Probability and Stochastic Analysis in Reproducing Kernels and Division by Zero Calculus

Tsutomu Matsuura, Hiroshi Okumura and Saburou Saitoh

November 3, 2020

Abstract: Professor Rolin Zhang kindly invited in The 6th Int'l Conference on Probability and Stochastic Analysis (ICPSA 2021), January 5-7, 2021 in Sanya, China as a Keynote speaker and so, we will state the basic interrelations with reproducing kernels and division by zero from the viewpoint of the conference topics. The connection with reproducing kernels and Probability and Stochastic Analysis are already fundamental and well-known, and so, we will mainly refer to the basic relations with our new division by zero $1/0 = 0/0 = z/0 = \tan(\pi/2) = \log 0 = 0, [(z^n)/n]_{n=0} = \log z$, $[e^{(1/z)}]_{z=0} = 1$.

Key Words: Division by zero, division by zero calculus, probability, stochastic analysis, data analysis, reproducing kernel, $1/0 = 0/0 = z/0 = \tan(\pi/2) = \log 0 = 0, [(z^n)/n]_{n=0} = \log z$, $[e^{(1/z)}]_{z=0} = 1$.

AMS Mathematics Subject Classification: 60A06, 46J10, 00A05, 00A09, 42B20, 30E20.

1 Introduction

For the long history of division by zero, see [4, 40]. S. K. Sen and R. P. Agarwal [54] quite recentry referred to our paper [16] in connection with division by zero, however, their understandings on the paper seem to be not suitable (not right) and their ideas on the division by zero seem to be traditional, indeed, they stated as the conclusion of the introduction of the book in the following way:

"Thou shalt not divide by zero" remains valid eternally.

However, in [43] we stated simply based on the division by zero calculus that

We Can Divide the Numbers and Analytic Functions by Zero with a Natural Sense.

For the long tradition on the division by zero, people may not be accepted our new results against many clear evidences on our division by zero, however, a physicist stated typically as follows:

Here is how I see the problem with prohibition on division by zero, which is the biggest scandal in modern mathematics as you rightly pointed out (2017.10.14.08:55).

The common sense on the division by zero with the long and mysterious history is wrong and our basic idea on the space around the point at infinity is also wrong since Euclid. On the gradient or on differential coefficients we have a great missing since $\tan(\pi/2) = 0$. Our mathematics is also wrong in elementary mathematics on the division by zero. In a new and definite sense, we will show and give various applications of the division by zero $0/0 = 1/0 = z/0 = 0$. In particular, we introduced several fundamental concepts in calculus, Euclidean geometry, analytic geometry, complex analysis and differential equations. We saw new properties on the Laurent expansion, singularity, derivative, extension of solutions of differential equations beyond analytical and isolated singularities, and reduction problems of differential equations. On Euclidean geometry and analytic geometry, we found new fields by the concept of the division by zero. We gave many concrete properties in mathematical sciences from the viewpoint of the division by zero. We will know that the division by zero is our elementary and fundamental mathematics.

The contents in ([46]) are as follows:
1. Introduction.
2. Division by zero.
3. Division by zero calculus.
4. We can divide the numbers and analytic functions by zero.
5. General division and usual division.

6. Division by zero calculus, applications.
7. Derivatives of functions.
8. Differential equations.
9. Euclidean spaces and division by zero calculus.
10. Analytic functions and division by zero calculus.
11. The Descartes circle theorem.
12. Horn torus models and division by zero calculus – a new world.

For the conference topics, after some minimum statement of the above historical discoveries, we would like to state the basic interrelations with the conference topics and reproducing kernels.

2 Division by zero

The division by zero with the mysterious and long history was indeed trivial and clear as in the followings.

By the concept of the Moore-Penrose generalized solution of the fundamental equation $ax = b$, the division by zero was trivial and clear as $b/0 = 0$ in the **generalized fraction** that is defined by the generalized solution of the equation $ax = b$. Here, the generalized solution is always uniquely determined and the theory is very classical. See [16] for example.

Recall the uniqueness theorem by S. Takahasi on the division by zero. See [16, 59]:

Proposition 2.1 *Let F be a function from $\mathbf{C} \times \mathbf{C}$ to $\mathbf{C}$ such that*

$$F(a, b)F(c, d) = F(ac, bd)$$

for all

$$a, b, c, d \in \mathbf{C}$$

and

$$F(a, b) = \frac{a}{b}, \quad a, b \in \mathbf{C}, b \neq 0.$$

Then, we obtain, for any $a \in \mathbf{C}$

$$F(a, 0) = 0.$$

On the long mysterious history of the division by zero, this proposition seems to be decisive.

