Statistics and Probability Letters 240 (2027) 110946

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A covariance greedy algorithm for scalar-on-function regression

Wei Qu [a], Hongyan Shen [a], Tao Qian [b,*]

[a] Department of Applied Mathematics, College of Sciences, China Jiliang University, Hangzhou, China
[b] Macau Center for Mathematical Sciences, School of Computer Science and Engineering, Macau University of Science and Technology, Macao Special Administrative Region of China

ARTICLE INFO

MSC:
62R10
41A45
62J05

Keywords:
Scalar-on-function regression
Greedy algorithm
Covariance selection
Supervised approximation

ABSTRACT

We introduce a covariance greedy algorithm (CGA) for scalar-on-function regression. The method builds a low-dimensional regression representation from a prescribed parameterized dictionary by selecting atoms according to their covariance with the current response residual. Unlike energy-based greedy methods, which target the marginal variation of the functional predictor, CGA uses a supervised selection criterion and can be viewed as a dictionary-constrained form of covariance-based dimension reduction. Each selected atom defines a normalized scalar feature and an explicit update of the slope function. We establish a residual-energy decomposition showing that the population residual energy decreases monotonically at each greedy step. Numerical examples indicate that CGA can yield parsimonious and interpretable representations when the response-relevant structure is localized or oscillatory.

1. Introduction

Scalar-on-function regression is a fundamental model in functional data analysis (Ramsay and Silverman, 2005; Morris and regression, 2015; Reiss et al., 2017), in which a scalar response is modeled using a functional predictor. Kernel-based methods for functional data have also been investigated; see, for example, Li and Wu (2022). Let y^ω be a centered real-valued random variable, and let $f^\omega(t) = f(\omega, t)$ be a centered random function defined on a domain T such that

$$\mathbb{E}\|f^\omega\|^2_{L^2(T)} < \infty.$$

A standard functional linear model seeks $\beta \in L^2(T)$ such that

$$y^\omega = \langle \beta, f^\omega \rangle_{L^2(T)} + \epsilon^\omega. \tag{1}$$

Equivalently, β may be characterized as a minimizer of the population risk

$$\beta = \underset{\widetilde{\beta} \in L^2(T)}{\arg\min}\, \mathbb{E}\left| y^\omega - \left\langle \widetilde{\beta}, f^\omega \right\rangle_{L^2(T)} \right|^2. \tag{2}$$

Dimension reduction is central to functional regression. Classical methods, including functional principal component regression and functional partial least squares, construct low-dimensional representations using variance or covariance-based criteria (Morris and regression, 2015; Preda and Saporta, 2005; Wang et al., 2016). However, when the response-relevant structure is localized, oscillatory, or nonstationary, global basis systems may require many components. This motivates the use of adaptive dictionary-based representations.

* Corresponding author.
E-mail addresses: quwei2math@qq.com (W. Qu), Shenhy_math@qq.com (H. Shen), tqian@must.edu.mo (T. Qian).

https://doi.org/10.1016/j.spl.2026.110946
Received 11 May 2026; Received in revised form 26 July 2026; Accepted 16 August 2026
Available online 22 August 2026
0167-7152/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Greedy approximation is a natural tool for constructing sparse expansions from a prescribed dictionary (Temlyakov, 2011). For regression, however, it is desirable to select dictionary elements according to their relevance to the scalar response, rather than only according to the marginal energy of the functional predictor. This leads to the covariance greedy algorithm studied in this note.

CGA is related to FPLS, classical greedy approximation, and component-wise boosting (Bühlmann and Hothorn, 2007), but differs from these methods in how it constructs candidate directions. Classical greedy methods select atoms directly in the target Hilbert space, whereas CGA associates each functional atom with a standardized scalar projection feature across the sample and selects the atom whose feature has the largest absolute covariance with the current response residual. Unlike the shrinkage updates commonly used in component-wise boosting, CGA uses the exact one-dimensional least-squares coefficient at each step. Unlike FPLS, it restricts the search to a prescribed parameterized functional dictionary. Each selected atom therefore yields an explicit slope-function increment, resulting in a sparse and interpretable representation of the slope.

