Finite-Dimensional RKHS

Ömür Uğur

Institute of Applied Mathematics
Middle East Technical University

RKHS (Learning) Seminars, March 2024

Table of Contents

Continue with . . .

Motivating Example

Consider a line passing through the origin in $\mathbb{R}^2$, with the parametrisation:

$$\ell(\theta) = \{(t\cos\theta, t\sin\theta) : t \in \mathbb{R}\} \subset \mathbb{R}^2.$$

It can easily be shown that

$$f(\theta) = (r(\theta)\cos\theta, r(\theta)\sin\theta),$$

where $r(\theta) = p_1\cos\theta + p_2\sin\theta$, is the *projection* of an arbitrary point $p = (p_1, p_2) \in \mathbb{R}^2$ *onto* the *one-dimensional* line (space):

$$\operatorname{proj}_{\ell(\theta)} p = \frac{p^\top v}{\|v\|^2}\, v,$$

where v is (any direction) vector of the line; for instance,

$$v = (\cos\theta, \sin\theta).$$

RKHS Way of Representation

RKHS theory uses, instead of *one*, *two* vectors to represent the line $\ell(\theta)$. It turns out that the *kernel* of $\ell(\theta)$, in matrix form, becomes

$$K(\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}.$$

Note that the columns,

$$k_1(\theta) = \cos\theta \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad k_2(\theta) = \sin\theta \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix},$$

of this kernel spans the the line: $\operatorname{span}\{k_1(\theta), k_2(\theta)\} = \ell(\theta) = V_\theta \subset \mathbb{R}^2$, although they are *linearly dependent*. Interestingly,

$$f(\theta) = p_1 k_1(\theta) + p_2 k_2(\theta) = K(\theta)p$$

turns out to be the projection.

Continue with ...

RKHS

Recall — Definitions

Definition (RKHS)

Let Ω be a set. We call a subset $\mathcal{H}$ of the set of all functions $\mathcal{F}(\Omega, \mathbb{F})$ from Ω to $\mathbb{F}$, that is, $\mathcal{H} \subseteq \mathcal{F}(\Omega, \mathbb{F})$, a *reproducing kernel Hilbert space* (RKHS) on Ω if,

1. $\mathcal{H}$ is a vector space of $\mathcal{F}(\Omega, \mathbb{F})$;
2. $\mathcal{H}$ is endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ with respect to which $\mathcal{H}$ is a Hilbert space;
3. for every $x \in \Omega$, the linear *evaluation functional* $\mathcal{L}_x : \mathcal{H} \to \mathbb{F}$, $f \mapsto \mathcal{L}_x f = f(x)$ is bounded.

Remark. If $\mathcal{H}$ is an RKHS, the by Riesz representation theorem the linear evaluation functional, for each $x \in \Omega$, is given by a unique vector $k_x \in \mathcal{H}$, such that for every $f \in \mathcal{H}$,

$$f(x) = \mathcal{L}_x f = \langle f, k_x \rangle_{\mathcal{H}} .$$

Definition (Reproducing Kernel)

The function k_x is called the *reproducing kernel for the point* x. The function $K(x, y) : \Omega \times \Omega \to \mathbb{F}$ defined by

$$K(x, y) = k_y(x)$$

is called the *reproducing kernel for* $\mathcal{H}$.

Remark. Note that

$$K(x, y) = k_y(x) = \langle k_y, k_x \rangle_{\mathcal{H}}$$

so that

$$K(x, y) = k_y(x) = \langle k_y, k_x \rangle_{\mathcal{H}} = \overline{\langle k_x, k_y \rangle_{\mathcal{H}}} = \overline{K(y, x)}.$$

Even, further we have for the linear *evaluation functional* $\mathcal{L}_x : \mathcal{H} \to \mathbb{F}$,

$$\|\mathcal{L}_x\|^2_{\mathcal{H}\to\mathbb{F}} = \|k_x\|^2_{\mathcal{H}} = \langle k_x, k_x \rangle_{\mathcal{H}} = K(x, x).$$