Following Proposition 2.1, we should **define**

$$F(b,0) = \frac{b}{0} = 0,$$

and we should consider that for the mapping

$$W = f(z) = \frac{1}{z}, \tag{2.1}$$

the image $f(0)$ of $z = 0$ is $W = 0$ (**should be defined from the form**). This fact seems to be a curious one in connection with our well-established popular image for the point at infinity on the Riemann sphere. As the representation of the point at infinity on the Riemann sphere by the zero $z = 0$, we will see some delicate relations between 0 and ∞ which show a strong discontinuity at the point of infinity on the Riemann sphere. We did not consider any value of the elementary function $W = 1/z$ at the origin $z = 0$, because we did not consider the division by zero $1/0$ in a good way. Many and many people consider its value at the origin by limiting like $+\infty$ and $-\infty$ or by the point at infinity as ∞. However, their basic idea comes from **continuity** with the common sense or based on the basic idea of Aristotele. – However, as the division by zero we will consider its value of the function $W = 1/z$ as zero at $z = 0$. We will see that this new definition is valid widely in mathematics and mathematical sciences, see ([20, 24]) for example. Therefore, the division by zero will give great impact to calculus, Euclidean geometry, analytic geometry, complex analysis and the theory of differential equations at an undergraduate level and furthermore to our basic idea for the space and universe.

The simple field structure containing division by zero was established by M. Yamada ([19]) in a natural way. For a simple introduction, H. Okumura [22] discovered the very simple essence that:

To divide by zero is to multiply by zero.

For the operator properties of the generalized fractions, see [59].

3 Background of division by zero calculus

We will recall the simple background on the division by zero calculus for differentiable functions based on ([50, 51]).

For a function $y = f(x)$ which is $n(n > 0)$ order differentiable at $x = a$, we will **define** the value of the function

$$\frac{f(x)}{(x-a)^n}$$

at the point $x = a$ by the value

$$\frac{f^{(n)}(a)}{n!}.$$

For the important case of $n = 1$,

$$\frac{f(x)}{x-a}|_{x=a} = f'(a). \tag{3.1}$$

In particular, the values of the functions $y = 1/x$ and $y = 0/x$ at the origin $x = 0$ are zero. **We write them as $1/0 = 0$ and $0/0 = 0$, respectively.** Of course, the definitions of $1/0 = 0$ and $0/0 = 0$ are not usual ones in the sense: $0 \cdot x = b$ and $x = b/0$. Our division by zero is given in this sense and is not given by the usual sense. However, we gave several definitions for $1/0 = 0$ and $0/0 = 0$. See, for example, [47].

In addition, when the function $f(x)$ is not differentiable, by many meanings of zero, we **should define** as

$$\frac{f(x)}{x-a}|_{x=a} = 0,$$

for example, since 0 **represents impossibility**. In particular, the value of the function $y = |x|/x$ at $x = 0$ is zero.

We will note its naturality of the definition.

Indeed, we consider the function $F(x) = f(x) - f(a)$ and by the definition, we have

$$\frac{F(x)}{x-a}|_{x=a} = F'(a) = f'(a).$$

Meanwhile, by the definition, we have

$$\lim_{x\to a} \frac{F(x)}{x-a} = \lim_{x\to a} \frac{f(x)-f(a)}{x-a} = f'(a). \tag{3.2}$$

For many applications, see the references cited in the reference.

The identity (3.1) may be regarded as an interpretation of the differential coefficient $f'(a)$ by the concept of the division by zero. Here, we do not use the concept of limitings. This means that NOT

$$\lim_{x \to a} \frac{f(x)}{x-a}$$

BUT

$$\frac{f(x)}{x-a}|_{x=a}.$$

Note that $f'(a)$ represents the principal variation of order $x-a$ of the function $f(x)$ at $x=a$ which is defined independently of $f(a)$ in (3.2). This is a basic meaning of the division by zero calculus $\frac{f(x)}{x-a}|_{x=a}$.

Following this idea, we can accept the formula, naturally, for also $n=0$ for the general formula; that is,

$$\frac{f(x)}{(x-a)^0}|_{x=a} = \frac{f^{(0)}(a)}{0!} = f(a).$$

In the expression (3.1), the value $f'(a)$ in the right hand side is represented by the point a, meanwhile the expression

$$\frac{f(x)}{x-a}|_{x=a} \tag{3.3}$$

in the left hand side, is represented by the dummy variable $x-a$ that represents the property of the function around the point $x=a$ with the sense of the division

$$\frac{f(x)}{x-a}.$$

For $x \neq a$, it represents the usual division.

When we apply the relation (3.1) to the elementary formulas for differentiable functions, we can imagine some deep results. For example, in the simple formula

$$(u+v)' = u' + v',$$

we have the result

$$\frac{u(x)+v(x)}{x-a}|_{x=a} = \frac{u(x)}{x-a}|_{x=a} + \frac{v(x)}{x-a}|_{x=a},$$

that is not trivial in our definition. This is a result from the property of derivatives.