The main contributions are as follows:

(i) A dictionary-constrained residual-covariance selection rule is formulated for scalar-on-function regression.
(ii) An explicit greedy update of the slope function is derived from normalized scalar projection features.
(iii) A residual energy decomposition is established, implying that the population residual energy decreases monotonically over the greedy iterations.

2. Covariance greedy algorithm

Let H be a real separable Hilbert space of functions on T that is continuously embedded in $L^2(T)$. Let

$$\mathcal{G} = \{g_\lambda : \lambda \in \Lambda\} \subset H$$

be a prescribed dictionary of candidate functions. Throughout this section, we assume that each dictionary element is normalized in $L^2(T)$, that is,

$$\|g_\lambda\|_{L^2(T)} = 1, \qquad \lambda \in \Lambda.$$

The dictionary may be finite or infinite. For example, it may consist of spline functions, wavelet functions, trigonometric atoms, kernel atoms, or other problem-specific atoms.

For each $g \in \mathcal{G}$, define the random projection

$$Z_g^\omega = \langle f^\omega, g\rangle_{L^2(T)}. \tag{3}$$

Since the functional predictor is centered, $\mathbb{E}f^\omega = 0$, and hence $\mathbb{E}Z_g^\omega = 0$.

Let

$$\sigma^2(g) = \mathrm{Var}(Z_g^\omega) = \mathbb{E}\left|\langle f^\omega, g\rangle_{L^2(T)}\right|^2. \tag{4}$$

Whenever $\sigma(g) > 0$, define the normalized scalar feature

$$X_g^\omega = \frac{Z_g^\omega}{\sigma(g)} = \frac{\langle f^\omega, g\rangle_{L^2(T)}}{\sigma(g)}. \tag{5}$$

Then $\mathbb{E}X_g^\omega = 0, \mathbb{E}|X_g^\omega|^2 = 1$.

We now construct the greedy regression procedure. Set the initial response residual to

$$y_1^\omega = y^\omega. \tag{6}$$

At the kth step, suppose that the current response residual y_k^ω has been obtained. For each candidate $g \in \mathcal{G}$, define the covariance score

$$C_k(g) = \mathrm{cov}(y_k^\omega, X_g^\omega) = \mathbb{E}\left[y_k^\omega X_g^\omega\right] = \langle y_k^\omega, X_g^\omega\rangle_{L^2(\Omega)}. \tag{7}$$

The next dictionary element is selected according to the maximal absolute covariance criterion:

$$g_k \in \arg\max_{g\in\mathcal{G}} |C_k(g)|. \tag{8}$$

This selection rule is supervised because the choice of g_k depends on the response residual y_k^ω, not merely on the marginal energy of the functional predictor.

Once g_k is selected, define the kth slope update by

$$\beta_k(t) = C_k(g_k)\frac{g_k(t)}{\sigma(g_k)}. \tag{9}$$

Then

$$\langle \beta_k, f^\omega\rangle_{L^2(T)} = C_k(g_k)X_{g_k}^\omega. \tag{10}$$

Thus, the update in (9) corresponds to the least-squares regression of the current response residual y_k^ω on the normalized feature $X_{g_k}^\omega$.

The response residual is updated by

$$y_{k+1}^\omega = y_k^\omega - \langle \beta_k, f^\omega \rangle_{L^2(T)} = y_k^\omega - \langle y_k^\omega, X_{g_k}^\omega \rangle_{L^2(\Omega)} X_{g_k}^\omega. \tag{11}$$

The projection theorem in a Hilbert space immediately implies that y_{k+1}^ω is orthogonal to $\langle y_k^\omega, X_{g_k}^\omega \rangle_{L^2(\Omega)} X_{g_k}^\omega$, and hence also to $\langle \beta_k, f^\omega \rangle_{L^2(T)}$. We therefore obtain the following orthogonal decomposition.

