi$. We can now find the least squares estimates of the unknown angles in two ways. The first method uses the constraint to write $\gamma = \pi - \alpha - \beta$ and reduces the number of unknown parameters from three to two, giving the model

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 - \pi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

We then minimize $(Y_1 - \alpha)^2 + (Y_2 - \beta)^2 + (Y_3 - \pi + \alpha + \beta)^2$ with respect to α and β, respectively. Unfortunately, this method is somewhat ad hoc and not easy to use with more complicated models.

An alternative and more general approach is to use the model

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

and minimize $(Y_1 - \alpha)^2 + (Y_2 - \beta)^2 + (Y_3 - \gamma)^2$ subject to the constraint $\alpha + \beta + \gamma = \pi$ using Lagrange multipliers. We consider this approach for a general model below. □

I

$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$= \|Y - \gamma z - \bar{Y} - \gamma \bar{z}\|_2^2$

$\Downarrow$

$$\hat{\gamma} = \frac{Y^T (z - \bar{z})}{\|z - \bar{z}\|_2^2}$$

Then by (i)

$$\hat{\theta} = \bar{Y} - \hat{\gamma} \bar{x}.$$