Reproducing Kernels Preserving Algebraic Structure: A Duality Approach

Tzvetan Ivanov*, Balázs Csanád Csáji†‡

***Abstract*— From the classical reproducing kernel theory of function spaces it is well-known that there is an inverse relationship between inner-products and kernels. In applications, such as linear system theory and machine learning, these kernels are often highly structured. In order to exploit algebraic structure, it is common to choose basis functions and fall back to matrix representations. However, the basis has to be chosen in a way that is compatible with the algebraic structure, which is itself a nontrivial task. We therefore choose a different approach and use standard duality theory where additional algebraic structures form no obstacle. This is demonstrated by examples from linear system theory, namely two variable polynomials given by Bézoutians and quadratic differential forms.**

Keywords: reproducing kernels, duality theory, shifts, Bézoutians, Hankel operators, asymptotic stability

I. Introduction

The theory of reproducing kernels developed impressively in the last decades and is of importance not only in functional analysis, but also in a large range of applied fields, such as statistics, control theory and machine learning [5]. An example of its numerous successful applications is the concept of support vector machines in statistical learning [8].

The classical theory of reproducing kernels is formulated in the framework of Hilbert spaces of scalar-valued functions. From the start the bijection between positive kernels and reproducing kernel Hilbert spaces was characterized essentially as an inverse relationship [5]. In recent years there were many attempts to extend this theory. Rather natural generalizations focused on the function space and allowed vector-valued functions [2]. More further-reaching generalizations relaxed the Hilbert space assumption and thus moved from positive kernels to indefinite kernels and reproducing kernel Kreĭn spaces [5]. The generalization to indefinite kernels has practical importance not only because testing Mercer's condition can be challenging, but also because some of the often applied kernels, for example, the hyperbolic tangent kernel, are indefinite [6].

In this paper we develop a novel framework of reproducing kernels suitable for applications with rich algebraic structure, e.g., modelling of dynamical systems. In these applications more often than not the linear space consists of relations or equivalence classes which are not necessarily functions. The novelty of our approach is that it deals with linear spaces which possess no function space structure. Also spaces with multiple function space structures are considered. We therefore extend the theory of reproducing kernel Kreĭn spaces from function spaces to general linear spaces.

The framework is developed in a coordinate-free manner and applies to non-degenerate sesquilinear forms on linear spaces. In other words, we present a flexible and unifying theory of kernels together with an algebraic version of the reproducing property. This algebraic reproducing property agrees with the classical one, whenever the space admits an evaluation, i.e., an identification with a function space.

In applications, such as linear system theory and machine learning, the kernels of interest are often highly structured. In order to exploit an algebraic structure, it is common to choose basis functions and fall back to matrix representations. For this to work the basis has to be chosen in a way that is compatible with the algebraic structure, which is itself a nontrivial task. We therefore choose a different approach and use standard duality theory where additional algebraic structures form no obstacle. The effectiveness of this theory is demonstrated by examples from linear system theory, namely two variable polynomials given by Bézoutians and quadratic differential forms.

This paper is structured as follows: In Section II we introduce the conjugate dual space and related constructs such as the conjugate dual map. In Section III we explain the relation between forms and kernels using the concept of primal dual operators and dual primal operators. In Section III-C we provide the link to classical integral-kernel representations on function spaces. In Section IV we discuss positivity, signature and truncation of forms and kernels. We discuss the resulting computational issues in Section IV-A using matrix representations. In Section V we focus on additional algebraic structure, more precisely, Hankel operators, Bézoutians and general intertwining maps. As an example we treat stability analysis for linear systems, in order to illustrate how this structure can be exploited. Finally, in Section VI we provide some concluding remarks.

Notations: The letter $\mathbb{K}$ denotes a field which is either $\mathbb{R}$ (real numbers) or $\mathbb{C}$ (complex numbers, where $\jmath$ denotes the imaginary unit). The notation k^* denotes the complex conjugate of $k \in \mathbb{C}$ whereas when restricted to $k \in \mathbb{R}$ it denotes the identity. Extentended to matrices $(\cdot)^*$ means conjugate transpose for $\mathbb{K} = \mathbb{C}$ and transpose for $\mathbb{K} = \mathbb{R}$. The term linear space means linear space over $\mathbb{K}$. If B is a set, $B^{\#}$ denotes its cardinality.

*Tzvetan Ivanov is with the Center for Systems Engineering and Applied Mechanics (CESAME), Universite Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium. email: tzvetan.ivanov@uclouvain.be

† Department of Electrical and Electronic Engineering (EEE), Melbourne School of Engineering, The University of Melbourne, 240 Grattan Street, VIC 3010, Melbourne, Australia, email: bcsaji@unimelb.edu.au

‡ Computer and Automation Research Institute (SZTAKI), Hungarian Academy of Sciences, Kende utca 13–17, Budapest, Hungary, H-1111

ISBN 978-963-311-370-7

II. Preliminaries on Duality Theory

A map $T : U \to V$ between two linear spaces is called *additive* if $T(u+v) = Tu + Tv$ for all $u, v \in U$.

