Proof of Proposition 1

For L_2Boost with cost function $C(y, u) = (y - u)^2/2$, the negative gradient in stage j is the classical residual vector $u_j = Y - \widehat{F}_{j-1}$ and $\widehat{F}_j = \widehat{F}_{j-1} + \hat{f}_j$ (there is no need for a line search) with $\hat{f}_j = \mathcal{S}u_j$. Thus

$$u_j = Y - \widehat{F}_{j-1} = u_{j-1} - \mathcal{S}u_{j-1} = (I - \mathcal{S})u_{j-1}, \quad j = 1, 2, \ldots, m,$$

implying that $u_j = (I - \mathcal{S})^j Y$ for $j = 1, 2, \ldots, m$. Because $\widehat{F}_0 = \mathcal{S}Y$ we obtain $\widehat{F}_m = \sum_{j=0}^{m} \mathcal{S}(I - \mathcal{S})^j Y$. Using a telescope-sumargument, this equals $(I - (I - \mathcal{S})^{m+1})Y$.

Proof of Proposition 3

The bias term is

$$\begin{aligned}\text{bias}^2(m, \mathcal{S}; f) &= (\mathbb{E}[\mathcal{B}_m Y] - f)^T (\mathbb{E}[\mathcal{B}_m Y] - f) \\ &= ((\mathcal{B}_m - I)f)^T ((\mathcal{B}_m - I)f).\end{aligned}$$

According to (7), using orthonormality of U,

$$\mathcal{B}_m - I = U(D_m - I)U^T = U \operatorname{diag}(-(1 - \lambda_k)^{m+1})U^T.$$

Thus, again by orthonormality of U, the formula for the bias follows.

For the variance, consider

$$\begin{aligned}\operatorname{cov}(\mathcal{B}_m Y) &= \mathcal{B}_m \operatorname{cov}(Y) \mathcal{B}_m^T = \sigma^2 \mathcal{B}_m \mathcal{B}_m^T \\ &= \sigma^2 U \operatorname{diag}((1 - (1 - \lambda_k)^{m+1})^2) U^T,\end{aligned}$$

using (7) and orthonormality of U. Then

$$\operatorname{var}(m, \mathcal{S}; \sigma^2) = tr[\operatorname{cov}(\mathcal{B}_m Y)] = \sigma^2 \sum_{k=1}^{n} (1 - (1 - \lambda_k)^{m+1})^2,$$

again using orthonormality of U.