Ridge reconstruction of partially observed functional data is asymptotically optimal

David Kraus, Marco Stefanucci

Ridge reconstruction of partially observed functional data is asymptotically optimal[*]

David Kraus[†] and Marco Stefanucci[‡]

May 28, 2020

Abstract: When functional data are observed on parts of the domain, it is of interest to recover the missing parts of curves. Kraus (2015) proposed a linear reconstruction method based on ridge regularization. Kneip and Liebl (2020) argue that an assumption under which Kraus (2015) established the consistency of the ridge method is too restrictive and propose a principal component reconstruction method that they prove to be asymptotically optimal. In this note we relax the restrictive assumption that the true best linear reconstruction operator is Hilbert–Schmidt and prove that the ridge method achieves asymptotic optimality under essentially no assumptions. The result is illustrated in a simulation study.

Key words and phrases: Functional data; partial observation; reconstruction; reproducing kernel Hilbert space; ridge regularization.

1 Introduction

We consider functional data consisting of continuous curves in the separable Hilbert space $L^2(\mathcal{I})$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$, where $\mathcal{I}$ is a compact interval. Each of the n independent curves $X_1, \ldots, X_n$ is observed on a set $O_i \subseteq \mathcal{I}$, while no information about the curves is available on the missing sets $M_i = \mathcal{I} \setminus O_i$. This type of partially observed functions was previously considered by, e.g., Bugni (2012), Delaigle and Hall (2013), Liebl (2013), Gellar et al. (2014), Goldberg et al. (2014), Kraus (2015), Delaigle and Hall (2016), Gromenko et al. (2017), Dawson and Müller (2018), Mojirsheibani and Shaw (2018), Stefanucci et al. (2018), Descary and Panaretos (2019), Kneip and Liebl (2020), Kraus (2019), Kraus and Stefanucci (2019) or Liebl and Rameseder (2019). In this paper we deal with function reconstruction (completion), that is, with the task to estimate, or rather predict the missing parts of curves from the observed parts.

We assume that the curves have finite fourth moments and that the observation domains O_i, consisting of finite unions of intervals, are mutually independent random sets that are independent of X_i.

To simplify the presentation, we study the reconstruction of the missing part X_M using the observed part X_O of a new curve X that is not part of the observed dataset $X_{1O_1}, \ldots, X_{nO_n}$ used for the training of the prediction method. Our results apply to the usual case, where one reconstructs the curves in the training sample, as well. To further simplify the presentation we assume without loss of generality that the curves have mean zero.

In general, the best reconstruction of X_M as a function of X_O is the conditional expectation, which is usually difficult to analyze. An optimal solution is, therefore, sought among linear reconstructions. That is, one seeks to minimize the mean squared reconstruction error $\mathbb{E}(\|X_M - \mathscr{A}(X_O)\|^2)$ over linear operators $\mathscr{A} : L^2(O) \to L^2(M)$.

Motivated by the similarity of the linear reconstruction problem with linear function-on-function regression, Kraus (2015) solved the theoretical optimization task over bounded linear operators. The solution to the normal equation $\mathscr{A}\mathscr{R}_{OO} = \mathscr{R}_{MO}$ is $\mathscr{A} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{-1}$, provided it is bounded. Here we denote by $\mathscr{R}_{OO} : L^2(O) \to L^2(O)$ and $\mathscr{R}_{MO} : L^2(O) \to L^2(M)$ the integral operators with kernels $\rho_{OO}(s,t) = \rho(s,t)$, $s,t \in O$ and $\rho_{MO}(s,t) = \rho(s,t)$, $s \in M$, $t \in O$, respectively, where $\rho(s,t) = \mathrm{cov}\{X(s), X(t)\}$ is assumed to be continuous. Since this linear inverse problem is ill-posed ($\mathscr{R}_{OO}^{-1}$ is unbounded because $\mathscr{R}_{OO}$ is compact), Kraus (2015) proposed to employ ridge regularization and use $\mathscr{A}^{(\alpha)} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{(\alpha)-1}$, where $\mathscr{R}_{OO}^{(\alpha)} = \mathscr{R}_{OO} + \alpha\mathscr{I}_O$ with $\alpha > 0$ being a parameter and $\mathscr{I}_O$ the identity operator on $L^2(O)$. In the empirical version of the problem, the covariance function is replaced by the estimator (4) defined below, leading to the estimator of the optimal reconstruction $\tilde{X}_M = \mathscr{A}(X_O)$ by

$$\hat{X}_M^{(\alpha)} = \hat{\mathscr{A}}^{(\alpha)}(X_O) = \hat{\mathscr{R}}_{MO}(\hat{\mathscr{R}}_{OO} + \alpha\mathscr{I}_O)^{-1}X_O. \tag{1}$$