If T is additive and satisfies $T(ku) = k \cdot Tu$ for all $k \in \mathbb{K}, u \in U$ we say that T is a linear map or simply an *operator*. The space of all (linear) operators $U \to V$ is denoted by $L(U, V)$. We denote the *kernel* and *image* of T by $\operatorname{Ker} T$ and $\operatorname{Im} T$, respectively. If $\ell : U \to \mathbb{K}$ is additive and $\ell(ku) = k^*\ell(u)$ we call it a *conjugate linear functional* on U. The space of conjugate linear functions on U is denoted by U^* and called the *conjugate dual* of U.

We use U^{**} to denote the conjugate dual space of U^*, i.e., the space $(U^*)^*$. For $u \in U$ we define

$$\hat{u} : U^* \to \mathbb{K}, \quad \ell \mapsto \hat{u}(\ell) = \ell(u)^*. \tag{1}$$

The map $U \to U^{**}, u \mapsto \hat{u}$ is linear and injective and called the *natural inclusion* of U in U^{**}. If $T : U \to V$ is an operator we define for $\ell \in V^*$

$$T^*\ell : U \to \mathbb{K}, \quad u \mapsto \ell(Tu), \tag{2}$$

and call $T^* : V^* \to U^*$ the *conjugate dual operator* of T. For a general overview on linear spaces, see [7].

A. Quotient Spaces and Annihilators

Linear equivalence relations are those whose equivalence classes form linear subspaces. These relations are crucial in the study of operators with nontrivial kernels.

Given a subspace $X \subseteq U$ we set denote the equivalence class of u mod X by $u + X := \{u + x \,|\, x \in X\}$ and define the *quotient space* via

$$U/X := \{u + X \,|\, u \in U\}. \tag{3}$$

Moreover, we define the *annihilator space* via

$$X^\perp := \{\ell \in U^* \,|\, \operatorname{Ker} \ell \supseteq X\}. \tag{4}$$

There exists a natural inclusion operator $(\cdot)^\iota$ from $X^\perp$ into $(U/X)^*$ given by $(\ell)^\iota(u+X) = \ell(u)^*$ for all $u \in U$. In other words, the annihilator of X forms a subset of the conjugate dual of the quotient space U/X. We shall use these facts later on when we discuss truncation of forms and integral-kernels.

III. Forms and Kernel Representations

In this section we introduce forms and integral-kernels from an operator perspective which makes it easier to see the connections between them. The approach we choose is algebraic in the sense that we do not assume that linear spaces form function spaces. Moreover, our approach is also geometric in the sense that we are working in a coordinate free setup, i.e., all results are independent of the chosen basis.

A. Operators from Primal to Dual and Vice Versa

A (linear) operator $G : U \to V^*$ is called *primal to dual operator* (pd-operator). Similarly a (linear) operator $K : U^* \to V$ is called a *dual to primal operator* (dp-operator). Pd-operators are related to forms and dp-operators are related to kernel-representations as we shall see next.

The *associated form* of a pd-operator $G : U \to V^*$ is

$$\langle u, v\rangle := (Gu)(v) \quad \text{for all} \quad u \in U, v \in V. \tag{5}$$

The associated form is called *bilinear* for $\mathbb{K} = \mathbb{R}$ and *sesquiliniear* for $\mathbb{K} = \mathbb{C}$. If G is bijective one the associated form is called *nondegenerate*. The *adjoint* of a pd-operator $G : U \to V^*$ is an operator from V to U^* and defined by

$$(\tilde{G}v)(u) := (Gu)(v)^* \quad \text{for all} \quad u \in U, v \in V. \tag{6}$$

A pd-operator with $G = \tilde{G}$ is called *self-adjoint*. Similarly, the *adjoint* of a dp-operator $K : U^* \to V$ is the dp-operator

$$\tilde{K} : V^* \to U \quad \text{with} \quad \eta(\tilde{K}\ell) := \ell(K\eta)^*, \tag{7}$$

for all $\eta \in U^*, \ell \in V^*$. If $K = \tilde{K}$, then K is called *self-adjoint*. In what follows we shall see that dp-operators admit an interpretation as kernel representations. In order to do this algebraically, we will apply tensor products.

B. Tensor Products and Abstract Kernel Representations

In order to avoid the need for topological constructs and other techniques from analysis, we shall constrain ourselves to the case where all kernels have finite rank.

The *conjugate tensor product* of two linear spaces U and V is given by the pair $(U \otimes V, \otimes)$ where the first component denotes the linear space

$$U \otimes V := \{K \in L(U^*, V) \,|\, \operatorname{rank} K < \infty\}, \tag{8}$$

and the second component denotes the conjugate linear map $\otimes : U \times V \to L(U^*, V)$ defined by

$$(u \otimes v)(\eta) := \eta(u) \cdot v \quad \text{for all} \quad \eta \in U^*. \tag{9}$$

Every element $K \in U \otimes V$ has a *finite sum* representation of the form $K = \sum_{ij} u_i \otimes v_j$ with $u_i \in U$ and $v_j \in V$.

Note that the elements of $U \otimes V$ are dp-operators.

Lemma 1 *Let $u, w \in U$ and $v \in V$. Then, there holds:*

1) $(u \otimes v)^* = v \otimes u$,
2) $(u + w) \otimes v = u \otimes v + w \otimes v$,
3) $(ku) \otimes v = k^*(u \otimes v)$,
4) $T \circ (u \otimes v) = u \otimes (Tv)$,
5) $(v \otimes u) \circ T^* = (Tv) \otimes u$,

for all operators $T : V \to V$ and scalars $k \in \mathbb{K}$.