Proposition 1. *The one-step update satisfies*

$$y_k^\omega = \langle \beta_k, f^\omega \rangle_{L^2(T)} + y_{k+1}^\omega \tag{12}$$

with

$$\mathbb{E}|y_k^\omega|^2 = \mathbb{E}\left|\langle \beta_k, f^\omega \rangle_{L^2(T)}\right|^2 + \mathbb{E}|y_{k+1}^\omega|^2. \tag{13}$$

After m steps, the resulting m-term slope approximation is

$$\beta^{(m)}(t) := \sum_{k=1}^m \beta_k(t) = \sum_{k=1}^m C_k(g_k) \frac{g_k(t)}{\sigma(g_k)}. \tag{14}$$

Consequently,

$$y^\omega = \left\langle \beta^{(m)}, f^\omega \right\rangle_{L^2(T)} + y_{m+1}^\omega. \tag{15}$$

Remark 1. The proposed selection rule is different from an unsupervised greedy approximation of f^ω. An unsupervised rule would select dictionary elements that explain the energy of the functional predictor. In contrast, (8) selects dictionary elements that are most correlated with the current response residual. Therefore, the selected atoms are directly adapted to the prediction task.

3. Residual energy decomposition

We next establish the basic residual energy identity of the proposed method. The identity is a direct consequence of the one-dimensional least-squares update in each greedy step.

Theorem 1. *For every $m \geq 1$, the greedy updates yield the stepwise orthogonal residual decomposition*

$$y^\omega = \sum_{k=1}^m C_k(g_k) X_{g_k}^\omega + y_{m+1}^\omega. \tag{16}$$

Moreover,

$$\mathbb{E}|y^\omega|^2 = \sum_{k=1}^m \mathbb{E}\left|\langle \beta_k, f^\omega \rangle_{L^2(T)}\right|^2 + \mathbb{E}|y_{m+1}^\omega|^2. \tag{17}$$

In particular, the population response-residual energy

$$\mathbb{E}|y_{m+1}^\omega|^2$$

is nonincreasing in m. Moreover, there exists $r^\omega \in L^2(\Omega)$ such that

$$r^\omega = \lim_{m\to\infty} y_m^\omega, \qquad \text{in } L^2(\Omega)$$

with $\mathbb{E}|r^\omega|^2 = \inf\{\mathbb{E}|y_m^\omega|^2\}$. Moreover, the response y^ω belongs to the closure of the span $\{\langle f^\omega, g\rangle : g \in \mathcal{G}\}$ in $L^2(\Omega)$ if and only if $\mathbb{E}|r^\omega|^2 = 0$.

Proof. The relations (7), (10) and (8) assert that the process given by (12) and (13) is just the classical greedy algorithm with $\{\langle f^\omega, g\rangle : g \in \mathcal{G}\}$ as a subset of $L^2(\Omega)$. By standard convergence results for greedy algorithms, the process converges to

$$\sum_{k=1}^\infty C_k(g_k) X_{g_k}^\omega = y_1^\omega - r^\omega,$$

where $r^\omega = 0$, a.s. if and only if y^ω belongs to the closure of the span of $\{\langle f^\omega, g\rangle : g \in \mathcal{G}\}$ in $L^2(\Omega)$. See Temlyakov (2011). □

Remark 2. Theorem 1 concerns the population residual energy generated by the greedy regression updates. It should not be interpreted as a monotonicity result for the test prediction error, which also depends on sample size, estimation error, and model selection.

Algorithm 1 Covariance Greedy Algorithm for scalar-on-function regression

Require: Centered samples $\{(y_i, f_i)\}_{i=1}^n$, dictionary $\mathcal{G}$, number of terms m.

Set $y_{1,i} = y_i$, $i = 1, \dots, n$.

for $k = 1, \dots, m$ **do**

For each $g \in \mathcal{G}$, compute

$$Z_{g,i} = \langle f_i, g \rangle_{L^2(T)}.$$

Estimate the feature standard deviation by

$$\widehat{\sigma}^2(g) = \frac{1}{n} \sum_{i=1}^{n} Z_{g,i}^2.$$

Form the normalized feature

$$X_{g,i} = \frac{Z_{g,i}}{\widehat{\sigma}(g)}.$$

Compute the empirical covariance score

$$\widehat{C}_k(g) = \frac{1}{n} \sum_{i=1}^{n} y_{k,i} X_{g,i}.$$

Select

$$g_k \in \operatorname*{argmax}_{g \in \mathcal{G}} |\widehat{C}_k(g)|.$$

Define the slope update

$$\widehat{\beta}_k(t) = \widehat{C}_k(g_k) \frac{g_k(t)}{\widehat{\sigma}(g_k)}.$$

Update the response residuals

$$y_{k+1,i} = y_{k,i} - \left\langle \widehat{\beta}_k, f_i \right\rangle_{L^2(T)}.$$

end for

return $\widehat{\beta}^{(m)} = \sum_{k=1}^{m} \widehat{\beta}_k$.