$\mathbb{F}^n$ as an RKHS

For $u, v \in \mathbb{F}^n$, we let the usual inner product be $\langle u, v\rangle = v^{\mathrm{H}} u = \sum_{i=1}^{n} \overline{v_i} u_i$.
We can also think an n-tuple, say $x = (x_1, \ldots, x_n)$, as a function, say

$$x : \Omega = \{1, \ldots, n\} \to \mathbb{F}, \qquad x : i \mapsto x(i) = x_i,$$

so that with this identification $\mathbb{F}^n$ becomes the vector space of all functions on Ω.
For the orthonormal basis $\{e_i\}_{i=1}^{n}$ we define the functions $e_i(j) = \delta_{ij}$; thence,

$$\mathcal{L}_i x = x(i) = x_i = \langle x, e_i\rangle_{\mathcal{H}} \,. \qquad (\star)$$

Therefore, $\mathbb{F}^n$ is an RKHS with the *reproducing kernel for the point* $i \in \Omega$ is $e_i \in \mathcal{H}$, and the *reproducing kernel for* $\mathcal{H}$ is (the identity matrix)

$$K(i, j) = \langle e_j, e_i\rangle_{\mathcal{H}} = \delta_{ij}. \qquad (\star)$$

Continue with . . .

Preparations

Definition (Gram Matrix)

For the given $\{v_1, \ldots, v_r\} \subset V \subset \mathbb{R}^n$ for an inner product (sub-)space, the Gram matrix $G = (G_{ij})$ is defined as

$$G_{ij} = \langle v_i, v_j \rangle_V = \langle v_j, v_i \rangle_V .$$

Let $\dim V = r$ and

$$\operatorname{span}\{v_1, \ldots, v_r\} = V.$$

This is perfectly fine to understand the inner product (sub-)space. Even further, we may choose an orthonormal basis $\{u_1, \ldots, u_r\} = V$ such that $G = I$.

Preparations

Lemma (Properties of Gram Matrix)

For the set of vectors $\{v_1, \dots, v_r\}$, *the Gram matrix is*

1. *positive semi-definite;*
2. *positive definite (and hence, nonsingular) if* $\{v_1, \dots, v_r\}$ *is linearly independent.*

Proof.

1. $x^\top G x = \sum_{i,j} x_i x_j G_{ij} = \sum_{i,j} x_i x_j \langle v_i, v_j \rangle = \sum_{i,j} \langle x_i v_i, x_j v_j \rangle = \left\langle \sum_{i,j} x_i v_i, \sum_{i,j} x_j v_j \right\rangle = \|\sum_s x_s v_s\|^2 \geq 0.$
2. Using above, $\|\sum_s x_s v_s\|^2 = 0 \iff \|\sum_s x_s v_s\| = 0 \iff x_s = 0$ since the v_s are linearly independent.

□

RKHS Way Configuration

Find (rather than the basis) a *unique, ordered, spanning* set $\{k_1, \dots, k_n\}$ for $V \subset \mathbb{R}^n$ by the rule that k_i is the *unique* vector in V satisfying (certain condition such as)

$$\langle v, k_i \rangle_V = \mathcal{L}_i v = v(i) = e_i^\top v = v_i, \qquad \text{for all } v \in V.$$

Notice the use of the (extrinsic) coordinates in $\mathbb{R}^n$ rather than the (intrinsic) coordinates in V for the vector $v \in V$.
Although the term $v_i = e_i^\top v$ looks like an inner product (the standard inner product in $\mathbb{R}^n$), we emphasise that

$$\mathcal{L}_i v = v_i = e_i^\top v = \langle v, e_i \rangle_{\text{standard}}$$

must be understood as the *point-evaluation* of the functional, or simply, the evaluation functional.

The Kernel — Definitions (with Matrices)

Let $V \subset \mathbb{R}^n$ be an inner product space. The kernel of V is the unique $K = [k_1, \ldots, k_n] \in \mathbb{R}^{n \times n}$ determined by any of the following three equivalent definitions.