Definition 2 Given two operators

$$G : U \to U^* \quad \text{and} \quad T : U \to V, \tag{10}$$

an element $K \in U \otimes V$ is called *kernel representation* of T with respect to G if for all $u \in U$ and $\ell \in V^*$ there holds[1]

$$\ell(Tu)^* = (Gu)(\tilde{K}\ell) \quad \text{for all} \quad u \in U, \ell \in V^*. \tag{11}$$

[1]Geometrically speaking (11) says that $\tilde{K}\ell$ is the gradient of the linear functional $u \mapsto \ell(Tu)^*$ with respect to the form G.

In the special case where $V = U$ and T is the identity we call K the *reproducing kernel of* U with respect to G.

Remark 3 Let $K \in U \otimes V$ be given by $\{v_i\}_{i=1}^n \subseteq V$, $\{u_j\}_{j=1}^m \subseteq U$ and $\alpha \in \mathbb{K}^{n\times m}$, i.e.,

$$K = \sum\nolimits_{j=1}^m \sum\nolimits_{i=1}^n \alpha_{ij} \cdot (u_j \otimes v_i). \tag{12}$$

For all $\ell \in V^*$ there holds

$$\tilde{K}\ell = \sum\nolimits_{j=1}^m \sum\nolimits_{i=1}^n \alpha_{ij}^* \cdot (u_j \cdot \ell(v_i)). \tag{13}$$

Moreover, for all $\eta \in U^*$ we have

$$\ell(K\eta) = \sum\nolimits_{j=1}^m \sum\nolimits_{i=1}^n \alpha_{ij}^* \cdot \eta(u_j)^* \ell(v_i). \tag{14}$$

Theorem 4 *Given three operators G, T as in (10) and a kernel $K \in U \otimes V$ the following statements are equivalent:*

1) $T = KG$,
2) $T^* = \tilde{G}\tilde{K}$,
3) *K is a kernel representation of T w.r.t. G.*

In particular, K is the reproducing kernel of U with respect to G if and only if $K = G^{-1}$.

C. Evaluation Structures and Function Spaces

In this section we link our algebraic constructions to the classical function space setup. We do this by using evaluations. Since evaluations are linear, we will have to conjugate them, in order to remain in the conjugate dual space. More precisely, given a functional $\eta : U \to \mathbb{K}$ we define $\bar{\eta}$ via $\bar{\eta}(u) = \eta(u)^*$. We call

$$\mathrm{ev}|_\Omega := \{\mathrm{ev}_z : U \to \mathbb{K} \,|\, z \in \Omega\}, \tag{15}$$

an *evaluation structure* on U if $\overline{\mathrm{ev}}_z \in U^*$ for all points $z \in \Omega$ in the *domain* Ω. Given that $\mathrm{ev}_z(u) \equiv 0$ for all $z \in \Omega$ implies $u = 0$, the pair $(U, \mathrm{ev}|_\Omega)$ is said to form a *linear function space*.

Remark 5 In linear function spaces it is common to identify vectors $u \in U$ with functions $\Omega \to \mathbb{K}, z \mapsto \mathrm{ev}_z(u)$ if the evaluation structure is clear. In particular, one writes $u(z)$ instead of $\mathrm{ev}_z(u)$. We follow this tradition in Theorem 6.

Theorem 6 *Assume $K \in U \otimes V$ given by (12) is the kernel representation of $T : U \to V$ with respect to G.*

Moreover, assume that $(U, \mathrm{ev}|_\Omega)$ and $(V, \mathrm{ev}|_\Xi)$ form linear function spaces and

$$(Gf)(g) = \int_\Omega f(z)\, g(z)^* \, \mathrm{d}\mu(z), \tag{16}$$

where μ denotes a signed measure on Ω. Then for all $w \in \Xi$ the following integral kernel representation holds

$$(Tf)(w) = \int_\Omega f(z)\, \kappa(z,w)^* \; \mathrm{d}\mu(z), \tag{17a}$$

$$\kappa(z,w) = \sum\nolimits_{j=1}^m \sum\nolimits_{i=1}^n \alpha_{ij}^* \quad u_j(z) \cdot v_i(w)^*, \tag{17b}$$

where the integral kernel κ is given by

$$\kappa : \Omega \times \Xi \to \mathbb{K}, \quad \kappa(z,w) = \overline{\mathrm{ev}}_w(K\overline{\mathrm{ev}}_z), \tag{18}$$

or, equivalently, $\kappa(z,w) = \mathrm{ev}_z(\tilde{K}\overline{\mathrm{ev}}_w)$.

IV. Signature and Congruence

Orthogonal projections and other approximation operations on inner-product spaces are possible due to the fact that inner-product spaces carry a natural 2-norm. In this section we shall introduce the notions of positivy, signature and congruence for pd- and dp-operators. After that, we will discuss how to truncate such operators in a fashion which preserves their signature.

Given a self-adjoint pd-operator $G : U \to U^*$ one writes $G \geq 0$ to indicate G is *positive-semidefinite*, i.e., satisfies

$$(Gu)(u) \geq 0 \quad \text{for all} \quad u \in U. \tag{19}$$

If additionally the associated form is nondegenerate, one writes $G > 0$ and says G is *positive definite*. If G is positive-definite we call the associated form an *inner-product*. For two self-adjoint pd-operators $G, H : U \to U^*$ one writes $G \geq H$ and $G > H$ if $G - H \geq 0$ and $G - H > 0$, respectively.

