Suppose $\widehat{X} = X - \mu$ has rank r. To overcome the second problem, note that if

$$\widehat{X} = VDU^\top$$

is an SVD for $\widehat{X}$, then

$$\widehat{X}^\top = UD^\top V^\top$$

is an SVD for $\widehat{X}^\top$, and the $r \times r$ submatrix of $D^\top$ consisting of the first r rows and r columns of $D^\top$ (and D), is the diagonal Σ_r matrix consisting of the singular values $\sigma_1 \geq \cdots \geq \sigma_r$ of $\widehat{X}$, so we can express the matrix U_r consisting of the first r columns u_k of U in terms of the matrix V_r consisting of the first r columns v_k of V $(1 \leq k \leq r)$ as

$$U_r = \widehat{X}^\top V_r \Sigma_r^{-1}.$$

Furthermore, $\sigma_1^2 \geq \cdots \geq \sigma_r^2$ are the nonzero eigenvalues of $\widehat{\mathbf{K}} = \widehat{X}\widehat{X}^\top$, and the columns of V_r are corresponding unit eigenvectors of $\widehat{\mathbf{K}}$. From

$$U_r = \widehat{X}^\top V_r \Sigma_r^{-1}$$

the kth column u_k of U_r (which is a unit eigenvector of $\widehat{X}^\top \widehat{X}$ associated with the eigenvalue σ_k^2) is given by

$$u_k = \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{X}_i^\top = \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{\varphi(x_i)}, \quad 1 \leq k \leq r,$$

so the projection of $\widehat{\varphi(x)}$ onto u_k is given by

$$\begin{aligned}
\langle \widehat{\varphi(x)}, u_k \rangle &= \left\langle \widehat{\varphi(x)}, \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{\varphi(x_i)} \right\rangle \\
&= \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \left\langle \widehat{\varphi(x)}, \widehat{\varphi(x_i)} \right\rangle = \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{\kappa}(x, x_i).
\end{aligned}$$

Therefore, the jth component of the principal component Y_k in the principal direction u_k is given by

$$(Y_k)_j = \langle X_j - \mu, u_k \rangle = \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{\kappa}(x_j, x_i) = \sum_{i=1}^n \sigma_k^{-1}(v_k)_i \widehat{\mathbf{K}}_{ij}.$$

The generalization of kernel PCA to a general embedding $\varphi\colon \mathcal{X} \to F$ of $\mathcal{X}$ in a (real) feature space $(F, \langle -, - \rangle)$ (where F is not restricted to be equal to $\mathbb{R}^d$) with the kernel matrix $\mathbf{K}$ given by

$$\mathbf{K}_{ij} = \langle \varphi(x_i), \varphi(x_j) \rangle,$$

goes as follows.