In the following well-known formulas, we have some **deep meanings** on the division by zero calculus.

$$(uv)' = u'v + uv',$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

and the famous laws

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

and

$$\frac{dy}{dx} \cdot \frac{dx}{dy} = 1.$$

Note also the logarithm derivative, for $u, v > 0$

$$(\log(uv))' = \frac{u\prime}{u} + \frac{v\prime}{v}$$

and for $u > 0$

$$(u^v)' = u^v \left(v' \log u + v\frac{u'}{u} \right).$$

We note the basic relation for analytic functions $f(z)$ for the analytic extension of $f(x)$ to complex variable z

$$\frac{f(x)}{(x-a)^n}|_{x=a} = \frac{f^{(n)}(a)}{n!} = Res._{\zeta=a}\left\{ \frac{f(\zeta)}{(\zeta-a)^{n+1}} \right\}.$$

We therefore see the basic identities among the division by zero calculus, differential coefficients and residues in the case of analytic functions. Among these basic concepts, the differential coefficients are studied deeply and so, from the results of the differential coefficient properties, we can derive another results for the division by zero calculus and residures. See [51].

With this assumption, we can obtain many new results and new concepts.

Typically, we found a beautiful and important circle by this division by zero calculus, see [27] and [31].

However, for this assumption we have to check the results obtained whether they are reasonable or not. By this idea, we can avoid any logical problem. – In this viewpoint, **the division by zero calculus may be considered as an axiom.**

4 Probability and division by zero

The conditional probability $P(A|B)$ for the probability of A under the condition that B happens is given by the formula

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

If $P(B) = 0$, then, of course, $P(A \cap B) = 0$ and from the meaning, $P(A|B) = 0$ and so, $0/0 = 0$.

For a line

$$x \cos\theta + y \sin\theta - p = 0$$

and for data (x_j, y_j), the minimum of $\sum_{j=1}^n d_j^2$ for the distance d_j of the line and the point (x_j, y_j) is attained for the case

$$\tan 2\theta = \frac{2\gamma_{xy}\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2},$$

where

$$\gamma_{xy} = \frac{n\sum_j x_j y_j - (\sum_j x_j)(\sum_j y_j)}{n^2\sigma_x\sigma_y}$$

and

$$\sigma_x = \frac{1}{n}\sqrt{n\sum_j x_j^2 - (\sum_j x_j)^2}.$$

If $\sigma_x^2 = \sigma_y^2$, then $\theta = \pi/4$ from $\tan 2\theta = 0$.

Here, of course, for

$$\mu_x = \frac{1}{n}\sum x_j, \quad \mu_y = \frac{1}{n}\sum y_j,$$

$$\sigma_x^2 = \frac{1}{n}\sum (x_j - \mu_x)^2, \quad \sigma_y^2 = \frac{1}{n}\sum (y_j - \mu_y)^2$$

and

$$\sigma_{xy}^2 = \frac{1}{n}\sum (x_j - \mu_x)(y_j - \mu_y),$$

$$\tan 2\theta = \frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}.$$

5 Inverse document frequency and $\log 0 = \log\infty = 0$

For any fixed complex number a, we will consider the sector domain $\Delta_a(\alpha, \beta)$ defined by

$$0 \le \alpha < \arg(z-a) < \beta < 2\pi$$

on the complex z plane and we consider the conformal mapping of $\Delta_a(\alpha, \beta)$ into the complex W plane by the mapping

$$W = \log(z-a).$$

Then, the image domain is represented by

$$S(\alpha, \beta) = \{W; \alpha < \Im W < \beta\}.$$

Two lines $\{W; \Im W = \alpha\}$ and $\{W; \Im W = \beta\}$ usually were considered as having the common point at infinity, however, in the division by zero, the point is represented by zero.

Therefore, $\log 0$ and $\log\infty$ **should be defined as zero**. Here, $\log\infty$ is precisely given in the sense of $[\log z]_{z=\infty}$. However, the properties of the logarithmic function should not be expected more, we should consider the value only. For example,

$$\log 0 = \log(2 \cdot 0) = \log 2 + \log 0$$

is not valid.

In particular, in many formulas in physics, in some expression, for some constants A, B

$$\log\frac{A}{B}, \tag{5.1}$$

if we consider the case that A or B is zero, then we should consider it in the form

$$\log\frac{A}{B} = \log A - \log B, \tag{5.2}$$

and we should put zero in A or B. Then, in many formulas, we will be able to consider the case that A or B is zero. For the case that A or B is zero, the identity (5.1) is not valid, then the expression $\log A - \log B$ may be valid in many physical formulas. However, the results are case by case, and **we**

should check the obtained results for applying the formula (5.2) **for** $A = 0$ **or** $B = 0$**.** Then, we will be able to enjoy the formula apart from any logical problems as in the applications of the division by zero and division by zero calculus. See [18] for many examples.

We will show a typical example.