1. K is such that each $k_i \in V$ and

$$\langle v, k_i \rangle = e_i^\top v, \qquad \text{for all } v \in V.$$

2. For an orthonormal basis $\{u_1, \ldots, u_r\}$ for V, the kernel is

$$K = u_1 u_1^\top + \cdots + u_r u_r^\top = \sum_{j=1}^{r} u_j u_j^\top .$$

3. K is such that the k_i span V, that is $\operatorname{span}\{k_1, \ldots, k_n\} = V$, and

$$\langle k_j, k_i \rangle = K_{ij} .$$

Example — Motivating

For a fixed θ, show (or calculate) that the kernel of

$$V = \ell(\theta) = \{(t\cos\theta, t\sin\theta) : t \in \mathbb{R}\} \subset \mathbb{R}^2,$$

where V is endowed with the standard inner product on $\mathbb{R}^2$, is

$$K(\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix},$$

where

$$k_1(\theta) = \cos\theta \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad k_2(\theta) = \sin\theta \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

Just verify that these satisfy the three definitions above!

RKHS

Example — Inner Product defined by ...

Let $V = \mathbb{R}^n$ with an inner product $\langle u, v \rangle_V = v^\top Q u$, where Q is *symmetric and positive definite*.

The requirement

$$v_i = \mathcal{L}_i v = e_i^\top v = \langle v, k_i \rangle_V = k_i^\top Q v$$

implies that (having the transpose of both sides)

$$k_i = Q^{-1} e_i.$$

Hence the kernel is

$$K = \left[Q^{-1} e_1, \ldots, Q^{-1} e_n \right] = Q^{-1} \left[e_1, \ldots, e_n \right] = Q^{-1}.$$

Example — Inner Product defined by ...

Alternatively, an eigendecomposition of the matrix $Q = XDX^\top$, where D is the diagonal matrix consisting of the eigenvalues of Q and the columns of X are the corresponding orthonormal eigenvectors, can be used to construct an *orthonormal* basis for V; that is,

$$\mathcal{B} = \left\{ XD^{-1/2}e_1, \ldots, XD^{-1/2}e_n \right\}.$$

Using this basis, in the *second* definition of the kernel, yields

$$K = \sum_{i=1}^{n} \left(XD^{-1/2}e_i \right) \left(XD^{-1/2}e_i \right)^\top = \sum_{i=1}^{n} XD^{-1/2} \underbrace{e_i e_i^\top}_{I_i} D^{-1/2} X^\top$$

$$= XD^{-1/2} \left(\sum_{i=1}^{n} I_i \right) D^{-1/2} X^\top = XD^{-1} X^\top = Q^{-1}$$

Note also that for $k_j = Q^{-1}e_j$, we have

$$\langle k_j, k_i \rangle_V = \left\langle Q^{-1}e_j, Q^{-1}e_i \right\rangle_V = e_i^\top Q^{-1} Q Q^{-1} e_j = e_i K e_j = K_{ij}.$$

RKHS

Example — A Trivial One!

Let $V \subset \mathbb{R}^n$ be a subspace spanned by the vector $v = (1, 1)$ and endowed with the inner product giving the vector v unit norm, that is, $\langle v, v \rangle_V = 1$.

No reference to "what exactly the inner product is" is given!

Since $\{v = (1, 1)\}$ is an *orthonormal basis* for V, the kernel is,

$$K = vv^\top = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = [k_1, k_2] \, .$$

What is the inner product therefore?

Note that $\langle k_i, k_j \rangle_V = 1$ for all $i, j = 1, 2$. Use this to complete the following exercise (consequence of the third definition): show that for $x, y \in V$, written as,

$$x = x_1 k_1 + x_2 k_2 = K\alpha, \qquad \alpha = [x_1, x_2]^\top,$$

$$y = y_1 k_1 + y_2 k_2 = K\beta, \qquad \beta = [y_1, y_2]^\top,$$

the inner product is,

$$\langle x, y \rangle_V = \langle K\alpha, K\beta \rangle_V = \beta^\top K\alpha.$$

Example — Yet Another Trivial One!

For the kernel

$$K = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

the associated configuration, the subspace V, can be found as follows: since K is 2×2, $V \subset \mathbb{R}^2$; since V is the span of $k_1 = (1,0)$ and $k_2 = (0,0)$, the subspace $V = \mathbb{R} \times \{0\} = \{(t,0) : t \in \mathbb{R}\} \subset \mathbb{R}^2$. Note also that the vector $k_1 = (1,0)$ has a unit norm in V:

$$\langle k_1, k_1 \rangle_V = K_{11} = 1.$$

Exercise. Find (calculate, reproduce) the inner product $\langle x, y \rangle_V$ for any $x = K\alpha \in V$ and $y = K\beta \in V$:

$$\langle x, y \rangle_V = \beta^\top K \alpha.$$

RKHS

The Kernel — Definitions (with Matrices) — Recall —

Let $V \subset \mathbb{R}^n$ be an inner product space. The kernel of V is the unique $K = [k_1, \ldots, k_n] \in \mathbb{R}^{n \times n}$ determined by any of the following three equivalent definitions.

