summary, $F(\mathcal{S}) \sim \mathcal{N}(0, K)$. This induces a predictive model via Bayesian model integration according to

(81)
$$p(y|x;S) = \int p(y|F(x,\cdot))p(F|S)dF,$$

where x is a test point that has been included in a sample (transductive setting). For an i.i.d. sample, the log-posterior for F can be written as

(82)
$$\ln p(F|S) = -\frac{1}{2}F^T K^{-1} F + \sum_{i=1}^{n}[f(x_i, y_i) - g(x_i, F)] + const.$$

Invoking the representer theorem for $\widehat{F}(S) := \arg\max_F \ln p(F|S)$, we know that

(83)
$$\widehat{F}(S)_{iy} = \sum_{j=1}^{n} \sum_{y' \in Y} \alpha_{iy} K_{iy,jy'},$$

which we plug into equation (82) to arrive at

(84)
$$\min_{\alpha} \alpha^T K \alpha - \sum_{i=1}^{n} (\alpha^T K e_{iy'} + \log \sum_{y \in Y} \exp[\alpha^T K e_i y]),$$

where e_{iy} denotes the respective unit vector. Notice that for $f(\cdot) = \sum_{iy} \alpha_{iy} k(\cdot, (x_i, y))$ the first term is equivalent to the squared RKHS norm of $f \in H$ since $\langle f, f \rangle_H = \sum_{i,j} \sum_{y,y'} \alpha_{iy} \alpha_{jy'} \langle k(\cdot, (x_i, y)), k(\cdot, (x_j, y')) \rangle$. The latter inner product reduces to $k((x_i, y), (x_j, y'))$ due to the reproducing property. Again, the key issue in solving (84) is how to achieve spareness in the expansion for $\widehat{F}$.