Remark 3. The finite expansion

$$\beta^{(m)}(t) = \sum_{k=1}^{m} C_k(g_k) \frac{g_k(t)}{\sigma(g_k)}$$

gives a sparse slope approximation. In practice, the number of terms m can be selected by cross-validation, an information criterion, or a prescribed residual-energy reduction threshold.

4. Numerical illustrations

For the quantitative comparisons in Examples 1 and 2, we use $n_{\text{train}} = 800$ training samples and $n_{\text{test}} = 200$ test samples, with results reported as means and standard deviations over 200 independent repetitions. The functions are evaluated on an equally spaced grid over $[0, 1]$, and all integrals are approximated by the trapezoidal rule. Complete dictionary definitions, parameter grids, data-generating mechanisms, and implementation details are provided in the Supplementary Material.

For all quantitative comparisons, prediction accuracy is measured by

$$\text{PE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \left(y_i - \int_0^1 \widehat{\beta}(t) f_i(t)\, dt \right)^2. \tag{18}$$

4.1. Example 1: comparison of different dictionaries

This example examines how the choice of candidate dictionary affects CGA. We consider three types of response-relevant structures: localized oscillatory, asymmetric local, and multiscale smooth. These labels describe the shape of the true slope function rather than the dictionary used by the algorithm.

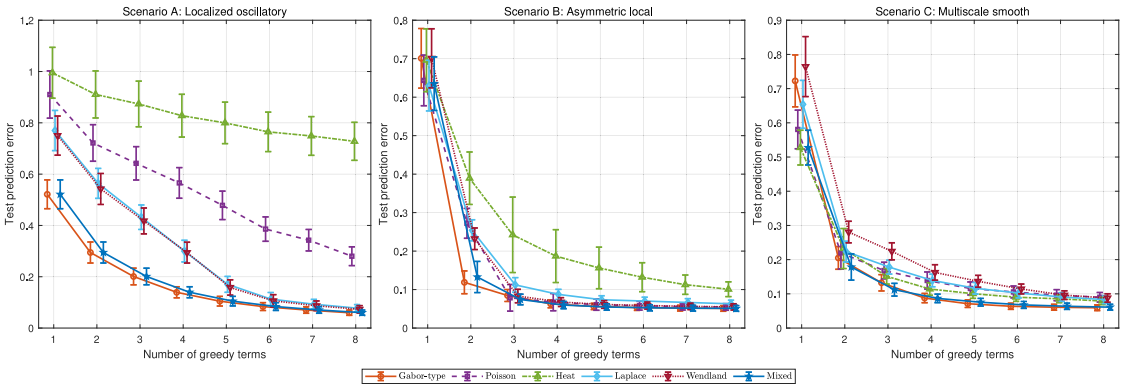


Fig. 1. Mean test prediction errors for six candidate dictionaries over 200 repetitions. Error bars represent one standard deviation.

For each scenario, the data are generated from

$$y_i = \int_0^1 \beta_0(t) f_i(t)\, dt + \eta_i, \qquad i = 1, \dots, n, \tag{19}$$

where η_i are independent Gaussian errors. The functional predictors contain response-relevant components and additional nuisance components.