1. K is such that each $k_i \in V$ and
$$\langle v, k_i \rangle = e_i^\top v, \qquad \text{for all } v \in V.$$
2. For an orthonormal basis $\{u_1, \ldots, u_r\}$ for V, the kernel is
$$K = u_1 u_1^\top + \cdots + u_r u_r^\top = \sum_{j=1}^{r} u_j u_j^\top.$$
3. K is such that the k_i span V, that is $\operatorname{span}\{k_1, \ldots, k_n\} = V$, and
$$\langle k_j, k_i \rangle = K_{ij}.$$

Some Remarks — Before the Boring Stuff!

- Existence and uniqueness of K follows most easily from Definition 1 (as it boils down to solving linear systems).
- Positive (semi-)definiteness of K is apparent from Definition 2; however, might be difficult from Definition 1. Also, uniqueness is not clear from Definition 2.
- Existence of K is not clear from Definition 3; however, it has a plausible implication: for $v, w \in V$, we have $v = \sum_{j=1}^{n} \alpha_j k_j = K\alpha$, $w = \sum_{j=1}^{n} \beta_j k_j = K\beta$, and further,

$$\langle K\alpha, K\beta \rangle_V = \beta^\top K\alpha.$$

These follows simply from the spanning set and the properties of the inner product:

$$\langle K\alpha, K\beta \rangle_V = \sum_{j,i}^{n} \alpha_j \beta_i \langle k_j, k_i \rangle_V = \sum_{j,i}^{n} \alpha_j \beta_i K_{ij} = \beta^\top K\alpha.$$

RKHS

Existence-Uniqueness of the Kernel

Lemma (Lemma 2.1)

Given an inner product space $V \subset \mathbb{R}^n$, there is precisely one matrix $K = [k_1, \dots, k_n]$ in $\mathbb{R}^{n\times n}$ for which each $k_i \in V$ and satisfies

$$\langle v, k_i \rangle_V = \mathcal{L}_i v = e_i^\top v, \qquad \text{for all } v \in V.$$

Proof. Let $\mathcal{B} = \{v_1, \dots, v_r\} \subset V$ be a basis for V, and let the $k_i = \sum_{j=1}^r \alpha_j^i v_j \in V$. Then the properties of the Gram matrix in solving the linear system of equations (obtained using $e_i^\top v_j = \mathcal{L}_i v_j = \langle v_j, k_i \rangle$):

$$b^i = \begin{bmatrix} e_i^\top v_1 \\ \vdots \\ e_i^\top v_r \end{bmatrix} = \begin{bmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_r \rangle \\ \vdots & \ddots & \vdots \\ \langle v_r, v_1 \rangle & \cdots & \langle v_r, v_r \rangle \end{bmatrix} \begin{bmatrix} \alpha_1^i \\ \vdots \\ \alpha_r^i \end{bmatrix} = G\alpha^i,$$

for $i = 1, \dots, n$, yields the existence and uniqueness.

Reproducing Property of the Kernel

Lemma (Lemma 2.2)

If K is such that the k_i span V and $\langle k_j, k_i \rangle_V = K_{ij}$, then

$$\langle v, k_i \rangle_V = \mathcal{L}_i v = e_i^\top v, \qquad \text{for all } v \in V.$$

Proof. Fix $v \in V$ and since k_i is in the span, we have $v = \sum_i \alpha_i k_i = K\alpha$. Therefore,

$$\langle v, k_i \rangle = \langle K\alpha, Ke_i \rangle = e_i^\top K\alpha = e_i^\top v = \mathcal{L}_i v$$

completes the proof.

