performance. It seems as if kernel methods are defying the curse of dimensionality [29], which requires the number of samples to increase with the dimensionality of the space in which estimation is performed. However, the distribution of capacity in these spaces is not isotropic (cf. Section 2.2.5).

Regularization Operator Viewpoint

The basic idea of the viewpoint described in the present section is simple: rather than dealing with an abstract quantity such as an RKHS, which is defined by means of its corresponding kernel k, we take the converse approach of obtaining a kernel via the corresponding Hilbert space. Unless stated otherwise, we will use $L_2(\mathcal{X})$ as the Hilbert space (cf. Section B.3) on which the regularization operators will be defined. Note that $L_2(\mathcal{X})$ is *not* the feature space $\mathcal{H}$.

Recall that in Section 2.2.2, we showed that one way to think of the kernel mapping is as a map that takes a point $x \in \mathcal{X}$ to a function $k(x, .)$ living in an RKHS. To do this, we constructed a dot product $\langle ., . \rangle_{\mathcal{H}}$ satisfying

$$k(x, x') = \langle k(x, .), k(x', .) \rangle_{\mathcal{H}}. \tag{4.15}$$

Physically, however, it is still unclear what the dot product $\langle f, g \rangle_{\mathcal{H}}$ actually does. Does it compute some kind of "overlap" of the functions, similar to the usual dot product between functions in $L_2(\mathcal{X})$? Recall that, assuming we can define an integral on $\mathcal{X}$, the latter is (cf. (B.60))

$$\langle f, g \rangle_{L_2(\mathcal{X})} = \int_{\mathcal{X}} f \cdot g. \tag{4.16}$$

Main Idea

In the present section, we will show that whilst our dot product in the RKHS is not quite a simple as (4.16), we can at least write it as

$$\langle f, g \rangle_{\mathcal{H}} = \langle \Upsilon f, \Upsilon g \rangle_{L_2} = \int_{\mathcal{X}} \Upsilon f(x) \Upsilon g(x) dx \tag{4.17}$$

in a suitable L_2 space of functions. This space contains transformed versions of the original functions, where the transformation Υ "extracts" those parts that should be affected by the regularization. This gives a much clearer physical understanding of the dot product in the RKHS (and thus of the similarity measure used by SVMs). It becomes particularly illuminating once one sees that for common kernels, the associated transformation Υ extracts properties like *derivatives* of functions. In other words, these kernels induce a form of regularization that penalizes non-smooth functions.

Definition 4.8 (Regularization Operator) *A regularization operator Υ is defined as a linear map from the space of functions $\mathcal{F} := \{f | f : \mathcal{X} \to \mathbb{R}\}$ into a space equipped with a dot product. The regularization term $\Omega[f]$ takes the form*

$$\Omega[f] := \frac{1}{2} \langle \Upsilon f, \Upsilon f \rangle. \tag{4.18}$$

Positive Definite Operator

Without loss of generality, we may assume that Υ is positive definite. This can be seen as follows: all that matters for the definition of $\Omega[f]$ is the positive definite operator $\Upsilon^*\Upsilon$ (since $\langle \Upsilon f, \Upsilon f \rangle = \langle f, \Upsilon^*\Upsilon f \rangle$). Hence we may always define a positive definite operator $\Upsilon_h := (\Upsilon^*\Upsilon)^{\frac{1}{2}}$ (cf. Section B.2.2) which has the same regulariza-