Kernel Analog Forecasting and its applications

Reviewing the state of theoretical research and applying the techniques
to a range of datasets

by

Kieran John Taylor

Thesis for the

Masters of Mathematics

Under the supervision of Prof. Dr. Anke Pohl

at

University of Bremen

Abstract

This is where I will write my abstract.

CONTENTS

1 Introduction

We will first work through proving the main theorems of [1].

2   Kernel Analog Forecasting

2.1   Forecasting Situation

2.1.1 Dynamical System

Let $(\Omega, \mathcal{F}, \mu)$ be a **Probability Space**, with Ω being the space of all possible states, $\mathcal{F}$ a σ-algebra of distinguished subsets of Ω, and $\mu : \mathcal{F} \rightarrow [0, 1]$ a **Probability Measure**. We now define a **Dynamical System** on Ω through the semigroup of measurable maps: $\{\Phi^t : \Omega \rightarrow \Omega\}_{t \geq 0}$, which evolve any initial state $\omega_0 \in \Omega$ to $\omega_t \in \Omega$. While the semigroup property stipulates that we can split these maps $\Phi^{t_1 + t_2} = \Phi^{t_1} \circ \Phi^{t_2}$, measurability ensures that for each $t \geq 0$ and $\forall A \in \mathcal{F}$

$$\Phi^{-t} A = (\Phi^t)^{-1} A := \{\omega \in \Omega : \Phi^t(\omega) \in A\} \in \mathcal{F}$$

Furthermore, we assume that these maps are also measure-preserving;

$$\forall B \in \mathcal{F}, t \geq 0, \mu(\Phi^{-t} B) = \mu(B)$$

Define the functional space $\mathbb{L}^2(\mu)$ as the space of square integrable functions from $\Omega \rightarrow \mathbb{C}$, with respect to the probability measure μ. Using the standard representation, we then define $L^2(\mu)$ as the Hilbert Space of μ.a.e. equivalence functions in $\mathbb{L}^2(\mu)$. The space is equipped with the inner product $\langle f_1, f_2 \rangle_{L^2(\mu)} = \int_\Omega \overline{f_1(\omega)} f_2(\omega) d\mu(\omega)$, which is conjugate-linear in the first argument.

Having defined our general probability space, we now move onto the forecasting task. We define a **Measurable Covariate Space** $(\mathcal{X}, \Sigma_{\mathcal{X}})$ as the tuple of a space, $\mathcal{X}$, which, while not necessarily linear, has the structure of a metric space, coupled with a Borel σ-algebra $\Sigma_{\mathcal{X}}$.

Furthermore, a **Measurable Response Space** $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$ is a Hilbert Space over the complex numbers $\mathcal{Y}$, with a Borel σ-algebra $\Sigma_{\mathcal{Y}}$. In order to use the kernel trick, which is central for KAF, $\mathcal{Y}$ is assumed to be $\mathbb{C}$, with $\Sigma_{\mathcal{Y}}$ being the Borel σ-algebra generated by the topology induced by open balls. The inner product $\langle y_1, y_2 \rangle_{\mathcal{Y}} = \overline{y_1} y_2$ is taken to be conjugate-linear in the first argument, as $L^2(\mu)$ above.

For any $\omega \in \Omega$, we then define the data-producing measurable covariance function:

$$X : \Omega \to \mathcal{X}$$
$$\omega \mapsto x_\omega := X(\omega)$$

and, similarly, the data-producing measurable response function:

$$Y : \Omega \to \mathcal{Y} = \mathbb{C}$$
$$\omega \mapsto y_\omega := Y(\omega)$$

Our aim is, for any given **Lead Time** $\tau > 0$, to find a target function $f_\tau : \mathcal{X} \to \mathbb{C}$, such that $f_\tau \circ X$ approximates $Y \circ \Phi^\tau$. We use the norm induced by the inner-product of $L^2(\mu)$ to evaluate the accuracy of such a target function:

$$\left\| f_\tau \circ X - Y \circ \Phi^\tau \right\|^2_{L^2(\mu)} = \int_\Omega \left\| \left(f_\tau \circ X_t \right)(\omega) - (Y \circ \Phi^\tau)(\omega) \right\|^2_{\mathbb{C}} d\mu(\omega) \tag{1}$$