Note that we used the fact that from Definition 3, it follows that

$$\langle K\alpha, Ke_i \rangle = e_i^\top K\alpha.$$

Entries of the Kernel

Lemma (Lemma 2.3)

If K is such that the $k_i \in V$ and $\langle v, k_i \rangle_V = \mathcal{L}_i v = e_i^\top v$ for all $v \in V$, then the k_i span V and

$$\langle k_j, k_i \rangle_V = K_{ij}.$$

Proof. The trivial part follows from

$$\langle k_j, k_i \rangle = e_i^\top k_j = K_{ij}.$$

To show $\operatorname{span}\{k_1, \ldots, k_n\} = V$, assume the contrapositive: there is a non-zero $k \in V$ which is orthogonal to each and every k_i; that is,

$$\langle k, k_i \rangle = 0, \qquad \text{for all } i.$$

However, the assumetions states that $\langle k, k_i \rangle = e_i^\top k$ for every i; hence $k = 0$ vector; hence a contradiction. Thus, $\operatorname{span}\{k_1, \ldots, k_n\} = V$.

Kernel as Outer Product

Lemma (Lemma 2.4)

If $\{u_1, \dots, u_r\}$ is an orthonormal basis for V, then

$$K = u_1 u_1^\top + \cdots + u_r u_r^\top = \sum_{i=1}^{r} u_i u_i^\top$$

and satisfies

$$\langle v, k_i \rangle_V = \mathcal{L}_i v = e_i^\top v, \qquad \text{for all } v \in V.$$

Proof. Let $U = [u_1, \dots, u_r] \in \mathbb{R}^{n \times r}$ so that $K = UU^\top$. Write $k_i = Ke_i = UU^\top e_i = U\beta$ and take an arbitrary $v = \sum_i \alpha_i u_i = U\alpha \in V$. Therefore

$$\begin{aligned} \langle v, k_i \rangle = \langle U\alpha, U\beta \rangle &= \sum_{i,j} \alpha_i \beta_j \langle u_i, u_j \rangle = \sum_{i,j} \alpha_i \beta_j = \beta^\top \alpha \\ &= e_i^\top U\alpha = e_i^\top v = v_i = \mathcal{L}_i v. \end{aligned}$$

RKHS

Existence of an Inner Product I

Lemma (Lemma 2.6)

Let $V = \operatorname{span}\{k_1, \ldots, k_n\}$ *be the space spanned by the columns of a positive semi-definite matrix* $K = [k_1, \ldots, k_n] \in \mathbb{R}^{n \times n}$*. Then, there exists and inner product on* V *satisfying* $\langle k_j, k_i \rangle_V = K_{ij}$*.*

Proof. For $x = K\alpha$ and $y = K\beta$ in V, define the inner product as

$$\langle x, y \rangle_V = \beta^\top K \alpha.$$

It is easy to show that this is a well-defined: in the sense that for other representations of $x = K\tilde{\alpha}$ and $y = K\tilde{\beta}$ we have

$$\begin{aligned}\beta^\top K\alpha - \tilde{\beta}^\top K\tilde{\alpha} &= (\beta - \tilde{\beta})^\top K\alpha + \tilde{\beta}^\top K(\alpha - \alpha^\top) \\ &= \left[K^\top(\beta - \tilde{\beta})\right]^\top \alpha + \tilde{\beta}^\top K(\alpha - \tilde{\alpha}) = 0\end{aligned}$$

RKHS

Existence of an Inner Product II

That is, no matter the representation; the inner product is

$$\langle x, y \rangle_V = \beta^\top K \alpha.$$

It is not difficult to show that this is really an inner product: liearity in the first argument is clear; for the positive definiteness (of the inner product), we calculate

$$\langle x, x \rangle_V = \langle K\alpha, K\alpha \rangle_V = \alpha^\top K \alpha = 0,$$

since K is positive semi-definite, which implies that $K\alpha = 0$.
Note. This is stated as Lemma 2.5:

Uniqueness of the Inner Product Space

Lemma (Lemma 2.6)

Let $V_1 \subset \mathbb{R}^n$ and $V_2 \subset \mathbb{R}^n$ be two inner product spaces having the same kernel K. Then, V_1 and V_2 are identical spaces: $V_1 = V_2$ and their inner products are the same (as above).