The inverse document frequency is a measure of how much information the word provides, i.e., if it's common or rare across all documents. It is the logarithmically scaled inverse fraction of the documents that contain the word (obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient):

$$idf(t, D) = \log \frac{N}{|\{d \in D; t \in d\}|}$$

where N is the total number of documents in the corpus $N = |D|$ and $\{d \in D; t \in d\}$ is the number of documents where the term appears (i.e., $tf(t, d) \neq 0$). If the term is not in the corpus, this will lead to a division-by-zero. It is therefore common to adjust the denominator to $idf(t, D) = \log N$, that is just our statement. See for the details

> tf–idf - Wikipedia https://en.wikipedia.org/wiki/Tf–idf Traduzir esta página In information retrieval, tf–idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a ... Motivations · Definition · Justification of idf · Example of tf–idf.

6 Identification of the image of a linear mapping: fundamental of reproducing kernels

For a basic reference on reproducing kernels, see [42].

Let $\mathcal{F}(E)$ be the linear space consisting of all complex-valued functions on any fixed abstract set E. Let $\mathcal{H}$ be a Hilbert space with the inner product

$\langle\cdot,\cdot\rangle_{\mathcal{H}}$. Let $\mathbf{h} : E \to \mathcal{H}$ be a fixed $\mathbf{h}$ valued mapping on E. Then, we consider the linear mapping L from $\mathbf{f} \in \mathcal{H}$ into $\mathcal{F}(E)$ defined by

$$L\mathbf{f}(p) \equiv \langle \mathbf{f}, \mathbf{h}(p)\rangle_{\mathcal{H}}. \tag{6.1}$$

The most fundamental problem in the linear mapping (6.1) is to characterize the image $L\mathbf{f}(p)$ of L and our next concern will be to grasp how the input $\mathbf{f}$ and the output $L\mathbf{f}(p)$ are related.

The key to consider these fundamental problems is to form the two variables complex-valued function; that is, a positive definite quadratic form function

$$K(p,q) \equiv \langle \mathbf{h}(q), \mathbf{h}(p)\rangle_{\mathcal{H}} \tag{6.2}$$

defined on $E \times E$. We denote by $\mathcal{R}(L)$ the linear function space consisting of all complex-valued functions of the images of $\mathbf{h}$ by L defined on E. In the image space $\mathcal{R}(L)$, we will introduce the norm by

$$\|f\|_{\mathcal{R}(L)} = \inf\{\|\mathbf{f}\|_{\mathcal{H}} : \mathbf{f} \in \mathcal{H}, \quad f = L\mathbf{f}\}, \tag{6.3}$$

and then the image space will form a Hilbert space; indeed, we obtain precisely

Fundamental Theorem for Linear Mappings;
Generalized Pythagorean Theorem;
Fundamental Theorem of Reproducung Kernels:

1. *By* (6.3) *we introduce the norm in* $\mathcal{R}(L)$ *and furthermore, from the norm we introduce the inner product* $\langle\cdot,\cdot\rangle_{\mathcal{R}(L)}$ *to* $\mathcal{R}(L)$ *to make* $\mathcal{R}(L)$ *into a Hilbert space.*

2. *The function* K *defined by* (6.2) *enjoys three properties*:

   (a) *For any* $q \in E$,
   $$K_q \in \mathcal{R}(L) \tag{6.4}$$
   where $K_q(p) = K(p,q)$ *for* $p \in E$.

   (b) *The function* K *has the reproducing property*;
   $$f(q) = \langle f, K_q\rangle_{\mathcal{R}(L)} \tag{6.5}$$
   for any function $f \in \mathcal{R}(L)$ *and for any point* $q \in E$.

(c) The mapping L is isomorphic from $\mathcal{H}$ to $(\mathcal{R}(L), \langle\cdot,\cdot\rangle_{\mathcal{R}(L)})$, that is,

$$\|L\mathbf{f}\|_{\mathcal{R}(L)} = \|\mathbf{f}\|_{\mathcal{H}} \quad (\mathbf{f} \in \mathcal{H}), \tag{6.6}$$

if and only if $\{\mathbf{h}(p) : p \in E\}$ spans a dense subspace of $\mathcal{H}$.

3. The function K satisfying (6.4) *and* (6.5) *is unique.*

For any positive definite quadratic form function K on E, we can construct a Hilbert space $\mathcal{H}$ and a mapping $\mathbf{h}$ from E into $\mathcal{H}$ satisfying (6.2) by using the Gaussian probability distribution with nth order co-distributions and zero mean that are given by $n \times n$ matrices $\{K(p_j, p_{j'})\}_{j,j'}$. For this construction, we use the **Kolmogorov theorem on measures** whose proof is not elementary [33]. This theorem is very important, from which we can derive simply the fundamental

Theorem: *For any positive definite quadratic form function $K : E \times E \to \mathbb{C}$, there exists a uniquely determined reproducing kernel Hilbert space $H_K = H_K(E)$ admitting the reproducing kernel K on E.*

Furthermore, this Kolmogorov factorization theorem [1, 15, 21] will essentially be indispensable when we introduce various operators on a Hilbert space and among Hilbert spaces.

