SOLUTION TO Ex. 5.13.

We know that $\hat{f}_\lambda$ minimizes

$$\sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int (f''(t))^2 dt$$

If we add the point $(x_0, \hat{f}_\lambda(x_0))$, then we are to minimize

$$\circledast \quad \underbrace{(\hat{f}_\lambda(x_0) - f(x_0))^2 + \sum_{i=1}^{N} (y_i - f(x_i))^2}_{N+1 \text{ terms now!}} + \lambda \int (f''(t))^2 dt$$

Now the old $\hat{f}_\lambda(x_0)$ minimizes both the first term, $(\hat{f}_\lambda(x_0) - f(x_0))^2$, and the second term $\sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int (f''(t))^2 dt$. But then it of course minimizes the sum in $\circledast$.

The goal is now to show that (see (5.26), (5.27))

$$\text{"} \; y_i - \hat{f}_\lambda^{(-i)}(x_i) = \frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_\lambda(i,i)} \; \text{"}$$ for a dataset $(y_i, x_i); i = 1, \dots, N$

To prove this we will instead assume that the data including x_0 is the full data set, while the data without the x_0 corresponds to the "(-i)" case (so we can write "(-0)").

Let now $S(\lambda)$ be the $(N+1) \times (N+1)$ smoothing matrix which includes the data point with x_0.

Let also $\hat{f}_\lambda(x_0)$ have the meaning of the first part of the exercise. Then

$$(**) \quad \hat{f}_\lambda(x_0) = \sum_{j=1}^{N} S_{0j}(\lambda) y_j + S_{00}(\lambda) \underbrace{\hat{f}_\lambda(x_0)}$$

"since this is the "y_0" for the observation at x_0 that was assumed in the beginning of the exercise.

(**) is simply the first element of (5.14), $\hat{\underset{\sim}{f}} = \underset{\sim}{S}_\lambda \underset{\sim}{y}$

Suppose now instead that the observed value of y at x_0 is y_0 (any number), instead of $\hat{f}_\lambda(x_0)$ as in the beginning of the exercise. This would not have changed the matrix $S(\lambda)$ [$(N+1)\times(N+1)$ matrix] since this depends only on the x-values and not y.

But then the $\hat{f}_\lambda(x_0)$ in (**) can be interpreted as $\hat{f}_\lambda^{(-0)}(x_0)$ in the full $(N+1)\times(N+1)$ case, and hence (**) implies

$$\hat{f}_\lambda^{(-0)}(x_0) = \sum_{j=1}^{N} S_{0j}(\lambda) y_j + S_{00}(\lambda) \hat{f}_\lambda^{(-0)}(x_0)$$

$$= \sum_{j=0}^{N} S_{0j}(\lambda) y_j - S_{00}(\lambda) y_0 + S_{00}(\lambda) \hat{f}_\lambda^{(-0)}(x_0)$$

$$= \underbrace{\hat{f}_\lambda(x_0)}_{\text{in the } (N+1) \text{ point model}} - S_{00}(\lambda) y_0 + S_{00}(\lambda) \hat{f}_\lambda^{(-0)}(x_0)$$

Hence

$$y_0 - \hat{f}_\lambda^{(-0)}(x_0) = y_0 - \hat{f}_\lambda(x_0) + S_{00}(\lambda)\left(y_0 - \hat{f}_\lambda^{(-0)}(x_0)\right)$$

so

$$y_0 - \hat{f}_\lambda^{(-0)}(x_0) = \frac{y_0 - \hat{f}_\lambda(x_0)}{1 - S_{00}(\lambda)}$$

which is what is needed for (5.27).