Ex. 6.10

We wish to show that C_λ is an unbiased estimate for $PE(\lambda)$ e.g.

$$(**) \quad \mathbb{E}(C_\lambda) = \mathbb{E}\left[ASR(\lambda) + \frac{2\sigma^2}{N}\operatorname{trace}(S_\lambda)\right] = PE(\lambda)$$

Start with $\mathbb{E}(ASR(\lambda))$

$$= 1/N\, \mathbb{E}\left[\sum_{i=1}^{N}(y_i - \hat{f}_\lambda(x_i))^2\right] \quad \text{where } y_i = f(x_i) + \varepsilon_i$$

$$= 1/N\, \mathbb{E}[(y - S_\lambda y)^T(y - S_\lambda y)] \quad \text{since } \hat{f} = S_\lambda y$$

$$= 1/N\, \mathbb{E}[y^T(\mathbb{I} - S_\lambda)^T(\mathbb{I} - S_\lambda)y]$$

$$= 1/N\, \mathbb{E}[y^T(\mathbb{I} - S_\lambda)^2 y] \quad \text{since } (\mathbb{I} - S_\lambda) = (\mathbb{I} - S_\lambda)^T$$

Now we will take advantage of some properties of the trace. Firstly, for a 1×1 matrix K_1 we have $K_1 = \operatorname{trace}(K_1)$, so:

$$= 1/N\, \mathbb{E}[\operatorname{trace}(y^T(\mathbb{I} - S_\lambda)^2 y)]$$

$$= 1/N\, \mathbb{E}[\operatorname{trace}((\mathbb{I} - S_\lambda)^2 y y^T)] \quad \text{using the "trace trick"}$$

$$(*) \quad = 1/N \operatorname{trace}((\mathbb{I} - S_\lambda)^2)\, \mathbb{E}(y y^T) \quad S_\lambda \text{ depends only on } X \text{ and } \lambda$$

Now $\mathbb{E}(y y^T)$

$$= \mathbb{E}[(f + \varepsilon)(f + \varepsilon)^T]$$

$$= \mathbb{E}[f f^T + \varepsilon f^T + f \varepsilon + \varepsilon \varepsilon^T]$$

$$= f f^T + \sigma^2 \mathbb{I}$$

Returning to (*) we obtain:

$$= \frac{1}{N}\,\text{trace}\big((\mathbb{I}-S_\lambda)^2\big)(ff^T + \sigma^2\mathbb{I})$$

$$= \frac{1}{N}\,\text{trace}\big((\mathbb{I}-S_\lambda)^2\big)ff^T + \frac{1}{N}\,\text{trace}\big((\mathbb{I}-S_\lambda)^2\big)\sigma^2\mathbb{I}$$

$$LHS = \frac{1}{N} f^T(\mathbb{I}-S_\lambda)^2 f$$ trace trick again in reverse.

$$RHS = \frac{\sigma^3}{N}\,\text{trace}\big((\mathbb{I}-S_\lambda)^2\big)$$

$$= \sigma^2/N\ \text{trace}(\mathbb{I} - 2S_\lambda + S_\lambda^2)$$

$$= \frac{\sigma^2}{N}\left[N - 2\,\text{trace}(S_\lambda) + \text{trace}(S_\lambda^2)\right]$$

$$= \sigma^2 - \frac{2\sigma^2}{N}\,\text{trace}(S_\lambda) + \frac{\sigma^2}{N}\,\text{trace}(S_\lambda^2)$$

Putting this all together we obtain:

$$\mathbb{E}(ASR(\lambda)) = \frac{1}{N} f^T(\mathbb{I}-S_\lambda)^2 f + \sigma^2 - \frac{2\sigma^2}{N}\,\text{trace}(S_\lambda) + \frac{\sigma^2}{N}\,\text{trace}(S_\lambda^2)$$

Therefore, Proving (**) reduces to showing

(***) $$\frac{1}{N} f^T(\mathbb{I}-S_\lambda)^2 f + \sigma^2 + \frac{\sigma^2}{N}\,\text{trace}(S_\lambda^2) = PE(\lambda)$$

and to solve this we need to evaluate $PE(\lambda)$ where

$$PE(\lambda) = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N}(y_i^* - \hat{f}_\lambda(x_i))^2\right]$$

Where this time $y_i^* = f(x_i) + \varepsilon_i^*$

$$= \frac{1}{N}\mathbb{E}[(f+\varepsilon^* - \hat{f})^T(f+\varepsilon^* - \hat{f})]$$

$$= \frac{1}{N}\mathbb{E}\left[(f-\hat{f})^T(f-\hat{f}) + 2\varepsilon^{*T}(f-\hat{f}) + \varepsilon^{*T}\varepsilon^*\right]$$

$$= \frac{1}{N}\mathbb{E}[(f-S_\lambda y)^T(f-S_\lambda y)] + 0 + \sigma^2 \quad \left(\begin{array}{l}\text{since } \mathbb{E}(\varepsilon^*)=0 \\ \text{and } \mathbb{E}(\varepsilon^{*T}\varepsilon^*) = N\sigma^2\end{array}\right)$$

$$= \frac{1}{N}\mathbb{E}[(f-S_\lambda f - S_\lambda\varepsilon)^T(f-S_\lambda f - S_\lambda \varepsilon)] + \sigma^2$$

(using $y = f+\varepsilon$)

$$= \frac{1}{N}\mathbb{E}[(f-S_\lambda f)^T(f-S_\lambda f) - (f-S_\lambda f)S_\lambda\varepsilon$$
$$- (S_\lambda\varepsilon)^T(f-S_\lambda f) + (S_\lambda\varepsilon)^T S_\lambda\varepsilon] + \sigma^2$$

① $$= \frac{1}{N}\mathbb{E}[(f-S_\lambda f)^T(f-S_\lambda f) + (S_\lambda\varepsilon)^T S_\lambda\varepsilon] + \sigma^2 \quad (\text{since } \mathbb{E}(\varepsilon)=0)$$

$$\Rightarrow LHS = \frac{1}{N} f^T(\mathbb{I}-S_\lambda)^2 f \quad (\text{using "trace trick" as before})$$

$$RHS = \frac{1}{N}\mathbb{E}[\varepsilon^T S_\lambda^2 \varepsilon] \quad (\text{since } S_\lambda^T = S_\lambda)$$

$$= \frac{1}{N}\mathbb{E}[\text{trace}(\varepsilon^T S_\lambda^2 \varepsilon)] \quad (\text{as } \varepsilon^T S_\lambda^2 \varepsilon = 1\times 1 \text{ matrix})$$

$$= \frac{1}{N}\mathbb{E}[\text{trace}(S_\lambda^2 \varepsilon\varepsilon^T)] \quad (\text{trace trick})$$

$$= \frac{1}{N}\text{trace}(S_\lambda^2)\sigma^2$$

Therefore ① is equal to:

$$= \frac{1}{N} f^T(\mathbb{I}-S_\lambda)^2 f + \frac{\sigma^2}{N}\text{trace}(S_\lambda^2) + \sigma^2$$

Satisfying (***) as required an thus proving C_λ is an unbiased estimate of $PE(\lambda)$.