Ex 5.13

We already know that for a fixed value of λ the function than minimises the penalized residual sum of squares is a natural cubic spline with knots at the unique values of x_i. ($f = S_\lambda y$)

Now suppose we remove the i^{th} observation from the data set and re-fit to the remaining $N-1$ observations.

Firstly, I will show that if we denote $f^{(-i)}$ as the vector with components $f_j^{(-i)} = \hat{f}_j^{(-i)}(x_j)$ (e.g. the predicted values from the spline fit to the $N-1$ observations) and define y^* by:

$$y_j^* = y_j \quad \text{for } i \neq j$$
$$y_i^* = \hat{f}^{(-i)}$$

then:

$$f^{(-i)} = S_\lambda y^*$$

where S_λ is the same smoother matrix as fit the full datset.

Proof//

For a general function g

$$RSS(g, \lambda) = \sum_{i=1}^{N} (y_j^* - g(x_j))^2 + \lambda \int g''(t)^2 dt \quad \text{(eqn 5.9)}$$

$$\geq \sum_{j \neq i}^{N} (y_j^* - g(x_j))^2 + \lambda \int g''(t)^2 dt \quad \text{(simply removing a single observation } i\text{)}$$

$$\geq \sum_{j \neq i}^{N} (y_j^* - \hat{f}^{(-i)}(x_j))^2 + \lambda \int \hat{f}^{(-i)''}(t)^2 dt$$

(Where $\hat{f}^{(-i)}$ is a natural cubic spline with knots at the $N-1$ observations. This was proved in exercise 5.7)

$$= \sum_{j=1}^{N} (y_j^* - \hat{f}^{(-i)}(x_j))^2 + \lambda \int \hat{f}^{(-i)''}(t)^2 dt$$

(Since $y_i^* = \hat{f}^{(-i)}(x_i)$)

Therefore $f^{(-i)} = S_\lambda^{(-i)} y^* = S_\lambda y^*$ as required.

Now we can re-write the expression inside the Parentheses in eqn 5.26

$$\Rightarrow \quad y_i - \hat{f}_\lambda^{(-i)}(x_i) = -\sum_{j=1}^{N} S_\lambda(i,j)\, y_j^* - y_i$$

$$= \sum_{j \neq i} y_i - S_\lambda(i,i)\, y_i^* - S_\lambda(i,j)\, y_j$$

$$= \sum_{j \neq i} y_i - S_\lambda(i,i)\, \hat{f}^{(-i)}(x_i) - S_\lambda(i,j)\, y_j \quad \left(\text{by def}^n \text{ of } y_i^*\right)$$

$$= \sum_{j=1}^{N} y_i - S_\lambda(i,j)\, y_j + S_\lambda(i,i)\left\{ y_i - \hat{f}^{(-i)}(x_i) \right\}$$

$$= y_i - \hat{f}_\lambda(x_i) + S_\lambda(i,i)\left\{ y_i - \hat{f}^{(-i)}(x_i) \right\}$$

$$\Rightarrow \quad \left(y_i - \hat{f}_\lambda^{(-i)}(x_i)\right)\left(1 - S_\lambda(i,i)\right) = y_i - \hat{f}_\lambda(x_i)$$

$$\Rightarrow \quad y_i - \hat{f}_\lambda^{(-i)}(x_i) = \frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_\lambda(i,i)}$$

Proving that 5.26 implies 5.27 as required.