Definition 7 Let $G : U \to U^*$ denote a self-adjoint pd-operator. Let $\mathcal{L}(U)$ denote the latice of all linear subspaces of U, $\iota_M : M \to U$ denote the natural inclusion and define the numbers

$$\mathrm{ind}_-(G) = \max_{M\in\mathcal{L}(U)} \{\dim M \,|\, -\iota_M^* G_{|M} > 0\}, \tag{20a}$$

$$\mathrm{ind}_+(G) = \max_{M\in\mathcal{L}(U)} \{\dim M \,|\, \iota_M^* G_{|M} > 0\}, \tag{20b}$$

$$\mathrm{ind}_0(G) = \max_{M\in\mathcal{L}(U)} \{\dim M \,|\, \iota_M^* G_{|M} = 0\}, \tag{20c}$$

which we call *negative index*, *positive index*, and *degree of degeneracy*. Moreover, we define the *signature*

$$\sigma(G) = \mathrm{ind}_+(G) - \mathrm{ind}_-(G). \tag{21}$$

Remark 8 Similar definitions apply to dp-operators. Instead of stating them twice we note that $K \in U \otimes U$ admits and interpretation as a pd-operator on U by looking at

$$K : U^* \to U \quad \text{as} \quad K : U^* \to U^{**} \mod U \hookrightarrow U^{**}.$$

In other words, we may think of K as a form on $U^* \times U^*$. All definitions for pd-operators such as positivity, signature, etc., apply *mutatis mutandis* in the context of dp-operators.

Definition 9 Let $G : U \to U^*$, $H : V \to V^*$ denote pd-operators, and $K : V^* \to V$ denote a dp-operator.

The operators G, H are called *congruent* if there exists a bijective operator $T : U \to V$ such that $G = T^*HT$. Similarly, the operators G and K are called *congruent* if $G = TK\tilde{T}$ for some bijective pd-operator $T : V \to U^*$.

Remark 10 Let $K = \sum_{i,j=1}^n \alpha_{ij} \, (y_j \otimes x_i) \in V \otimes V$ and $G : U \to U^*$ be congruent, i.e., $G = TK\tilde{T}$, for some bijective operator $T : V \to U^*$. There holds

$$(Gu)(v) = \sum_{i,j=1}^n \alpha_{ij}^* \, (Ty_j)(u)^* (Tx_i)(v). \tag{22}$$

In Theorem 11 we quote the celebrated *law of inertia* of Sylvester in this coordinate free context.

Theorem 11 *Let $G : U \to U^*$ denote a pd-operator congruent to $H : V \to V^*$ and $K : V^* \to V$. Then*

$$\sigma(G) = \sigma(H) = \sigma(K), \tag{23}$$

and all operators have the same degree of degeneracy.

Definition 12 Given a dp-operator $K : U^* \to V$ with

$$\operatorname{Im} K = Y \quad \text{and} \quad \operatorname{Im} \tilde{K} = X, \tag{24}$$

we denote the *truncated version of* K by K_* where

$$K_* : X^* \to Y \quad \text{with} \quad K_*(\ell_{|X}) = K\ell, \tag{25}$$

for all $\ell \in U^*$ is indeed well-defined. If $Y = X$ one says that K *lifts* X *into* U.

Theorem 13 *If $K : U^* \to U$ lifts X into U, then the truncated dp-operator $K_* \in X \otimes X$ is the reproducing kernel of the subspace $X \subseteq U$ with respect to $G = (K_*)^{-1}$. Moreover, the signatures of K and K_* are equal.*

Definition 14 Given a pd-operator $G : U \to V^*$ we define the *factor map* via

$$G_{\%} : U/X \to (V/Y)^*, \quad G_{\%}(u + X) := Gu + Y, \tag{26}$$

for all $u \in U$ where

$$X := \operatorname{Ker} G \quad \text{and} \quad Y := \operatorname{Ker} \tilde{G}. \tag{27}$$

Let $T_X : (U/X)^* \to X^\perp$ and $T_X : (V/Y)^* \to Y^\perp$ denote the natural linear identifications. The dp-operator

$$G_* : X^{\perp,*} \to Y^\perp, \quad \text{with} \quad G_* = T_Y G_{\%} \tilde{T}_X, \tag{28}$$

is called *the truncated version of* G. If $X^\perp = Y^\perp$ one says that G *lifts* $X^\perp$ *into* U^*.

Theorem 15 *If $G : U \to U^*$ lifts $X^\perp$ into U^*, then the truncated dp-operator $G_* \in X^\perp \otimes X^\perp$ is the reproducing kernel of the subspace $X^\perp \subseteq U^*$ with respect to $(G_*)^{-1}$. Moreover, the signatures of G and G_* are equal.*

