Note that even if k happens to be a smooth function (this turns out to be a reasonable assumption), the actual realizations $t(x)$, as drawn from the Gaussian process, need not be smooth at all. In fact, they may be even pointwise discontinuous.

Let us have a closer look at the prior distribution resulting from these assumptions. The standard setting is $\mu = 0$, which implies that we have no prior knowledge about the particular value of the estimate, but assume that small values are preferred. Then, for a given set of $(t(x_1), \ldots, t(x_m)) =: \mathbf{t}$, the prior density function $p(\mathbf{t})$ is given by

$$p(\mathbf{t}) = (2\pi)^{-\frac{m}{2}} (\det K)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{t}^\top K^{-1}\mathbf{t}\right). \tag{16.38}$$

In most cases, we try to avoid inverting K. By a simple substitution,

$$\mathbf{t} = K\boldsymbol{\alpha}, \tag{16.39}$$

RKHS Regularization

we have $\boldsymbol{\alpha} \sim \mathcal{N}(0, K^{-1})$, and consequently

$$p(\boldsymbol{\alpha}) = (2\pi)^{-\frac{m}{2}} (\det K)^{+\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\alpha}^\top K\boldsymbol{\alpha}\right). \tag{16.40}$$

Taking logs, we see that this term is identical to $\Omega[f]$ from the regularization framework (4.80). This result thus connects Gaussian process priors and estimators using the Reproducing Kernel Hilbert Space framework: Kernels favoring smooth functions, as described in Chapters 2, 4, 11, and 13, translate immediately into covariance kernels with similar properties in a Bayesian context.

16.3.3 Simple Hypotheses

Let us analyze in more detail which functions are considered simple by a Gaussian process prior. As we know, hypotheses of low complexity correspond to vectors $\mathbf{y}$ for which $\mathbf{y}^\top K^{-1}\mathbf{y}$ is small. This is in particular the case for the (normalized) eigenvectors v_i of K with large eigenvalues λ_i, since

$$Kv_i = \lambda_i v_i \text{ yields } v_i^\top K^{-1} v_i = \lambda_i^{-1}. \tag{16.41}$$

In other words, the estimator is biased towards solutions with small λ_i^{-1}. This means that the spectrum and eigensystem of K represent a practical means of actually *viewing* the effect a certain prior has on the degree of smoothness of the estimates.

Let us consider a practical example: For a Gaussian covariance kernel (see also (2.68)),

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\omega^2}\right), \tag{16.42}$$

where $\omega = 1$, and under the assumption of a uniform distribution on $[-5, 5]$, we obtain the functions depicted in Figure 16.4 as simple base hypotheses for our estimator. Note the similarity to a Fourier decomposition: This means that the kernel has a strong preference for slowly oscillating functions.