This important result was interestingly derived from the theory of stochastic theory independent of the theory of reproducing kernels. Furthermore, such factorization representation of a positive definite quadratic form function may be used conversely to realize reproducing kernel Hilbert spaces admitting reproducing kernels.

For some deep theory on the Kolmogorov factorization theorem, see, for example, [6, 7, 8]. Here, note that for a positive definite quadratic form function $K(p,q)$ on E, when we know its reproducing kernel Hilbert space, it is easy to show that K can be written in the form (6.2). Indeed. it is given simply by

$$K(p,q) = \langle K_q, K_p \rangle_{H_K(E)} \tag{6.7}$$

and such a factorization is determined among isometric Hilbert spaces.

7 Random fields estimations

We will consider the following situation;

1. X is a set,

2. $(\Omega, \mathcal{F}, P)$ is a probability space.

We assume that the random field is of the form

$$u(x) = s(x) + n(x) \quad (x \in X), \tag{7.1}$$

where, for each $x \in X$, $s(x) = s(x; \cdot) : \Omega \to \mathbb{R}$ is the useful signal and $n(x) = n(x; \cdot) : \Omega \to \mathbb{R}$ is a noise. Note that $s(x)$ and $n(x)$ are not necessarily independent for each $x \in X$. Without loss of generality, we can assume that the mean values of $u(x)$ and $n(x)$ are zero. We assume that the covariance functions

$$R(x, y) = E[u(x)u(y)] \quad (x, y \in X) \tag{7.2}$$

and the information

$$f(x, y) = E[u(x)s(y)] \quad (x, y \in X) \tag{7.3}$$

are known.

Here and below, we identify X with the structure of the measure space: let (X, dm) be a measure space.

In addition, we shall consider the general form of a linear estimation $\hat{u}$ of u in the form

$$\hat{u}(x) = \int_X u(y)h(x, y)dm(y) = \langle u, h(x, \cdot)\rangle_{L^2(X,dm)} \tag{7.4}$$

for an $L^2(X, dm)$ space and for a function $h(x, \cdot)$ belonging to $L^2(X, dm)$ for any fixed $x \in E$. For the desired information $As : X \times \Omega \to \mathbb{R}$, which satisfies

$$A(ks) = kAs,\ A(s_1 + s_2) = As_1 + As_2$$

for all $s, s_1, s_2 : X \times \Omega \to \mathbb{R}$ and $k \in \mathbb{C}$, we wish to determine the function $h(x, t)$ attaining

$$\inf\{E[(\hat{u}(x) - As(x))^2] : \hat{u} \text{ is given by (7.4) and } h(x, \cdot) \in L^2(X, dm)\} \tag{7.5}$$

which gives the minimum of the variance by the least squares method.

Many topics in filtering and estimation theory in signal and image processing, underwater acoustics, geophysics, optical filtering, etc., which were initiated by N. Wiener (1894–1964), will be presented in this framework.

Then, we see that the linear transform $h(x,t)$ is given by the integral equation

$$\int_X R(x',y)h(x,y)dm(y) = f(x',x). \tag{7.6}$$

Therefore, our random fields estimation problems will be reduced to finding the inversion formula

$$f \mapsto h \tag{7.7}$$

in our framework. So, our general method for integral transforms will be applied to these problems. For this situation and other topics and methods for the inversion formulas, see [38] for the details.

8 Support vector machines and probability theory for data analysis

Let T be a set. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $L^2(\Omega, \mathcal{B}, P)$ the Hilbert space of composing of the square integrable random variables on Ω with the inner product $E[X\overline{Y}]$. Let $X(t) = X(t,\cdot), t \in T$ be a square integrable stochastic process defined on the probability space $(\Omega, \mathcal{B}, P)$. We set the mean value function as $m(t) = E[X(t)]$. Then, the second moment function

$$R(t,s) = E[X(t)\overline{Y(s)}] \quad t,s \in T \tag{8.1}$$

and the covariance function

$$K(t,s) = \mathrm{Cov}[X(t),Y(s)] = E[(X(t)-m(t))\overline{(Y(s)-m(s))}] \quad t,s \in T \tag{8.2}$$

are positive definite quadratic form functions on Ω and so, the both theories of stochastic processes and reproducing kernels have the fundamental relationship. A typical result is the **Loéve' theorem:** The Hilbert space $H(X)$ generated by the process $X(t)$, t on a set T with the covariance function R is *congruent* to the reproducing kernel Hilbert space admitting the kernel R.

The support vector machine is a powerful computational method for solving learning and function estimating problems such as pattern recognition, density and regression estimation and operator inversion. The basic idea can be found in E. Parzen [36], A. Berlinet and C. Thomas-Agnan [2, p. 248] and Vapnik [60]. See also the recent book Steinwart and Christmann [58].