A. Matrix Representations and Gramians

Matrix representations, especially in spaces that admit a natural choice of basis, provide a way to do computations in the context of forms. In particular they allow to compute signature and the the reproducing kernel of form. Before we discuss this we shall first fix the notation. The unique *matrix represenation* of an operator $T : U \to V$ with respect to the bases $B = \{b_1, \dots, b_m\}$ and $C = \{c_1, \dots, c_n\}$, of U and V, respectively, is denoted by $[T|^C_B] \in \mathbb{K}^{n\times m}$ and defined by

$$Tb_j = [T|^C_B]_{1j}\, c_1 + \dots + [T|^C_B]_{nj}\, c_n, \tag{29}$$

for all $j = 1, \dots, m$. For $u \in U$ we denote by $[u|^B_1] \in \mathbb{K}^m$ its *coordinate vector* with respect to the basis B or equivalently

$$u = [u|^B_1]_1 b_1 + \dots + [u|^B_1]_m b_m. \tag{30}$$

Definition 16 Let $B = \{b_1, \dots, b_n\}$ denote a basis of U. The *conjugate dual basis* of U^* with respect to $B \subseteq U$ is

$$\tilde{b}_i(u) = [u|^B_1]^* \quad \text{for all} \quad u \in U, \tag{31}$$

and denoted by $\tilde{B} = \{\tilde{b}_1, \dots, \tilde{b}_n\}$.

Theorem 17 *Let $K : U^* \to V$ denote a dp-operator. For any two bases B and C of U and V respectively, there holds*

$$K = \sum\nolimits_{i=1}^{C^\#} \sum\nolimits_{j=1}^{B^\#} [K|^C_{\tilde{B}}]_{ij} \cdot (b_j \otimes c_i), \tag{32}$$

and $[\tilde{K}|^B_{\tilde{C}}] = [K|^C_{\tilde{B}}]^*$. Moreover, there holds

$$\tilde{c}_i(K\tilde{b}_j) = [K|^C_{\tilde{B}}]_{ij} \quad \textit{and} \quad [\eta|^{\tilde{B}}_1]^* \cdot [\tilde{K}|^B_{\tilde{C}}] \cdot [\ell|^{\tilde{C}}_1]. \tag{33}$$

In particular, K self-adjoint if and only if $[K|^B_{\tilde{B}}]_{ij} = [K|^B_{\tilde{B}}]^*_{ji}$. In this case, $[K|^B_{\tilde{B}}]$ is diagonalizable and

$$\sigma(K) = \{\lambda \in \Lambda \mid \lambda > 0\}^\# - \{\lambda \in \Lambda \mid \lambda < 0\}^\#, \tag{34}$$

where Λ denotes the eigenvalues of $[K|^B_{\tilde{B}}]$. In particular the kernel $K \geq 0$ if and only if $[K|^B_{\tilde{B}}] \geq 0$.

Theorem 18 *Let B and C denote bases of the linear spaces U and V, respectively. Given a pd-operator $G : U \to V^*$ and $j \leq B^\#, i \leq C^\#$ there holds*

$$[G|^{\tilde{C}}_B]_{ij} = \langle b_j, c_i \rangle\,, \tag{35}$$

i.e., $[G|^{\tilde{C}}_B]$ is the Gramian of the form $\langle\cdot,\cdot\rangle : U \times V \to \mathbb{K}$ associated with the pd-operator G. If the associated form is non-degenerate, there holds

$$G^{-1} = \sum\nolimits_{i=1}^{C^\#} \sum\nolimits_{j=1}^{B^\#} ([G|^{\tilde{C}}_B]^{-1})_{ji} \cdot (c_i \otimes b_j).$$

Corollary 19 *Let $G : U \to U^*$ denote a pd-operator whose associated form $\langle\cdot,\cdot\rangle$ is an inner-product. The reproducing kernel of U with respect to G is given by*

$$K = \sum\nolimits_{i=1}^{n} b_i \otimes b_i, \tag{36}$$

where $b_1, \dots, b_n$ is arbitrary basis of U which is orthonormal with respect to the inner-product $\langle\cdot,\cdot\rangle$.

V. Exploiting Structured Spaces and Forms

Historically speaking, Hermitian forms, and therefore inner-products, have been developed by Hermite for the study of root location problems. The treatment back then was algebraic and centered around highly structured quadratic forms such as Bézoutians. In section V-A we introduce the language of intertwining operators to describe the structure of these forms. In Section V-B we study Hankel and Bézout-operators from this abstract point of view.

Quite recently the algebraic approach had a revival in linear system theory with the introduction of quadratic differential forms. These form inner-products on the state space. However, the state space admits no natural interpretation as a function space. Still, these quadratic differential forms admit an interpretation as kernels as we shall see in Section V-C.

A. Intertwining Maps and Forms

Definition 20 An operator $T : U \to V$ is said to *intertwine* the operators $A : U \to U$ and $B : V \to V$ if

$$TA = BT. \tag{37}$$

If T is bijective, then A and B are called *similar*. A subspace $X \subseteq U$ is called A*-invariant* if $Ax \in X$ for all $x \in X$. For A-invariant subspace $X \subseteq U$ the operator

$$A_{\%X} : U/X \to U/X \quad \text{via} \quad A_{\%X}(u+X) = Au+X, \tag{38}$$

is well-defined and called A *modulo* X.

Theorem 21 *Let A, B denote two operators $A : U \to U$ and $B : V \to V$ and consider a dp-operator K such that*

$$K : U^* \to V \quad \textit{with} \quad KA^* = BK. \tag{39}$$

Then, $\tilde{K}B^* = A\tilde{K}$. Moreover,

$$X = \operatorname{Im}\tilde{K} \subseteq U \quad \textit{and} \quad Y = \operatorname{Im}K \subseteq V, \tag{40}$$

are A- and B-invariant, respectively. The truncated kernel, given in Definition 12, satisfies

$$K_* \, A_{|X} = (B_{|Y})^* \, K_*. \tag{41}$$

In other words, the truncated kernel intertwines A restricted to X and the adjoint of B restricted to Y.