We apply CGA using the following dictionaries:

$$\mathcal{G}_{\mathrm{Gab}}, \quad \mathcal{G}_{\mathrm{P}}, \quad \mathcal{G}_{\mathrm{H}}, \quad \mathcal{G}_{\mathrm{L}}, \quad \mathcal{G}_{\mathrm{W}}, \quad \mathcal{G}_{\mathrm{mix}}. \tag{20}$$

Here $\mathcal{G}_{\mathrm{Gab}}$, $\mathcal{G}_{\mathrm{P}}$, $\mathcal{G}_{\mathrm{H}}$, $\mathcal{G}_{\mathrm{L}}$ and $\mathcal{G}_{\mathrm{W}}$ denote the Gabor-type, Poisson, heat, Laplace and Wendland dictionaries, respectively. The mixed dictionary is defined by

$$\mathcal{G}_{\mathrm{mix}} = \mathcal{G}_{\mathrm{Gab}} \cup \mathcal{G}_{\mathrm{P}} \cup \mathcal{G}_{\mathrm{H}} \cup \mathcal{G}_{\mathrm{L}} \cup \mathcal{G}_{\mathrm{W}}. \tag{21}$$

Thus, only the candidate dictionary is changed across the experiments.

In addition to the prediction error in (18), the slope recovery error is measured by

$$\left\| \widehat{\beta}^{(K)} - \beta_0 \right\|_{L^2}. \tag{22}$$

The number of greedy terms ranges from $K = 1$ to $K = 8$, while the slope recovery error is reported at $K = 5$.

Fig. 1 shows the mean test prediction error with one-standard-deviation error bars. The results show that the dictionary choice has a clear effect on the performance of CGA. In particular, the Gabor-type dictionary is competitive in the localized oscillatory setting, while the relative performance of the other dictionaries varies across scenarios.

Fig. 2 reports the mean L^2 slope recovery error at $K = 5$. The results further confirm that the accuracy of recovering β_0 depends on the compatibility between the dictionary and the slope structure. The mixed dictionary is therefore not claimed to be uniformly optimal, but is used as a general-purpose choice when the appropriate dictionary is unknown.

4.2. Example 2: localized oscillatory regression under dictionary match and mismatch

This example compares CGA-Gabor and CGA-Mixed with FPLS, functional ridge regression, and Fourier regression in two settings: a localized oscillatory slope and a deliberately dictionary-mismatched slope.

The data are generated from

$$y_i = \int_0^1 \beta_0(t) f_i(t)\, dt + \eta_i, \qquad i = 1, \dots, n, \tag{23}$$

where $\eta_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\eta^2)$, independently of the predictors, and σ_η is set to 0.25 times the empirical signal standard deviation. The predictors combine global Fourier components, smooth Gaussian perturbations, and two local random components.

In the localized oscillatory setting, the slope function is chosen as

$$\begin{aligned} \beta_0(t) = {} & \exp\{-80(t-0.35)^2\}\cos(12\pi t) \\ & + 0.5\exp\{-60(t-0.70)^2\}\sin(16\pi t). \end{aligned} \tag{24}$$

The resulting function is centered and normalized to have unit L^2 norm.

To examine the effect of dictionary mismatch, we also replace β_0 in (24) with

$$\begin{aligned} \widetilde{\beta}_{\mathrm{mis}}(t) = {} & 0.55\sin(3\pi t) + 0.45\cos(5\pi t) - 0.35\sin(7\pi t) \\ & + 0.30\cos(9\pi t) + 0.40(2t-1) + 0.25(6t^2 - 6t + 1). \end{aligned} \tag{25}$$