Under the additional assumption that $\mathscr{A} = \mathscr{R}_{MO}\mathscr{R}_{OO}^{-1}$ is a Hilbert–Schmidt operator, Kraus (2015) proved that the estimated reconstruction $\hat{X}_M^{(\alpha_n)}$ consistently (in the mean square sense) estimates the best linear reconstruction $\tilde{X}_M$, that is, $\mathbb{E}(\|\hat{X}_M^{(\alpha_n)} - \tilde{X}_M\|^2) \to 0$ as $n \to \infty$, provided that $\alpha_n \to 0$ at an appropriate rate, namely, α_n satisfies $\alpha_n n^{1/3} \to \infty$. In the general case of processes with non-zero mean the reconstruction is $\hat{X}_M^{(\alpha)} = \hat{\mu}_M + \hat{\mathscr{R}}_{MO}(\hat{\mathscr{R}}_{OO} + \alpha\mathscr{I}_O)^{-1}(X_O - \hat{\mu}_O)$, where $\hat{\mu}$ is defined in (3) and $\hat{\mu}_O(t) = \hat{\mu}(t)$, $t \in O$, $\hat{\mu}_M(t) = \hat{\mu}(t)$, $t \in M$.

Kneip and Liebl (2020) revisit the problem of function reconstruction and argue that optimal linear reconstruction generally need not be achievable by a Hilbert–Schmidt or even bounded linear operator. Therefore, they consider general linear operators from $L^2(O)$ to $L^2(M)$ with the minimal restriction that $\mathrm{var}\,\mathscr{A}(X_O)(u) < \infty$ for all $u \in M$ (they call such operators reconstruction operators). They show that any such operator can be represented as $\mathscr{A}(X_O)(u) = \langle b_u, X_O\rangle_{\mathscr{R}_{OO}}$, where b_u is a unique element of the Reproducing Kernel Hilbert Space $\{f \in L^2(O) : \|f\|_{\mathscr{R}_{OO}} < \infty\}$ with inner product $\langle f, g\rangle_{\mathscr{R}_{OO}} = \sum_{j=1}^{\infty}\langle f, \varphi_{OOj}\rangle\langle g, \varphi_{OOj}\rangle/\lambda_{OOj}$ and norm $\|f\|_{\mathscr{R}_{OO}} = \langle f, f\rangle_{\mathscr{R}_{OO}}^{1/2}$. Here λ_{OOj} and φ_{OOj} are the eigenvalues and eigenfunctions of $\mathscr{R}_{OO}$. The optimal reconstruction operator then takes the form

$$\tilde{X}_M(u) = \mathscr{A}(X_O)(u) = \langle \rho_{uO}, X_O\rangle_{\mathscr{R}_{OO}} = \sum_{j=1}^{\infty}\langle \rho_{uO}, \varphi_{OOj}\rangle\langle X_O, \varphi_{OOj}\rangle/\lambda_{OOj}, \tag{2}$$

where $\rho_{uO}(t) = \rho(u,t)$, $t \in O$. See also Section 2.4.1 of Berlinet and Thomas-Agnan (2004) for such results in the general setting and Shin and Hsing (2012) in the context of linear prediction from functional data.