Proof. Since the columns of K span both V_1 and V_2, then $V = V_1 = V_2$. Since the inner products on V_1 and V_2 are uniquely determined from the Gram matrix $K_{ij} = \langle k_j, k_i \rangle_V$ corresponding to $k_1, \dots, k_n$, and both V_1 and V_2 have the same matrix, their inner products are identical and further for any $u = K\alpha$ and $v = K\beta$

$$\langle u, v \rangle_V = \beta^\top K \alpha.$$

RKHS

Continue with . . .

Extrinsic versus Intrinsic

- In applications, knowing $V \subset \mathbb{R}^n$ allows us working with V using the *extrinsic* coordinates by writing an element of V as a vector in $\mathbb{R}^n$.
- Note that if V and W are two r-dimensional linear subspaces of $\mathbb{R}^n$, then their *intrinsic* geometry is the same, but their *extrinsic* geometry may differ unless $V = W$.

Example (The Problem at Hand)

Endow $V \subset \mathbb{R}^n$ with an inner product. Fix $i \in \{1, \ldots, n\}$ and consider *how to find* $x \in V$ *satisfying*

$$f(x) = e_i^\top x = 1$$

and having the *smallest* norm (induced by the inner product in V).

The Problem — Interpolation

In other words, we wish to solve

Example (The Problem at Hand)

$$\underset{x \in V}{\text{minimise}} \;\; \|x\|_V^2 = \langle x, x\rangle_V$$

$$\text{subject to } f(x) = e_i^\top x = 1$$

The nature of the problem is better understood when the extrinsic coordinates in $\mathbb{R}^n$ are considered, whether or not $V = \mathbb{R}^n$.
Geometrically, such a solution $x \in V$ must be *orthogonal* to any vector $v \in V$ satisfying

$$f(v) = e_i^\top v = 0;$$

otherwise, its norm could be decreased! That is,

$$\langle x, v\rangle_V = 0, \qquad \text{for all } v \in V \text{ subject to } e_i^\top v = 0.$$

The Problem — Exercise

Exercise

Take $V = \{x \in \mathbb{R}^3 : x_1 + x_2 = 1, x_3 \in \mathbb{R}\} \subset \mathbb{R}^3$ and $i = 1$ (or, $i = 2, 3$) fixed, to visualise the problem (using the standard inner product in $\mathbb{R}^n$). Then, solve the resulting problem.

Particularly, show a special attention to those vectors $v \in V$ which are perpendicular (respectively, not perpendicular) to the solution vector $x \in V$; namely, $\langle x, v \rangle_V = 0$ (respectively, $\langle x, v \rangle_V \neq 0$).

In an attempt to solve the problem, you might need a basis for V; take any (suitable one)!

The Problem — Solution using Kernel I

Assume that we know the kernel $K = [k_1, \ldots, k_n]$ of the subspace $V \subset \mathbb{R}^n$. Then, let $x = K\alpha$ and $v = K\beta$.
Then for the first constraint,

$$e_i^\top x = 1 \implies 1 = e_i^\top K\alpha = k_i^\top \alpha$$

and, for the second constraint, with $v = K\beta$,

$$\langle x, v\rangle_V = 0 \text{ whenever } e_i^\top v = 0 \implies \langle K\alpha, K\beta\rangle_V = \beta^\top K\alpha = 0,$$

whenever

$$0 = e_i^\top K\beta = k_i^\top \beta = \beta^\top k_i = \beta^\top K e_i.$$

That is, the constraints turns to

$$k_i^\top \alpha = 1 \qquad \text{(first)}$$

$$\beta^\top K\alpha = 0 \text{ whenever } \beta^\top K e_i = 0. \qquad \text{(second)}$$

The Problem — Solution using Kernel II

The second constraint is satisfied when

$$\alpha = ce_i, \qquad c \in \mathbb{R}.$$

Hence by the first constraint, we get

$$1 = k_i^\top ce_i = cK_{ii} \implies c = \frac{1}{K_{ii}}.$$

Therefore,

$$x = K\alpha = Kce_i = ck_i = \frac{1}{K_{ii}}k_i.$$

Also note that

$$\|x\|_V^2 = \langle x, x\rangle_V = \frac{1}{K_{ii}^2}\langle k_i, k_i\rangle_V = \frac{1}{K_{ii}}$$

is the minimum norm of such a solution $x \in V$.