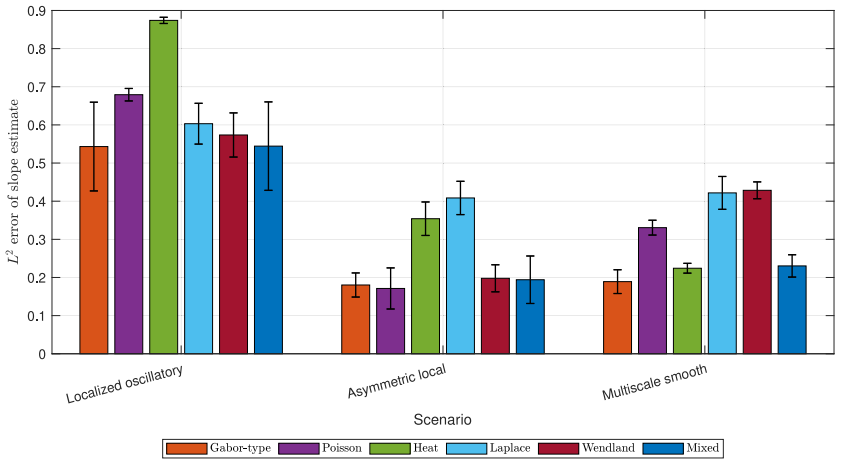
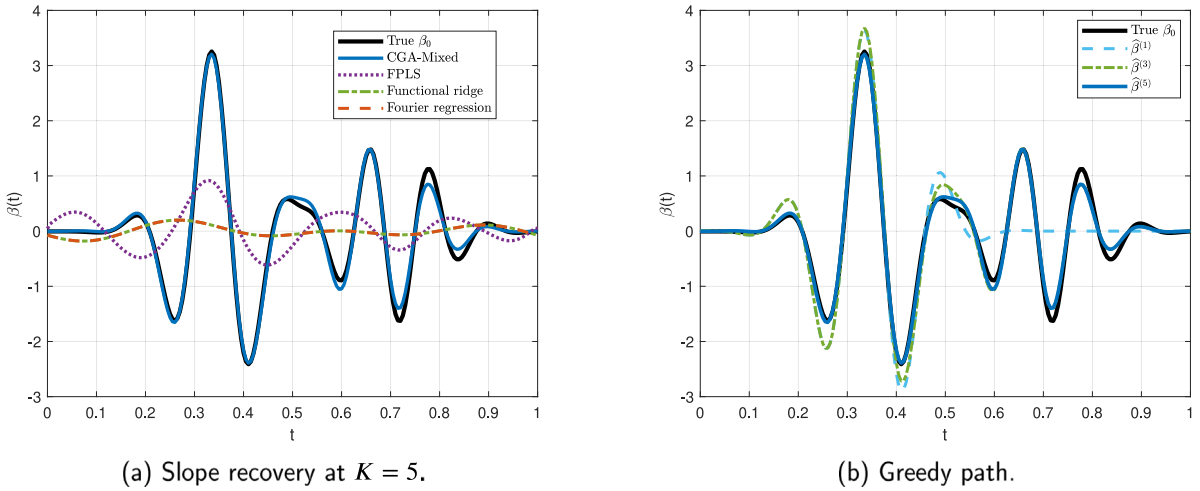


Fig. 2. Mean L^2 slope-recovery errors at $K = 5$ over 200 repetitions, with one-standard-deviation error bars.



(a) Slope recovery at $K = 5$. (b) Greedy path.

Fig. 3. Illustrative results from one prespecified repetition under the localized oscillatory setting. Panel (a) compares the slope estimates at $K = 5$. Panel (b) shows the cumulative CGA-Mixed estimates after 1, 3, and 5 selected atoms.

This function is also centered and normalized to have unit L^2 norm.

For CGA, we use the Gabor-type dictionary $\mathcal{G}_{\text{Gab}}$ and the mixed dictionary $\mathcal{G}_{\text{mix}}$ defined in (21). The former is structurally compatible with the localized oscillatory slope, whereas the latter serves as a general-purpose candidate class.

Fig. 3 provides a representative illustration under the localized oscillatory setting. In panel (a), the CGA-Mixed estimate closely follows the main localized peaks and oscillations of the true slope, whereas the competing estimates are smoother and capture these local features less accurately in this particular run. Panel (b) shows how the cumulative CGA-Mixed estimate progressively captures the response-relevant structure as additional atoms are selected.

The repeated data-driven results are summarized in Table 1, while the corresponding fixed-complexity results are reported in Supplementary Table S2.

For a fair data-driven comparison, we tune the method-specific representation size using a common threshold for training-sample explained response variation (ERV). Here, K denotes the method-specific representation size: the number of selected atoms for CGA, the number of latent components for FPLS, and the number of Fourier basis functions for Fourier and functional ridge regression. Each method selects the smallest $K \in \{1, \dots, 50\}$ satisfying

$$\text{ERV}(K) = 1 - \frac{\|y_c - \hat{y}_c^{(K)}\|_2^2}{\|y_c\|_2^2} \geq 0.92.$$

Table 1
Mean test prediction errors and median representation sizes obtained using the common training-sample ERV threshold of 0.92. Standard deviations are shown in parentheses, based on 200 independent repetitions.