Theorem 22 *Let A, B denote two operators $A : U \to U$ and $B : V \to V$ and consider a pd-operator G such that*

$$G : U \to V^* \quad \textit{with} \quad GA = B^*G. \tag{42}$$

Then, $\tilde{G}B = A^*\tilde{G}$. Moreover,

$$X = \operatorname{Ker}G \subseteq U \quad \textit{and} \quad Y = \operatorname{Ker}\tilde{G} \subseteq V, \tag{43}$$

are A- and B-invariant, respectively. The factor map $G_\%$, given in Definition 14, satisfies

$$G_\% A_{\%X} = (B_{\%Y})^* G_\%, \tag{44}$$

i.e., the factor map intertwines A modulo $\operatorname{Ker}G$ and adjoint of B modulo $\operatorname{Ker}\tilde{G}$.

Theorem 23 *If $K \in X \otimes X$ is the reproducing kernel of a subspace $X \subseteq U$ with respect to $G : X \to X^*$, then K is intertwines A^*, B if and only if G is intertwines A, B^*.*

B. Hankel Forms and Bézoutian Kernels

Let $U := \mathbb{K}[x]$ denote the polynomial ring in the indeterminant x with coefficients in $\mathbb{K}$. Moreover, let S denote the *shift* S given by

$$S : U \to U \quad \text{with} \quad (Sp)(x) := x \cdot p(x). \tag{45}$$

For $q \in U$ let $\tilde{q}$ denote the unique polynomial such that

$$q(S)^* = \tilde{q}(S^*). \tag{46}$$

Then, $U \to U, q \mapsto \tilde{q}$ defines a conjugate linear map which conjugates the coefficients of its argument. By using the natural evalation we have $\tilde{q}(z) = q(z^*)^*$ for all $z \in \mathbb{K}$.

Theorem 24 *Let $Z \subseteq U$ denote a subspace and define the conjugate space $\tilde{Z} := \{\tilde{\zeta} \,|\, \zeta \in Z\}$. The following conditions are equivalent:*

1) *Z is S-invariant.*
2) *$\tilde{Z}$ is S-invariant.*
3) *$Z = \operatorname{Im} q(S)$ for some $q \in U$.*
4) *$\tilde{Z} = \operatorname{Im} \tilde{q}(S)$ for some $q \in U$.*

Definition 25 We call $H : U \to U^*$ a *Hankel operator* if it satisfies the Hankel functional equation

$$HS = S^*H. \tag{47}$$

Similarly, we call $K \in U \otimes U$ a *Bézoutian kernel* if K lifts $X = \operatorname{span}\{x^0, \ldots, x^n\}$ into U and

$$SK - KS^* \in \tilde{Z} \otimes Z, \tag{48}$$

for some S-invariant subspace $Z \subseteq U$ which is complementary to X, i.e., satisfies $X + Z = U$ with $X \cap Z = \{0\}$.

Theorem 26 *Let $H : U \to U^*$ denote a Hankel operator and $q \in U$ denote a polynomial with $\operatorname{Ker}H = \operatorname{Im}q(S)$. Then, $H = 0$ or H is of the form*[2]

$$(Hf)(g) = \lim_{\rho\to\infty} \frac{1}{2\pi\jmath} \oint_{|z|=\rho} \frac{f(z)\cdot\tilde{g}(z)}{q(z)}\, p(z)\mathrm{d}z, \tag{49}$$

where $p \in \mathbb{K}[x]$ is such that p and q are coprime and p/q is strictly proper. The rational function $p/q \in \mathbb{K}(x)$ is uniquely determined by H and called the symbol *of H. The adjoint of H is of the form*

$$(\tilde{H}f)(g) = \lim_{\rho\to\infty} \frac{1}{2\pi\jmath} \oint_{|z|=\rho} \frac{f(z)\cdot\tilde{g}(z)}{\tilde{q}(z)}\, \tilde{p}(z)\mathrm{d}z, \tag{50}$$

and $\operatorname{Ker}\tilde{H} = \operatorname{Im}\tilde{q}(S)$. The matrix $[H|^{\mathrm{st}}_{\mathrm{st}}]$ representing H w.r.t. the standard basis $x^0, \ldots, x^{n-1}$, with $n = \deg q$, is called Hankel matrix and satisfies,

$$[H|^{\mathrm{st}}_{\mathrm{st}}]_{i,j} = f_{i+j-1} \quad \textit{with} \quad f(z) = \sum\nolimits_{i=1}^{\infty} f_i z^{-i}, \tag{51}$$

for all $i, j = 1, \ldots, n$. In particular H is determined by the first $2n-1$ coefficients of the Laurent expansion of its symbol f around infinity.

In Theorem 27 we characterize Bézoutian kernels and comment on their relation to Hankel operators.