Since Hilbert–Schmidt operators from $L^2(O)$ to $L^2(M)$ (considered in Kraus, 2015) correspond to integral operators, such a reconstruction operator takes the form $\tilde{X}_M = \mathscr{A}(X_O) = \int_O a(\cdot, t) X_O(t) dt$, where a is a square integrable function on $M \times O$. In particular, the reconstruction of $X_M(u)$ is the bounded linear functional $\tilde{X}_M(u) = \mathscr{A}(X_O)(u) = \langle a_u, X_O \rangle$, where $a_u = a(u, \cdot) \in L^2(O)$. If the optimal reconstruction operator $\mathscr{A} = \mathscr{R}_{MO} \mathscr{R}_{OO}^{-1}$ is Hilbert–Schmidt, then its kernel can be written in terms of principal components of $\mathscr{R}_{OO}$ as $a(u, t) = \sum_{j=1}^{\infty} \langle \rho_{uO}, \varphi_{OOj} \rangle \varphi_{OOj}(t) / \lambda_{OOj}$. Hence

$$\tilde{X}_M(u) = \langle a_u, X_O \rangle = \langle \mathscr{R}_{OO}^{-1} \rho_{uO}, X_O \rangle = \sum_{j=1}^{\infty} \langle \rho_{uO}, \varphi_{OOj} \rangle \langle X_O, \varphi_{OOj} \rangle / \lambda_{OOj}.$$

The latter form of the series is the same as (2) in the general case but the property $a_u = \mathscr{R}_{OO}^{-1} \rho_{uO} \in L^2(O)$ in the Hilbert–Schmidt case is stronger than the general property $\rho_{uO} \in \mathrm{RKHS}(\mathscr{R}_{OO})$, which is equivalent to $\mathscr{R}_{OO}^{-1/2} \rho_{uO} \in L^2(O)$. Equivalently, for the optimal reconstruction operator to be Hilbert–Schmidt it must hold that $\sum_{j=1}^{\infty} \langle \rho_{uO}, \varphi_{OOj} \rangle^2 / \lambda_{OOj}^2 < \infty$, while in general the less restrictive condition $\sum_{j=1}^{\infty} \langle \rho_{uO}, \varphi_{OOj} \rangle^2 / \lambda_{OOj} < \infty$ is needed. Kneip and Liebl (2020) discuss a prominent example of non-Hilbert–Schmidt optimal reconstruction depending on the evaluation of the observed function at a single impact point.

Motivated by the series form (2) of the optimal reconstruction, Kneip and Liebl (2020) offer a solution based on spectral truncation, i.e., they take $K < \infty$ terms in the series in (2) and replace unknown quantities by consistent estimators (namely local linear kernel estimators). Under regularity assumptions they establish that such a reconstruction, say $\hat{X}_M^{(K)}$, consistently estimates the true optimal reconstruction as $n \to \infty$ and $K \to \infty$ at an appropriate rate.

It is interesting to notice that the spectrally truncated reconstruction $\hat{X}_M^{(K)}$ is in fact Hilbert–Schmidt but it nevertheless consistently estimates the optimal reconstruction $\tilde{X}_M$ even if the latter is not even a bounded operator of X_O. The ridge reconstruction $\hat{X}_M^{(\alpha)}$ of Kraus (2015) given in (1) is also Hilbert–Schmidt but its asymptotic optimality was established in that paper under the additional assumption that the optimal reconstruction is Hilbert–Schmidt too. It is worth investigating whether this additional assumption is redundant and ridge reconstruction method also asymptotically achieves optimality in the general case. The answer we give in the next section is positive. We show that from this point of view the ridge method is not inferior to the principal component method of Kneip and Liebl (2020).

Section 2 gives the main theoretical result on the asymptotic optimality of ridge reconstruction and Section 3 illustrates this property by simulation under scenarios not covered by the theory in Kraus (2015).

2   Main result

We provide a proof of the asymptotic optimality result that, unlike in Kraus (2015), does not require the restrictive assumption of point reconstructions being bounded linear functionals (or $\mathscr{A}$ being a Hilbert–Schmidt operator). The reconstruction method based on ridge regularization that we analyze here is the same as in Kraus (2015) but the theoretical result is stronger than the one in Kraus (2015).