2.1.2 KOOPMAN OPERATOR

We are now presented with the problem that for any $t \geq 0$, Φ^t is not an intrinsically linear operator. We therefore use the **Koopman Operator** $U^t : L^2(\mu) \to L^2(\mu)$, named after the author of the paper in which it was first introduced [2]. The Koopman Operator allows us to express, for any $f \in L^2(\mu)$, the dynamical action in Ω as:

$$U^t f = f \circ \Phi^t$$

Unlike Φ^t, U^t is an intrinsically linear operator on $L^2(\mu)$, as $\forall a, b \in \mathbb{C}, f, g \in L^2(\mu)$

$$U^\tau(af + bg) = aU^\tau f + bU^\tau g$$

Lemma 2.1. *As the dynamical system is measure-preserving, the Koopman operator is a unitary operator on $L^2(\mu)$.*

Proof. TODO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

As U^t are unitary operators for all $t \geq 0$, we can freely shift the starting point of our dynamical system ω_0 forwards by any arbitrary amount $\tau \geq 0$:

$$\left\| f_t \circ X - U^t Y \right\|_{L^2(\mu)}^2 = \left\| U^\tau \left(f_t \circ X - U^t Y \right) \right\|_{L^2(\mu)}^2$$
$$= \left\| f_t \circ X \circ \Phi^\tau - U^t Y \circ \Phi^\tau \right\|_{L^2(\mu)}^2$$

As such, the starting point ω_0 for our KAF predictions need not be fixed.

2.1.3 SUB SIGMA-ALGEBRA

Let us now define a measure μ_X on $\mathcal{X}$, as the pushforward of μ onto $\mathcal{X}$ through X:

$$\forall S \in \Sigma_X : \mu_X(S) = \mu(X^{-1}S) \tag{2}$$

This is well-defined, as X is a μ-measurable function. We can therefore also define the Hilbert-Space $L^2(\mu_X)$ of equivalence classes of square-integrable functions from $\mathcal{X}$ to $\mathbb{C}$. Furthermore, X also induces a Sub σ-Algebra $\mathcal{G} \subseteq \mathcal{F}$ on Ω, with $\mathcal{G} := X^{-1}(\Sigma_X)$. We now define $\mathbb{L}^2_\mathcal{G}(\mu)$ (and subsequently $L^2_\mathcal{G}(\mu)$) as the set of square-integrable functions that take constant values on subsets of Ω where X is constant. We label such a function $g \in \mathbb{L}^2_\mathcal{G}(\mu)$ as being $\mathcal{G}$-**measurable** and

$$\forall T \in \mathcal{F} \setminus \mathcal{G}, g(T) = 0$$

Lemma 2.2. *For any $\mathcal{G}$-measurable function $g \in \mathbb{L}^2_\mathcal{G}(\mu)$, there exists an $f : \mathcal{X} \to \mathcal{Y}$ s.t. $g = f \circ X$.*

Proof. TODO $\qquad\qquad\square$

We can therefore define an **Isometric Embedding** with range $L^2_\mathcal{G}(\mu)$:

$$\Xi : L^2(\mu_X) \to L^2(\mu)$$
$$f \mapsto f \circ X$$

We also define the accompanying adjoint operator:

$$\Xi^* : L^2(\mu) \to L^2(\mu_X)$$

Such that $\forall g \in L^2_\mathcal{G}(\mu), g := f \circ X, \Xi^*(g) = f$ and $\ker \Xi^* = \left(L^2_\mathcal{G}(\mu) \right)^\perp$