From some data input space E we consider a general non-linear mapping to a feature space F that is a pre-Hilbert space with the inner product $\langle \cdot, \cdot \rangle_F$:

$$\Phi : E \to F; \quad x \mapsto \Phi(x). \tag{8.3}$$

Then, we form the positive definite quadratic form function

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_F. \tag{8.4}$$

The important point of this method is that we can apply this kernel to the problem of construction of the optimal hyperplanes in the space F not by using the explicit values of the transformed data $\Phi(x)$. See B.E. Boser, I.M. Guyon, and V.N. Vapnik [5] and V.N. Vapnik [60].

A new method appeared as the *kernel method*. We consider the transform of the data in the probability space $(\Omega, \mathcal{B}, P)$ for a reproducing kernel Hilbert space $H_K(\Omega)$ admitting a kernel K on Ω:

$$\Psi : \Omega \to \langle \cdot, \cdot \rangle_F; \quad x \mapsto K_x \tag{8.5}$$

and we can apply the theory of reproducing kernels to the probability problems on the space $(\Omega, \mathcal{B}, P)$.

On the whole space $\mathbb{R}^m$ the following kernels are typical:

1. The usual inner product is given by $k(x_1, x_2) = x_1^T x_2$.

2. For $c \geq 0$ and for a positive integer d

$$k_{d,c}^{\mathrm{poly}}(x_1, x_2) = (x_1^T x_2 + c)^d \quad x_1, x_2 \in \mathbb{R}^m.$$

3. The Gauss kernel, for $\sigma > 0$

$$k_\sigma^G(x_1, x_2) = \exp\Bigl(-\frac{|x_1 - x_2|^2}{2\sigma^2}\Bigr) \quad x_1, x_2 \in \mathbb{R}^m.$$

In statistical learning theory, reproducing kernel Hilbert spaces are used basically as the hypothesis space in the approximation of the regression functions. See, for example, the books [9, 60]. Here, in connection with a basic formula by F. Cucker and S. Smale [10] which is fundamental in the approximation error estimates, we shall consider a general formula based on the general theory of reproducing kernels. For related results, see also [52, 53, 61].

Acknowledgements

The authors wish to express their deep thanks the organizing committee and in particular Professor Rolin Zhang of The 6th Int'l Conference on Probability and Stochastic Analysis (ICPSA 2021) for their honorable invitation and kind suggestions and arrangements for our presentation.

References

[1] G. D. Allen, F. J. Narcowich and J.P. Williams, *An operator version of a theorem of Kolmogorov,* Pacific J. Math. **61** (1975), no. 2, 305–312.

[2] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics,* Kluwer Akademic Publishers, Boston/Dordrecht/London (2004).

[3] A. Berlinet, *Reproducing kernels in probability and statistics,* More Progresses In Analysis, Proceedings of the 5th International ISAAC Congress, (2010), 153–162.

[4] C. B. Boyer, An early reference to division by zero, The Journal of the American Mathematical Monthly, **50** (1943), (8), 487- 491. Retrieved March 6, 2018, from the JSTOR database.

[5] B. E. Boser, I. M. Guyon, and V. N. Vapnik. *A training algorithm for optimal margin classifiers,* In D. Haussler, editor, *Fifth Annual ACM Workshop on Computational Learning Theory*, 144–152, Pittsburgh, PA, 1992. ACM Press.

[6] T. Constantinescu and A. Gheondea, *Representations of Hermitian kernels by means of Krein spaces. II. Invariant kernels,* Commun. Math. Phys. **216**, No.2 (2001), 409–430.

[7] T. Constantinescu and A. Gheondea, *Invariant Hermitian kernels and their Kolmogorov decompositions,* (English. Abridged French version) C. R. Acad. Sci., Paris, Ser. I, Math. **331**, No.10 (2000), 797–802.

[8] T. Constantinescu and A. Gheondea, *Representations of Hermitian kernels by means of Krein spaces,* Publ. RIMS Kyoto Univ. **33** (1997), 917–951.

[9] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-based Learning Methods,* Cambridge Univ. Press, Cambridge (2001).

[10] F. Cucker and S. Smale, *On the mathematical foundations of learning,* Bull. Amer. Math. Soc., **39** (2002), 1–49.

[11] M. Cui and Y. Lin, *Nonlinear Numerical Analysis in the Reproducing Kernel Space,* Nova Science Publisher: New York. (2009).

[12] M. Cuturi, K. Fukumizu and J. P. Vert, *Semigroup kernels on measures,* Journal of Machine Learning Research, **6** (2005), 1169–1198.

[13] W. W. Däumler, H. Okumura, V. V. Puha and S. Saitoh, Horn Torus Models for the Riemann Sphere and Division by Zero, viXra:1902.0223 submitted on 2019-02-12 18:39:18.