We denote by $O_i(t)$ the indicator that the value of $X_i(t)$ is observed and use the following estimators in the empirical reconstruction operator. The mean function is estimated by the cross-sectional average of available observations

$$\hat{\mu}(t) = \frac{J(t)}{N(t)} \sum_{i=1}^{n} O_i(t) X_i(t), \tag{3}$$

where $N(t) = \sum_{i=1}^{n} O_i(t)$ and $J(t) = 1_{[N(t)>0]}$. The covariance function is estimated by the empirical covariance using pairwise complete observations

$$\hat{\rho}(s,t) = \frac{I(s,t)}{M(s,t)} \sum_{i=1}^{n} U_i(s,t)\{X_i(s) - \hat{\mu}_{st}(s)\}\{X_i(t) - \hat{\mu}_{st}(t)\}, \tag{4}$$

where $U_i(s,t) = O_i(s)O_i(t)$, $M(s,t) = \sum_{i=1}^{n} U_i(s,t)$, $I(s,t) = 1_{[M(s,t)>0]}$ and

$$\hat{\mu}_{st}(s) = \frac{1_{[M(s,t)>0]}}{M(s,t)} \sum_{i=1}^{n} U_i(s,t) X_i(s).$$

The estimator $\hat{\hat{\mathscr{R}}}$ with kernel $\hat{\rho}$ has been shown to be root-n consistent for $\mathscr{R}$ in Kraus (2015, Proposition 1) under the assumption on the observation mechanism that

$$\text{there exists } \delta > 0 \text{ such that } \sup_{(s,t)\in\mathcal{I}^2} P\left(n^{-1} \sum_{i=1}^{n} U_i(s,t) \leq \delta\right) = O(n^{-2}) \text{ as } n \to \infty. \tag{5}$$

A further study of the asymptotic distribution of $\hat{\hat{\mathscr{R}}}$ and related tests can be found in Kraus (2019).

Our main result is as follows. Note that it holds without any restrictive assumptions on the optimal reconstruction operator or, for example, on the eigenvalue sequence.

Theorem 1. *Assume that $\mathbb{E}(\|X_i\|^4) < \infty$ and the observation pattern satisfies (5). Then, with $\hat{X}_M^{(\alpha_n)}$, $\tilde{X}_M$ defined in (1), (2), respectively, for $\alpha_n \to 0$ such that $\alpha_n n^{1/3} \to \infty$ as $n \to \infty$,*

$$\mathbb{E}[\{\hat{X}_M^{(\alpha_n)}(u) - \tilde{X}_M(u)\}^2] \to 0$$

for all $u \in M$, and also $\mathbb{E}\{\|\hat{X}_M^{(\alpha_n)} - \tilde{X}_M\|^2\} \to 0$. The result holds regardless of whether the optimal linear reconstruction operator $X_O \mapsto \tilde{X}_M$ is Hilbert–Schmidt (i.e., whether the optimal linear reconstruction functionals $X_O \mapsto \tilde{X}_M(u)$ are bounded) or even bounded.

Proof. Denote $\tilde{X}_M^{(\alpha_n)} = \mathscr{A}^{(\alpha_n)} X_O$. Compute the regularization error

$$\begin{aligned}
\mathbb{E}[\{\tilde{X}_M^{(\alpha_n)}(u) - \tilde{X}_M(u)\}^2] &= \mathbb{E}\{\langle \mathscr{R}_{OO}\mathscr{R}_{OO}^{(\alpha_n)-1}\rho_{uO} - \rho_{uO}, X_O\rangle_{\mathscr{R}_{OO}}^2\} \\
&= \|\mathscr{R}_{OO}\mathscr{R}_{OO}^{(\alpha_n)-1}\rho_{uO} - \rho_{uO}\|_{\mathscr{R}_{OO}}^2 \\
&= \sum_{j=1}^{\infty} \frac{\langle \mathscr{R}_{OO}\mathscr{R}_{OO}^{(\alpha_n)-1}\rho_{uO} - \rho_{uO}, \varphi_{OOj}\rangle^2}{\lambda_{OOj}} \\
&= \sum_{j=1}^{\infty} \left(\frac{\lambda_{OOj}}{\lambda_{OOj} + \alpha_n} - 1\right)^2 \frac{\langle \rho_{uO}, \varphi_{OOj}\rangle^2}{\lambda_{OOj}} \\
&= \sum_{j=1}^{\infty} \left(\frac{\alpha_n}{\lambda_{OOj} + \alpha_n}\right)^2 \frac{\langle \rho_{uO}, \varphi_{OOj}\rangle^2}{\lambda_{OOj}}.
\end{aligned}$$

