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

$$p(y|x;\mathcal{S}) = \int p(y|F(x,\cdot))p(F|\mathcal{S})dF, \tag{80}$$

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

$$\ln p(F|\mathcal{S}) = -\frac{1}{2}F^T \boldsymbol{K}^{-1} F + \sum_{i=1}^{n}[f(x_i, y_i) - g(x_i, F)] + const. \tag{81}$$

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

$$\hat{F}(\mathcal{S})_{iy} = \sum_{j=1}^{n}\sum_{y'\in\mathcal{Y}} \alpha_{iy} K_{iy,jy'}, \tag{82}$$

which we plug into equation (**??**) to arrive at

$$\min_{\alpha} \alpha^T \boldsymbol{K}\alpha - \sum_{i=1}^{n}\left(\alpha^T \boldsymbol{K} e_{iy'} + \log \sum_{y\in\mathcal{Y}} \exp[\alpha^T \boldsymbol{K} e_i y]\right), \tag{83}$$

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

4.2 Markov networks and kernels. In Section 4.1 no assumptions about the specific structure of the joint kernel defining the model in equation (70) has been made. In the following, we will focus on a more specific setting with multiple outputs, where dependencies are modeled by a conditional independence graph. This approach is motivated by the fact that independently predicting individual responses based on marginal response models will often be suboptimal and explicitly modeling these interactions can be of crucial importance.

4.2.1 Markov networks and factorization theorem. Denote predictor variables by X, response variables by Y and define $Z := (X, Y)$ with associated sample space $\mathcal{Z}$. We use Markov networks as the modeling formalism for representing dependencies between covariates and response variables, as well as interdependencies among response variables.