2.1.4 Regression Function

Lemma 2.3. *For $U^\tau \in L^2(\mu)$ there exists a unique $\mathcal{G}$-measurable element $Z_\tau \circ X \in L^2(\mu)$ such that for all $g \in L^2_{\mathcal{G}}(\mu)$*

$$\left\langle g, U^\tau Y \right\rangle_{L^2(\mu)} = \left\langle g, Z_\tau \circ X \right\rangle_{L^2(\mu)} = \int_\Omega \overline{g(\omega)} \cdot (Z_\tau \circ X)(\omega) d\mu(\omega)$$

Proof. TODO

- Paper says to use Radon-Nikodym Theorem □

We define the **Orthogonal Projection** $\Pi_{\mathcal{G}} : L^2(\mu) \to L^2(\mu)$ as the mapping into the closed and convex subspace $L^2_{\mathcal{G}} \subseteq L^2(\mu)$. We can therefore also observe the unique element $Z_\tau \circ X$, which we label the **Conditional Expectation**, as the orthogonal projection of $U^\tau Y$ onto $L^2_{\mathcal{G}}(\mu)$.

$$Z_\tau \circ X = \Pi_{\mathcal{G}} U^\tau Y \tag{3}$$

Furthermore, as $Z_\tau \circ X \in L^2_{\mathcal{G}}$, then $\Xi Z_\tau = Z_\tau \circ X$. Consequntly, the adjoint of the isometric embedding Ξ^* defines a unique $Z_\tau \in L^2(\mu_X)$, which from now on will be known as the **Regression Function**:

$$Z_\tau = \Xi^* U^\tau Y \tag{4}$$

The regression function is the unique element that minimizes the Mean Squared Error in Equation 1:

$$\underset{f \in L^2_{\mathcal{G}}(\mu)}{\arg\min} \| f \circ X - U^\tau Y \|_{L^2(\mu)} = Z_\tau$$

We define the **Intrinsic System Error** $\sigma_\tau = \| Z_\tau \circ X - U^\tau Y \|_{L^2(\mu)}$ as the error that is inherent to our system, or the amount of information that is lost due to our choice of covariate and response functions. This error will always appear and is independent of our search for suitable forecasting functions. Consequently, our task can be formulated as trying to produce functions that consistently approximate the regression function Z_τ.

2.2 Reproducing Kernel Hilbert Spaces

In the following section we define and construct a subspace of functions $\mathcal{H} \subseteq \mathbb{L}^2(\mu)$ with a convenient structure to simplify the search for optimal approximations of the regression function. We must first, however, define the **Inclusion Map**:

$$\iota : \mathbb{L}^2(\mu_X) \to L^2(\mu_X)$$

This linear map associates a function f to its equivalence class ιf. We may now use this to adapt the mean squared error (Equation 1) to the **Generalization Error** $\mathscr{E}_\tau : \mathbb{L}^2(\mu_X) \to \mathbb{R}$:

$$\mathscr{E}_\tau(f) = \|\iota f \circ X - U^\tau Y\|_{L^2(\mu)}^2 \tag{5}$$

To ease the formulation of the following lemma, let us define the **Excess Generalization Error** as:

$$\mathscr{A}_\tau(f) = \|\iota f - Z_\tau\|_{L^2(\mu_X)}^2 \tag{6}$$

Lemma 2.4. *We can decompose the generalization error into the excess generalization error and the intrinsic system error:*

$$\mathscr{E}_\tau(f) = \mathscr{A}_\tau(f) + \sigma_\tau$$

Proof. TODO ☐

As the intrinsic system error is independent of our prediction function f, the task of minimizing $\mathscr{E}_\tau$ is equivalent to minimizing $\mathscr{A}_\tau$.