This converges to 0 as $n \to \infty$ by the dominated convergence theorem because $\sum_{j=1}^{\infty} \langle \rho_{uO}, \varphi_{OOj} \rangle^2 / \lambda_{OOj} < \infty$. We now turn to the estimation error

$$
\begin{aligned}
\mathbb{E}[\{\hat{X}_M^{(\alpha_n)}(u) - \tilde{X}_M^{(\alpha_n)}(u)\}^2] &= \mathbb{E}\{\langle \hat{\mathscr{R}}_{OO}^{(\alpha_n)-1} \hat{\rho}_{uO} - \mathscr{R}_{OO}^{(\alpha_n)-1} \rho_{uO}, X_O \rangle^2\} \\
&= \mathbb{E}\{\|\mathscr{R}_{OO}^{1/2}(\hat{\mathscr{R}}_{OO}^{(\alpha_n)-1} \hat{\rho}_{uO} - \mathscr{R}_{OO}^{(\alpha_n)-1} \rho_{uO})\|^2\} \\
&\leq \mathbb{E}[\{\|\mathscr{R}_{OO}^{1/2} \hat{\mathscr{R}}_{OO}^{(\alpha_n)-1}(\hat{\rho}_{uO} - \rho_{uO})\| + \|\mathscr{R}_{OO}^{1/2}(\hat{\mathscr{R}}_{OO}^{(\alpha_n)-1} - \mathscr{R}_{OO}^{(\alpha_n)-1})\rho_{uO}\|\}^2].
\end{aligned}
$$

First, using $\|\cdot\|_\infty$ for the operator norm, we see that

$$
\mathbb{E}\{\|\mathscr{R}_{OO}^{1/2} \hat{\mathscr{R}}_{OO}^{(\alpha_n)-1}(\hat{\rho}_{uO} - \rho_{uO})\|^2\} \leq \mathbb{E}\{\|\mathscr{R}_{OO}^{1/2}\|_\infty^2 \|\hat{\mathscr{R}}_{OO}^{(\alpha_n)-1}\|_\infty^2 \|\hat{\rho}_{uO} - \rho_{uO}\|^2\} = \lambda_{OO1} \alpha_n^{-2} \mathbb{E}\{\|\hat{\rho}_{uO} - \rho_{uO}\|^2\},
$$

which converges to 0 because $\mathbb{E}\{\|\hat{\rho}_{uO} - \rho_{uO}\|^2\} = O(n^{-1})$. Second, compute

$$
\begin{aligned}
\mathbb{E}\{\|\mathscr{R}_{OO}^{1/2}(\hat{\mathscr{R}}_{OO}^{(\alpha_n)-1} - \mathscr{R}_{OO}^{(\alpha_n)-1})\rho_{uO})\|^2\} &= \mathbb{E}\{\|\mathscr{R}_{OO}^{1/2} \hat{\mathscr{R}}_{OO}^{(\alpha_n)-1}(\hat{\mathscr{R}}_{OO} - \mathscr{R}_{OO})\mathscr{R}_{OO}^{(\alpha_n)-1}\rho_{uO}\|^2\} \\
&\leq \mathbb{E}\{\|\mathscr{R}_{OO}^{1/2} \hat{\mathscr{R}}_{OO}^{(\alpha_n)-1}\|_\infty^2 \|\hat{\mathscr{R}}_{OO} - \mathscr{R}_{OO}\|_\infty^2 \|\mathscr{R}_{OO}^{(\alpha_n)-1}\|_\infty^2 \|\rho_{uO}\|^2\} \\
&= \alpha_n^{-1} \mathbb{E}\{\|\hat{\mathscr{R}}_{OO} - \mathscr{R}_{OO}\|_\infty^2\} \alpha_n^{-2} \|\rho_{uO}\|^2,
\end{aligned}
$$

which also converges to 0 due to the $n^{1/2}$-consitency of $\hat{\mathscr{R}}_{OO}$. This completes the proof for point reconstructions. The fact that $\mathbb{E}\{\|\hat{X}_M^{(\alpha_n)} - \tilde{X}_M\|^2\} \to 0$ follows from the dominated convergence theorem. $\qquad\square$

The theorem shows that even if the optimal solution is not a regression (i.e., Hilbert–Schmidt integral) operator, one can proceed as if it was, using ridge regularization, and asymptotically achieve optimality. The method of Kneip and Liebl (2020) is also of this kind: they use principal component regression (with some modifications) to construct a sequence of regression operators that asymptotically performs optimally. A similar phenomenon was observed in Kraus and Stefanucci (2019) for linear classification of functional data. The common property of all these situations is that the goal is prediction, that is, one is interested in the value of a linear functional or operator at a random process rather than in estimating the (possibly non-existent) regression functional or operator itself.