[14] R. Kaplan, THE NOTHING THAT IS A Natural History of Zero, OXFORD UNIVERSITY PRESS (1999).

[15] A. N. Kolmogoroff, *Stationary sequences in Hilbert's space*, (in Russian) Bolletin Moskovskogo Gosudarstvenogo Universiteta, Matematika **2** (1941). 40pp.

[16] M. Kuroda, H. Michiwaki, S. Saitoh and M. Yamane, New meanings of the division by zero and interpretations on $100/0 = 0$ and on $0/0 = 0$, Int. J. Appl. Math. **27** (2014), no 2, pp. 191-198, DOI: 10.12732/ijam.v27i2.9.

[17] T. Matsuura and S. Saitoh, Matrices and division by zero $z/0 = 0$, Advances in Linear Algebra & Matrix Theory, **6**(2016), 51-58 Published Online June 2016 in SciRes. http://www.scirp.org/journal/alamt http://dx.doi.org/10.4236/alamt.2016.62007.

[18] T. Matsuura, H. Michiwaki and S. Saitoh, $\log 0 = \log \infty = 0$ and applications, Differential and Difference Equations with Applications, Springer Proceedings in Mathematics & Statistics, **230** (2018), 293-305.

[19] H. Michiwaki, S. Saitoh and M.Yamada, Reality of the division by zero $z/0 = 0$, IJAPM International J. of Applied Physics and Math. **6**(2015), 1–8. http://www.ijapm.org/show-63-504-1.html

[20] H. Michiwaki, H. Okumura and S. Saitoh, Division by Zero $z/0 = 0$ in Euclidean Spaces, International Journal of Mathematics and Computation, **2**8(2017); Issue 1, 1-16.

[21] K. Nakade, T. Ohwada and K. S. Saito, *Kolmogorov's factorization theorem for von Neumann algebras*, J. Math. Anal. Appl. **401** (2013), no. 1, 289–292.

[22] H. Okumura, To Divide by Zero is to Multiply by Zero, viXra:1811.0132., (2018).

[23] H. Okumura, A Chain of Circles Touching a Circle and Its Tangent and Division by Zero, viXra:2001.0034 submitted on 2020-01-03 01:08:58.

[24] H. Okumura, S. Saitoh and T. Matsuura, Relations of 0 and ∞, Journal of Technology and Social Science (JTSS), **1**(2017), 70-77.

[25] H. Okumura and S. Saitoh, The Descartes circles theorem and division by zero calculus, https://arxiv.org/abs/1711.04961 (2017.11.14).

[26] H. Okumura and S. Saitoh, Harmonic Mean and Division by Zero, Dedicated to Professor Josip Pečarić on the occasion of his 70th birthday, Forum Geometricorum, **18** (2018), 155—159.

[27] H. Okumura and S. Saitoh, Remarks for The Twin Circles of Archimedes in a Skewed Arbelos by H. Okumura and M. Watanabe, Forum Geometricorum, **18**(2018), 97-100.

[28] H. Okumura and S. Saitoh, Applications of the division by zero calculus to Wasan geometry, GLOBAL JOURNAL OF ADVANCED RESEARCH ON CLASSICAL AND MODERN GEOMETRIES" (GJARCMG), **7**(2018), 2, 44–49.

[29] H. Okumura and S. Saitoh, Wasan Geometry and Division by Zero Calculus, Sangaku Journal of Mathematics (SJM), **2** (2018), 57–73.

[30] H. Okumura and S. Saitoh, Values of the Riemann Zeta Function by Means of Division by Zero Calculus viXra:1907.0437 submitted on 2019-07-23 20:48:54.

[31] H. Okumura, Pappus Chain and Division by Zero Calculus, viXra:2006.0095 submitted on 2020-06-11 17:00:55.

[32] H. Okumura and S. Saitoh, Division by Zero Calculus and Euclidean Geometry - Revolution in Euclidean Geometry viXra:2010.0228 submitted on 2020-10-28 21:39:06.

[33] K. R. Parthasarathy and K. Schmidt, *Positive Definite Kernels, Continuous Tensor Products, and Central Limit Theorems of Probability Theory*, Springer Lecture Notes in Math., Berlin, **107** (1972).

[34] E. Parzen, *Extraction and detection problems and reproducing kernel Hilbert spaces*, J. SIAM control Ser. A 1 (1962), 35–62.

[35] E. Parzen, *Probability density functionals and reproducing kernel Hilbert spaces*, 1963 Proc. Sympos. Time Series Analysis (Brown Univ., 1962) pp. 155–169 Wiley, New York.

[36] E. Parzen, *Statical inference on time series by RKHS methods.* Proc. of 12th Biennial Seminar of the Canadian Mathematical Congress, Amer. Math. Soc., Providence, R. I. (1971), 1–37.