2.2.1 RKHS

Definition 2.5 (RKHS). *For each $x \in \mathcal{X}$, let $L_x : \mathcal{K} \to \mathbb{C}$ be the evaluation function defined by $L_x f = f(x)$. The space $\mathcal{K}$ is an RKHS if L_x is bounded at every $x \in \mathcal{X}$.*

Theorem 2.6 (Riesz Representation Theorem). *For every $x \in \mathcal{X}$, there exists some function $k_x \in \mathcal{K}$ such that $\forall f \in \mathcal{K}$*

$$L_x f = \langle k_x, f \rangle_{\mathcal{K}}$$

Proof. BLACK BOX ☐

Due to this reproducing property of $\mathcal{K}$, we are able to define the **Reproducing Kernel** function:

$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$$
$$(x_1, x_2) \mapsto \langle k_{x_1}, k_{x_2} \rangle_{\mathcal{K}}$$

Lemma 2.7. *The reproducing kernel is*
(i) Conjugate Symmetric
(ii) Positive Definite

Proof. TODO ☐

Theorem 2.8 (Moore-Aronszajn Theorem)**.** *For any conjugate-symmetric, positive-definite kernel function k, there exists a unique RKHS on $\mathcal{X}$ for which k is the reproducing kernel.*

There is a one-to-one correspondence between kernels and RKHSs.

Proof. BLACK BOX $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

2.3 Convergence of Target Function in RKHS

Theorem 2.9. *Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be an $L^2(\mu_X)$-strictly-positive kernel with corresponding RKHS $\mathcal{H}$. Then, for any response variable $U^\tau Y \in L^2(\mu)$ and lead time $\tau \geq 0$, as $l \to \infty$, the target functions $f_{\tau,l}$ converge to the conditional expectation $Z_\tau \circ X$, in the sense of the excess generalization error (Equation 6)*

$$\lim_{l\to\infty} \|\iota f_{\tau,l} - Z_\tau\|^2_{L^2(\mu_X)} = 0.$$

Proof. TODO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

2.4 Main Theory

Theorem 2.10 (Uniform Convergence of KAF Target Function)**.** *Under the basic assumptions for convergence, for every lead time $\tau = q\Delta t, q \in \mathbb{N}_0$, and hypothesis space dimension l such that $\lambda_l > 0$, the KAF target function $f_{\tau,l,n} \in \mathcal{H}_n$ converges as $n \to \infty$ to the ideal target function $f_{\tau,l} \in \mathcal{H}$, uniformly on $\mathcal{U}$. Moreover, if the reproducing kernel k of $\mathcal{H}$ is $L^2(\mu_X)$-strictly-positive-definite, then by Theorem 2.9, $f_{\tau,l,n}$ converges to the regression function Z_τ associated with the conditional expectation, $Z_\tau \circ X$, in the sense of the iterated limit*

$$\lim_{l\to\infty}\lim_{n\to\infty} f_{\tau,l,n} = \lim_{l\to\infty} f_{\tau,l} = Z_\tau$$

Here, the $n \to \infty$ and $l \to \infty$ limits are taken in $C(\mathcal{U})$ and $L^2(\mu_X)$ norm, respectively. Moreover, the convergence is uniform with respect to τ lying in compact sets.

Proof. TODO $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

Acronyms

KAF   Kernel Analog Forecasting

RKHS   Reproducing Kernel Hilbert Space

Glossary

$\mathbb{R}$ The set of real numbers

$(\Omega, \mathscr{F}, \mu)$ Probability space

Ω Space of all possible initial states

$\mathscr{F}$ $\sigma-$algebra of distinguished subsets of Ω

$\mu : \mathscr{F} \to \infty$ Probability Measure

BIBLIOGRAPHY

1. R. Alexander and D. Giannakis. "Operator-Theoretic Framework for Forecasting Nonlinear Time Series with Kernal Analog Techniques". *Physica D: Nonlinear Phenomena* 409, 2020, p. 132520. ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2020.132520. URL: https://www.sciencedirect.com/science/article/pii/S016727891930377X.

2. B. O. Koopman. "Hamiltonian Systems and Transformation in Hilbert Space". *Proceedings of the National Academy of Sciences* 17:5, 1931, pp. 315–318. DOI: 10.1073/pnas.17.5.315. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.17.5.315. URL: https://www.pnas.org/doi/abs/10.1073/pnas.17.5.315.