In addition to the revealing discussion of optimal reconstruction in RKHS terms and development and analysis of the principal component method, Kneip and Liebl (2020) proposed certain adjustments beyond the direct empirical plug-in approach to mainly ensure continuity of the reconstructed trajectory. First, they use local linear kernel estimators of the mean and covariance function rather than the empirical moment estimators (3), (4). Second, they vertically shift the reconstructed missing parts of curves to achieve continuity at the endpoints of observation intervals. Third, they consider applying principal component regularization also to the observed parts of curves to improve their estimation. All these modifications can be used in connection with ridge regularization as well.

3 Numerical illustration

Our simulation study illustrates the asymptotic optimality of the ridge reconstruction method in situations that are covered by the present results but not by the results in Kraus (2015). Under two scenarios the optimal reconstruction operator is a bounded linear operator but not a Hilbert–Schmidt integral operator (hence point reconstructions are not bounded linear

functionals), and under another scenario the optimal reconstruction operator is not even bounded.

We consider functional data on $\mathcal{I} = [0, 1]$ evaluated on a grid of 500 equidistant points. For different training sample sizes n (e.g., $n = 50, 100, 200, 400$) we repeat $10\,000$ times the following steps. We generate a function X to be reconstructed on the missing set $M = (0.5, 1]$. We generate a set of n independent random trajectories X_i and independent observation sets O_i. We estimate the mean and covariance function using the fragments X_{iO_i} and build the ridge reconstruction operator (using generalized cross-validation as described in Kraus, 2015) for predicting functions on M from their observations on the complement. Then we compute the prediction of X_M. Finally we combine the results of the $10\,000$ replications to estimate the mean squared reconstruction error at each $u \in M$ by the average of the squared differences of the reconstructed value and the true value.

The observation sets O_i for $20\,\%$ of the training data are equal to $[0, 1]$, i.e., these curves are observed completely. For the remaining curves we independently generate U_{i1} and U_{i2} from the uniform distribution on $[0, 1]$ and denote by A_i and B_i their minimum and maximum, respectively. Then we set with equal probability either $O_i = [A_i, B_i]$ or $O_i = [0, 1] \setminus (A_i, B_i)$. Among the incomplete curves, the percentage of complete pairs of observations at pairs of time points ranges from $25\,\%$ for $|s - t| = 0.5$ to $50\,\%$ for $s = t$.

We consider three models for training and test random functions. Let $U(t)$ be the Ornstein–Uhlenbeck process with mean 0 and covariance $\exp(-|s-t|)$ and $W(t)$ the Brownian motion with covariance $\min(s, t)$ independent of $U(t)$.

Case 1 Set $X(t) = U(t) + U(1 - t)$. Since the process is symmetric about 0.5, the optimal reconstruction of X_M is given by $\mathscr{A}(X_O)(u) = X_O(1 - u)$, $u \in M$. The reconstruction operator $\mathscr{A}$ is bounded but not Hilbert–Schmidt and point reconstructions $\mathscr{A}(X_O)(u)$, being evaluation functionals, are not bounded linear functionals. The optimal reconstruction error is 0 (perfect reconstruction) for all $u \in M$.

Case 2 Set $X(t) = U(t) + U(1-t) + W(t-0.5)1_{\{t \geq 0.5\}}$. This is a symmetric process about 0.5 with additional independent noise on $[0.5, 1]$, hence the optimal reconstruction of X_M is again given by $\mathscr{A}(X_O)(u) = X_O(1-u)$, $u \in M$. The reconstruction operator is bounded but not Hilbert–Schmidt and point reconstruction functionals are not bounded. The optimal reconstruction error is $\operatorname{var} W(u - 0.5) = u - 0.5$ for $u \in M$.