Theorem 27 *The kernel $K \in U \otimes U$ with*

$$K = \sum\nolimits_{i,j=1}^{n} \alpha_{ij}\,(y^j \otimes x^i) \tag{52a}$$

$$K(x,y) = \sum\nolimits_{i,j=1}^{n} \alpha_{ij}\, x^i y^j, \tag{52b}$$

and $\alpha_{ij} \in \mathbb{K}$ is called a Bézoutian kernel if and only if it satisfies one of the two equivalent conditions

$$(x-y)K(x,y) = q(x)d(y) - d(x)q(y), \quad \textit{or} \tag{53a}$$

$$SK - KS^* = \tilde{d}(y) \otimes q(x) - \tilde{q}(y) \otimes d(x), \tag{53b}$$

[2] The associated form of the Hankel operator is called *residual form* because $(Hf)(g)$ denotes the sum of the residues of $p(z)f(z)\tilde{g}(z)/q(z)$ at the zeroes of $q(z)$.

with $d, q \in U$ such that $\deg(q) > \deg(d)$. If this is the case, then K intertwines $S_{\tilde{q}}^*$ and S_q where $S_q, S_{\tilde{q}} : U \to U$ are operators given by

$$(S_q f)(x) = x \cdot f(x) \mod q(x), \tag{54a}$$
$$(S_{\tilde{q}} f)(x) = x \cdot f(x) \mod \tilde{q}(x), \tag{54b}$$

respectively. Moreover, K lifts $X = \operatorname{span}\{x^0, \dots, x^{n-1}\}$, with $n = \deg(q)$, into the ambient space U. In particular the truncated version K_* is the reproducing kernel of X with respect to K_*^{-1}. If d and q are coprime, then K_*^{-1} is congruent to a truncated Hankel operator H_* with symbol p/q where

$$p \cdot d \equiv 1 \mod q, \tag{55}$$

and $\operatorname{Ker} H = \operatorname{Im} q(S)$. In fact $(K_*^{-1} f)(g)$ is given by the right hand side of equation (49) for all $f, g \in X$.

Remark 28 The classical notation for (53) is given by

$$K(x, y) = \frac{q(x)d(y) - d(x)q(y)}{x - y}, \tag{56}$$

where one says that (56) is the generating function of the Bézoutian and the matrix $[K|_{\mathrm{\bar{s}t}}^{\mathrm{st}}]$ is called *Bézoutian matrix*.

C. Stability Theory

Given a polynomial $q \in \mathbb{K}[x]$ we define the corresponding *autonomous behavior* via

$$\mathcal{B} = \{w \in C^\infty(\mathbb{R}, \mathbb{K}) \mid q(\partial)w(t) = 0\}, \tag{57}$$

where $(\partial w)(t) := \dot{w}(t)$ for all $t \in \mathbb{R}$. Moreover, we define the time shift $\sigma^\tau : \mathcal{B} \to \mathcal{B}$ via

$$(\sigma^\tau w)(t) = w(t - \tau) \quad \text{for all} \quad w \in \mathcal{B}. \tag{58}$$

The behavior $\mathcal{B}$ is called *asymptotically stable* if $(\sigma^\tau w)(0) \to 0$ as $\tau \to \infty$ for all $w \in \mathcal{B}$. The natural question emerges, on what condition the coefficients of q have to satisfy in order for $\mathcal{B}$ to be asymptotically stable.

The origininal approach by Fuhrmann and Willems in [1] is centered around the Hankel operator H with symbol $1/q$ defined in (49) and restricted to

$$X := \{f \in \mathbb{K}[s] \mid \deg(f) < \deg(q)\}. \tag{59}$$

A bijective operator $\mathcal{F} : \mathcal{B} \to X^*$ which links the space $\mathcal{B}$ and the conjugate dual of X is given by

$$(\mathcal{F}w)(f) := (\tilde{f}(\partial)w)(0) \quad \text{for all} \quad f \in X, w \in \mathcal{B}. \tag{60}$$

Theorem 29 *Given $S_q, S_{\tilde{q}}$ as in (54). The operator defined in (60) satisfies*

$$(\mathcal{F}\partial w)(f) = (\mathcal{F}w)(S_{\tilde{q}} f) \quad \textit{and} \tag{61a}$$
$$(\mathcal{F}\sigma^\tau w)(f) = (\mathcal{F}w)(e^{S_{\tilde{q}}\tau} f), \tag{61b}$$

for all $\tau \in \mathbb{R}$, $f \in X$ and $w \in \mathcal{B}$. In particular

$$e^{S_{\tilde{q}}\tau} f \to 0 \quad (\tau \to \infty) \quad \textit{for all} \quad f \in X, \tag{62}$$

if and only if $\mathcal{B}$ is asymptotically stable.

Using the classical Lyapunov's stability test and state-space theory one verifies that (62) holds if and only if for some, and then any, operator $C : X \to \mathbb{K}$ where (A, C) is an obsevable pair, there exists a $Q : X \to X^*$ s.t.[3]

$$Q > 0 \quad \text{and} \quad S_{\tilde{q}}^* Q + Q S_{\tilde{q}} = -C^* C. \tag{63}$$

There exists unique $K \in X \otimes X$ and $\tilde{r} \in X$ such that $Q = H K \tilde{H}$ and $C = (1 \otimes \tilde{r})\tilde{H}$ which yields

$$C^* C = H(\tilde{r} \otimes 1)(1 \otimes \tilde{r})\tilde{H}, \tag{64a}$$
$$S_{\tilde{q}}^* Q + Q S_{\tilde{q}} = S_{\tilde{q}}^* H K \tilde{H} + H K \tilde{H} S_{\tilde{q}}. \tag{64b}$$