Yet, Towards Another Definition of the Kernel

- In the above example problem, as V changes, both the kernel K and the solution $x \in V$ change. Yet, the relationship between x and K remains the same:

$$x = \frac{1}{K_{ii}} k_i.$$

- There is a geometric explanation for the columns of K solving the *single-point* interpolation problem: let $L_i : v \mapsto e_i^\top v$ denote the ith coordinate function. Then,

$$L_i(v) = \langle v, k_i \rangle_V$$

  means that k_i is the *gradient* of L_i. In particular, the line determined by k_i meets the level set $\{v : L_i(v) = 1\}$ at *right angles*, showing that k_i meets the *orthogonality* condition for *optimality*.

Geometric Definition of the Kernel

Definition (Lemma 2.8 — Geometric Interpretation)

Let $H_i = \left\{ z \in \mathbb{R}^n : f(z) = e_i^\top z = 1 \right\}$ be the hyperplane consisting of all vectors whose ith coordinate is unity.

- If $V \cap H_i$ is empty then define $k_i = 0$;
- Otherwise, let $\tilde{k}_i$ be the point in the intersection $V \cap H_i$ that is closest to the origin. Define k_i to be $k_i = c^2 \tilde{k}_i$, where $c^2 = \left\langle \tilde{k}_i, \tilde{k}_i \right\rangle_V^{-1}$.

Proof. Let $v \in V$ such that $a = e_i^\top v = v_i$; define $w = v - a\tilde{k}_i \in V$. Then for $k_i = c^2 \tilde{k}_i$, we must have

$$
\begin{aligned}
a = v_i = \langle v, k_i \rangle_V &= \left\langle w + a\tilde{k}_i, c^2 \tilde{k}_i \right\rangle_V \\
&= c^2 \left\langle w, \tilde{k}_i \right\rangle_V + ac^2 \left\langle \tilde{k}_i, \tilde{k}_i \right\rangle_V = ac^2 \left\langle \tilde{k}_i, \tilde{k}_i \right\rangle_V = a.
\end{aligned}
$$

The choice of c^2 guarantees in particular: $\langle k_i, k_i \rangle_V = K_{ii}$.

Visualising Geometric Definition of the Kernel

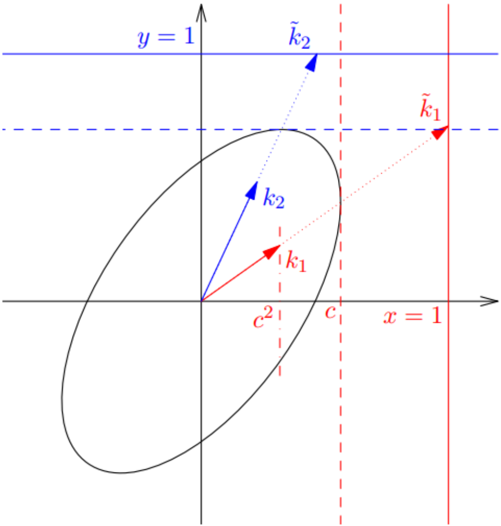

Figure 2.1: The ellipse comprises all points one unit from the origin. It determines the chosen inner product on $\mathbb{R}^2$. The vector $\tilde{k}_1$ is the closest point to the origin on the vertical line $x = 1$. It can be found by enlarging the ellipse until it first touches the line $x = 1$, or equivalently, as illustrated, it can be found by shifting the line $x = 1$ horizontally until it meets the ellipse tangentially, represented by the dashed vertical line, then travelling radially outwards from the point of intersection until reaching the line $x = 1$. The vector k_1 is a scaled version of $\tilde{k}_1$. If the dashed vertical line intersects the x-axis at c then $k_1 = c^2\tilde{k}_1$. Equivalently, k_1 is such that its tip intersects the line $x = c^2$. The determination of k_2 is analogous but with respect to the horizontal line $y = 1$.