Case 3 Set $X(t) = W(t)$. In this case the best reconstruction is given by $\mathscr{A}(X_O)(u) = X_O(0.5)$, $u \in M$. The reconstruction functionals are the evaluation functional at 0.5, hence they are unbounded. Unlike in the previous cases, the reconstruction operator is not even bounded, since the whole reconstructed trajectory depends on one impact point in the observation set. The error of the optimal reconstruction is $\operatorname{var}\{W(u) - W(0.5)\} = u - 0.5$, $u \in M$.

None of these cases is covered by the theory in Kraus (2015) because the optimal reconstruction operator is not a Hilbert–Schmidt operator. Case 3 was considered by Kneip and Liebl (2020) as an example of situation, where the optimal reconstruction operator is not a regression operator. Despite the same optimal reconstruction error in Cases 2 and 3 the properties of the optimal reconstruction operators are different (the former is bounded, the latter is not).
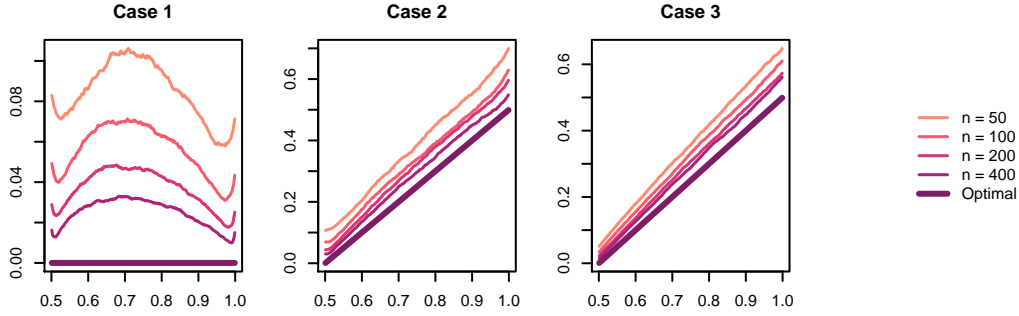
Figure 1: Mean squared reconstruction error for different sample sizes and the optimal mean squared reconstruction error

The results of the simulation are plotted in Figure 1. It confirms the theoretical findings of this paper that the error of the ridge reconstruction method converges to the optimal error as the sample size grows. This empirical study further illustrates the remarkable theoretical fact that a sequence of Hilbert–Schmidt operators asymptotically achieves the performance of the optimal operator that may not be Hilbert–Schmidt or even bounded.

Acknowledgements

References

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer.

Bugni, F. A. (2012). Specification test for missing functional data. *Econometric Theory*, 28(5):959–1002.

Dawson, M. and Müller, H.-G. (2018). Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data. *Journal of the American Statistical Association*, 113(524):1612–1624.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *Journal of the American Statististical Association*, 108(504):1269–1283.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika*, 103(4):779–799.

Descary, M.-H. and Panaretos, V. M. (2019). Recovering covariance from functional fragments. *Biometrika*, 106(1):145–160.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2014). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association*, 109(508):1425–1439.

Goldberg, Y., Ritov, Y., and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference*, 147:53–65.

Gromenko, O., Kokoszka, P., and Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*, 11(2):898–918.

Kneip, A. and Liebl, D. (2020). On the optimal reconstruction of partially observed functional data. *The Annals of Statistics*. To appear.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 77(4):777–801.

Kraus, D. (2019). Inferential procedures for partially observed functional data. *Journal of Multivariate Analysis*, 173:583–603.

Kraus, D. and Stefanucci, M. (2019). Classification of functional fragments by regularized linear classifiers with domain selection. *Biometrika*, 106(1):161–180.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Annals of Applied Statistics*, 7(3):1562–1592.

Liebl, D. and Rameseder, S. (2019). Partially observed functional data: The case of systematically missing parts. *Computational Statistics & Data Analysis*, 131:104 – 115.

Mojirsheibani, M. and Shaw, C. (2018). Classification with incomplete functional covariates. *Statistics & Probability Letters*, 139:40–46.

Shin, H. and Hsing, T. (2012). Linear prediction in functional data analysis. *Stochastic Processes and their Applications*, 122(11):3680–3700.

Stefanucci, M., Sangalli, L. M., and Brutti, P. (2018). PCA-based discrimination of partially observed functional data, with an application to AneuRisk65 data set. *Statistica Neerlandica*, 72(3):246–264.