The observability of C is equivalent to the coprimeness of r and q. Multiplying (64a) and (64b) by H^{-1} from the left and $\tilde{H}^{-1}$ from the right we obtain

$$-\tilde{r} \otimes \tilde{r} = H^{-1} S_{\tilde{q}}^* H K + K \tilde{H} S_{\tilde{q}} \tilde{H}^{-1} \tag{65a}$$
$$= H^{-1} H S_q K + K S_q^* \tilde{H} \tilde{H}^{-1} \tag{65b}$$
$$= S_q K + K S_q^*. \tag{65c}$$

If such a $K \in X \otimes X$ exists, then $K(x, y) := \sum \alpha_{ij}\, x^i y^j$ is given by

$$K(x, y) = \frac{q(x)\tilde{d}(y) + d(x)\tilde{q}(y) - \tilde{r}(x)r(y)}{x + y}, \tag{66}$$

where the polynomial $d \in X$ must solve

$$q(x)\tilde{d}(-x) + d(x)\tilde{q}(-x) = \tilde{r}(x)r(-x), \tag{67}$$

to ensure that (66) indeed defines a polynomial in x and y. We arrive at the following result.

Theorem 30 *The behavior $\mathcal{B}$ is asymptotically stable if and only if for one, and then any, polynomial $r \in X$, such that q and r are coprime, the equation (67) admits a solution and $K(x, y) = \sum \alpha_{ij}\, x^i y^j$ given by (66) satisfies $\alpha > 0$.*

Since Q and K are congruent we have $Q > 0$ if and only if $K > 0$, i.e., the coefficient matrix α is positive definite.

In order to relate this to the quadratic differential form approach, initiated in [9] by Willems and Trentleman, we note that the differential form $\Psi : \mathcal{B} \to \mathcal{B}^*$ given by

$$(\Psi v)(w) = \sum \alpha_{ij}^* \, (\partial^i v)(0)\, (\partial^j w)(0)^* \quad \forall v, w \in \mathcal{B}, \tag{68}$$

is congruent to K via $\Psi = \tilde{\mathcal{F}} K \mathcal{F}$. Again we have $\Psi > 0$ if and only if α is positive definite. In particular $Q > 0$, $K > 0$, $\Psi > 0$ and $\alpha > 0$ are all equivalent.

Clearly, checking stability with Ψ instead of $Q = H K \tilde{H}$ is conceptualy more sound because it avoids the construction of the state space X and the auxiliuary Hankel operator H. The advantage of the Hankel form method is that standard results from state space theory can be directly applied to higher order differential equations without destroying structure. We did not discuss this, but we want to mention, that the Hankel form method is also more closely related to realization theory and standard companion matrices. The connection between both approaches is the congruence transformation $\mathcal{F}$.

[3]The map $C^* : \mathbb{K}^* \to X^*$ is regarded as a map $\mathbb{K} \to X^*$ by $\mathbb{K} \cong \mathbb{K}^*$.

VI. Conclusions

Reproducing kernels have many important applications not only in functional analysis, but in statistics, control theory and machine learning, as well. Their theory developed impressively in the last decades, but the classical results on kernels concentrated on Hilbert spaces of functions.

In the paper we presented a purely algebraic approach to reproducing kernels. This approach is versatile in the sense that it can handle bilinear and sesquilinear forms that are not necessarily positive definite. Moreover, it extends the existing reproducing kernel theory from linear function spaces to general linear spaces. We showed that our approach makes it easy to exploit algebraic structures induced by intertwining relations. As opposed to a function space point of view, in our approach quotient space structures form no obstacle. The Classical Bézoutian is shown to be a bona fide kernel in this context. Additionally, the importance of the Hankel operator as a congruence transformation has been highlighted. Future work is planned on adding suitable topological regularity conditions to handle kernels of infinite rank. Another point of interest is the extension of the theory to operator valued kernels. This means switching from the scalar-valued reproducing property to a vector valued one.

Acknowledgments

The authors would like to thank Quentin Rentmeester for the interesting discussions.

References

[1] P. A. Fuhrmann and J. C. Willems. Stability theory for high order equations. *Linear Algebra and its Applications*, 167:131–149, 1992.
[2] M. Hein and O. Bousquet. Kernels, associated structures and generalizations. Technical Report 127, Max Planck Institute for Biological Cybernetics, July 2004.
[3] P. Jordan and J. V. Neumann. On inner products in linear, metric spaces. *Annals of Mathematics*, 36(3):719–723, July 1935.
[4] E. H. Moore. On properly positive Hermitian matrices. *Bull. Amer. Math. Soc.*, 23(59):66–67, 1916.
[5] C. S. Ong. *Kernels: Regularization and Optimization*. PhD thesis, The Australian National University, April 2005.
[6] C. S. Ong and S. Canu and A. J. Smola. Learning with Non-Positive Kernels. *21st International Conference on Machine Learning*, 2004.
[7] S. Roman. *Advanced Linear Algebra*. 3rd edition, Springer, 2008.
[8] B. Schölkopf and A.J. Smola. *Learning with Kernels*. MIT Press, 2002.
[9] J. C. Willems and H. L. Trentelman. On quadratic differential forms. *SIAM Journal of Control and Optimization*, 36(5):1703–1749, 1998.