Visualising Columns of the Kernel

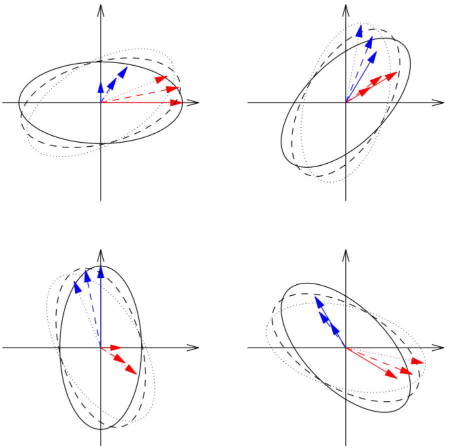

Figure 2.2: Shown are the vectors k_1 (red) and k_2 (blue) corresponding to rotated versions of the inner product $\langle u, v\rangle = v^\top Qu$ where $Q = \mathrm{diag}\{1, 4\}$. The magnitude and angle of k_1 and k_2 are plotted in Figure 2.3.

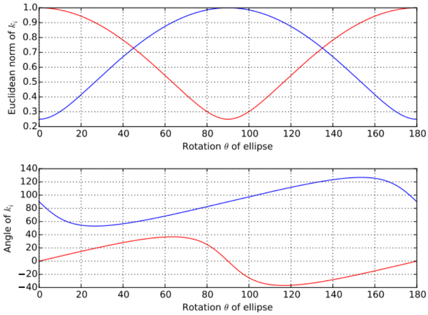

Figure 2.3: Plotted are the magnitude and angle of k_1 (red) and k_2 (blue) corresponding to rotated versions of the inner product $\langle u, v\rangle = v^\top Qu$ where $Q = \mathrm{diag}\{1, 4\}$, as in Figure 2.2.

Visualising the Kernel of 1D subspace

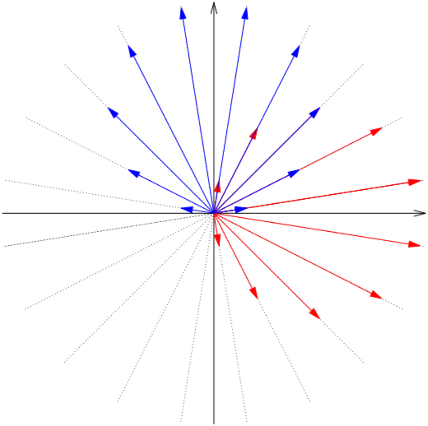

Figure 2.4: Shown are k_1 (red) and k_2 (blue) for various one-dimensional subspaces (black) of $\mathbb{R}^2$. In all cases, the inner product is the standard Euclidean inner product. Although not shown, k_2 is zero when V is horizontal, and k_1 is zero when V is vertical. Therefore, the magnitude of the red vectors increases from zero to a maximum then decreases back to zero. The same occurs for the blue vectors.

RKHS

Visualising Convergence to 1D subspace

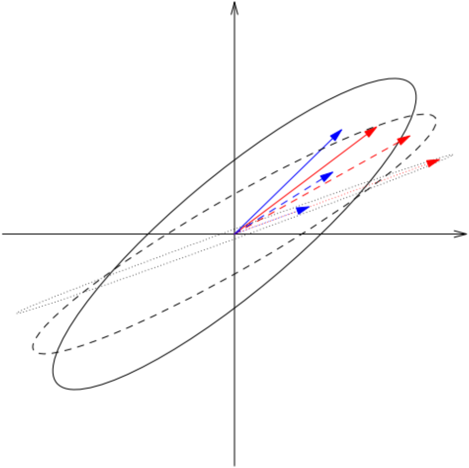

Figure 2.5: Illustration of how, as the ellipse gets narrower, the two-dimensional inner product space $V = \mathbb{R}^2$ converges to a one-dimensional inner product space. The kernels of the subspaces are represented by red (k_1) and blue (k_2) vectors.

Bibliography

Jonathan H. Manton and Pierre-Olivier Amblard.
A primer on reproducing kernel Hilbert spaces, 2015.

Il Shan Ng.
Reproducing kernel Hilbert spaces & machine learning, 2024.
accessed: March 2024.

Vern I. Paulsen and Mrinal Raghupathi.
An Introduction to the Theory of Reproducing Kernel Hilbert Spaces.
Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.

Jesse Perla, Thomas J. Sargent, and John Stachurski.
Orthogonal projections and their applications, 2024.
accessed: March 2024.

RKHS