For (b), the finite evaluation matrix is $H \in \mathbb{R}^{N\times M}$ and $K = HH^\top$. With $\lambda > 0$, $(H^\top H+\lambda I)H^\top = H^\top(K+\lambda I)$, hence

$$\widehat{\beta} = H^\top(K + \lambda I)^{-1}y, \qquad \boxed{\widehat{f} = K(K + \lambda I)^{-1}y.}$$

Both shifted matrices are invertible by positive definiteness, so all multiplications are legitimate even when H is rank deficient. The empirical SVD can be obtained from an eigendecomposition of K; this is a finite-sample alternative to the global Mercer construction.

Structural or analytic derivation

The push-through identity shows that the solution lies in the span of the observed feature vectors. The infinite or very large basis disappears from the prediction calculation once its inner products are known. The two meanings of eigen-decomposition—global kernel operator and empirical Gram matrix—are kept separate.

Comparative cost and preferred method

For $M \gg N$, a dense feature normal-equation solve has $O(M^3)$ factorization cost and $O(M^2)$ storage. The dual system has N unknowns and $O(N^3)$ factorization cost, plus kernel construction and prediction. With explicit H, constructing its Gram matrix costs $O(N^2M)$; this setup cost is included. A thin-SVD conventional implementation achieves the same dimensional reduction.