[37] S. Pinelas and S. Saitoh, Division by zero calculus and differential equations. Differential and Difference Equations with Applications. Springer Proceedings in Mathematics & Statistics, **230** (2018), 399-418.

[38] A. G. Ramm, *Random fields estimation theory*, Pitman Monographs and Surveys in Pure and Applied Mathematics, Longman Scientific and Technical, UK, **48**(1990).

[39] A. G. Ramm, *Multidimensional inverse scattering problems*, Monographs and Surveys in Pure and Applied Mathematics, **51** Longman Scientific and Technical, UK (1992).

[40] H. G. Romig, Discussions: Early History of Division by Zero, American Mathematical Monthly, **3**1, No. 8. (Oct., 1924), 387-389.

[41] S. Saitoh and Y. Sawano, Theory of Reproducing Kernels and Applications, Developments in Mathematics **44**, Springer (2016).

[42] S. Saitoh, A reproducing kernel theory with some general applications, Qian,T./Rodino,L.(eds.): Mathematical Analysis, Probability and Applications - Plenary Lectures: Isaac 2015, Macau, China, Springer Proceedings in Mathematics and Statistics, **177**(2016), 151-182.

[43] S. Saitoh, We Can Divide the Numbers and Analytic Functions by Zero with a Natural Sense, viXra:1902.0058 submitted on 2019-02-03 22:47:53.

[44] S. Saitoh, Zero and Infinity; Their Interrelation by Means of Division by Zero, viXra:1902.0240 submitted on 2019-02-13 22:57:25.

[45] S. Saitoh, Division by Zero Calculus in Trigonometric Functions, viXra:1903.0566 submitted on 2019-03-31 15:59:03.

[46] S. Saitoh, What Was Division by Zero?; Division by Zero Calculus and New World, viXra:1904.0408 submitted on 2019-04-22 00:32:30.

[47] S. Saitoh, Fundamental of Mathematics; Division by Zero Calculus and a New Axiom, viXra:1908.0100 submitted on 2019-08-06 20:03:01.

[48] S. Saitoh, Division by Zero Calculus in Multiply Dimensions and Open Problems (An Extension), viXra:1906.0185 submitted on 2019-06-11 20:12:46.

[49] S. Saitoh, Essential Problems on the Origins of Mathematics; Division by Zero Calculus and New World Authors, viXra:1912.0300 submitted on 2019-12-16 18:37:53.

[50] S. Saitoh, Division by Zero Calculus for Differentiable Functions - L'Hôpital's Theorem Versions, viXra:2001.0091 submitted on 2020-01-06 17:52:07.

[51] S. Saitoh, Division by Zero Calculus, Derivatives and Laurent's Expansion, viXra:2001.0586 submitted on 2020-01-27 16:28:38.

[52] B. Schölkopf, R. Herbrich and A. J. Smola, A Generalized Representer Theorem, D. Helmbold and B. Williamson (Eds.): COLT/EuroCOLT 2001, LNAI 2111, pp. 416–426, 2001. Springer-Verlag Berlin Heidelberg 2001.

[53] B. Schölkopf and A. J. Smola. *Learning with Kernels.* MIT Press, (2002).

[54] S.K.S. Sen and R. P. Agarwal, ZERO A Landmark Discovery, the Dreadful Volid, and the Unitimate Mind, ELSEVIER (2016).

[55] F. Soddy, The Kiss Precise, Nature **137**(1936), 1021. doi:10.1038/1371021a0.

[56] E. Sondheimer and A. Rogerson, NUMBERS AND INFINITY A Historical Account of Mathematical Concepts, Dover (2006) unabridged republication of the published by Cambridge University Press, Cambridge (1981).

[57] J. Stewart, *Positive definite functions and generalizations, an historical survey*, Rocky Mountain J. Math. **6** (1976), no. 3, 409–434.

[58] I. Steinwart and A. Christmann, *Support Vector Machine*, Springer (2008).

[59] S.-E. Takahasi, M. Tsukada and Y. Kobayashi, Classification of continuous fractional binary operations on the real and complex fields, Tokyo Journal of Mathematics, **38**(2015), no. 2, 369-380.

[60] V. Vapnik, *Statistical Learning Theory,* John Wiley & Sons (1998).

[61] D. X. Zhou, *Capacity of reproducing kernel spaces in learning theory,* IEEE Trans. Inform. Theory, **49** (2003), 1743–1752.

Tsutomu Matsuura: Gunma University Center for Mathematics and Data Science,
Showa-machi, 3-39-22, Maebashi-shi, 371-8511, Japan
matsuura@gunma-u.ac.jp

Hiroshi Okumura: Maebashi-shi, 371-0123, Japan
hokmr@yandex.com
and

Saburou Saitoh: Institute of Reproducing Kernels,
Kawauchi-cho, 5-1648-16, Kiryu 376-0041, Japan
saburou.saitoh@gmail.com