Method	Localized slope		Mismatched slope	
	Test PE	K	Test PE	K
CGA-Gabor	$\mathbf{2.95 \times 10^{-4}}$ (5.97×10^{-5})	**2**	1.77×10^{-2} (3.03×10^{-3})	4
CGA-Mixed	$\mathbf{2.95 \times 10^{-4}}$ (5.97×10^{-5})	**2**	1.72×10^{-2} (2.64×10^{-3})	**3**
FPLS	3.54×10^{-4} (5.28×10^{-5})	8	$\mathbf{1.68 \times 10^{-2}}$ (2.35×10^{-3})	**3**
Functional ridge	3.49×10^{-4} (4.53×10^{-5})	14	1.83×10^{-2} (2.82×10^{-3})	5
Fourier regression	3.39×10^{-4} (4.41×10^{-5})	14	1.79×10^{-2} (2.89×10^{-3})	5

Functional ridge regression additionally considers

$$\lambda \in \{10^{-5}, 10^{-4}, \dots, 10^{2}\}.$$

Thus, all methods are compared under the same minimum training-sample ERV requirement rather than at a fixed representation size.

As shown in Table 1, under the localized oscillatory setting, CGA-Gabor and CGA-Mixed not only reach the common training-sample ERV threshold with a median of two atoms, compared with eight FPLS components and fourteen basis functions for both functional ridge and Fourier regression, but also achieve the lowest mean test prediction errors. These results indicate the advantage of CGA when the dictionary is well matched to the response-relevant structure, yielding both lower prediction error and a more parsimonious representation. Under dictionary mismatch, however, FPLS achieves the lowest mean prediction error, indicating that the performance gain of CGA depends on dictionary compatibility.

4.3. Example 3: mechanism visualization of the mixed-greedy algorithm

This example uses representative runs to illustrate the internal behavior of mixed-dictionary CGA, rather than providing another repeated quantitative comparison. We focus on the decay of the response residual and the types of atoms selected during the greedy iterations.

We consider two slope structures. Case A is a localized oscillatory slope, while Case B is a non-oscillatory local slope. For Case A, the slope function is generated as

$$\begin{aligned}\beta_0^A(t) = c_A \Big[& \exp\{-120(t-0.30)^2\}\cos(12\pi t) \\ & - 0.80\exp\{-100(t-0.62)^2\}\sin(16\pi t) \\ & + 0.55\exp\{-160(t-0.78)^2\}\cos(10\pi t) \Big],\end{aligned} \tag{26}$$

where c_A is a normalization constant. For Case B, the slope function is generated as

$$\beta_0^B(t) = c_B\{g_{\mathrm{P},1}(t) - 0.75 g_{\mathrm{H},1}(t) + 0.60 g_{\mathrm{L},1}(t) - 0.45 g_{\mathrm{W},1}(t)\}, \tag{27}$$

where $g_{\mathrm{P},1}$, $g_{\mathrm{H},1}$, $g_{\mathrm{L},1}$ and $g_{\mathrm{W},1}$ denote representative Poisson, heat, Laplace and Wendland local atoms, respectively, and c_B is a normalization constant. The data are generated from

$$y_i = \int_0^1 \beta_0(t) f_i(t)\, dt + \eta_i, \qquad i = 1, \dots, n. \tag{28}$$

The predictors contain response-relevant components and additional nuisance components.

We apply CGA using the mixed dictionary $\mathcal{G}_{\mathrm{mix}}$ defined in (21), with atoms selected at each step according to their empirical covariance with the current response residual.

Fig. 4 summarizes the results. Column (a) shows the normalized residual energy

$$\frac{\|y_k\|_n^2}{\|y_1\|_n^2}, \qquad \|y_k\|_n^2 = \frac{1}{n}\sum_{i=1}^{n} y_{k,i}^2. \tag{29}$$

In both cases, the residual energy decreases during the greedy iterations, which is consistent with the residual energy decomposition. Column (b) reports the absolute covariance scores of the selected atoms, with colors indicating the corresponding sub-dictionaries. For Case A, most selected atoms are from the Gabor-type dictionary, while for Case B, atoms from other local dictionaries also appear in the early steps. This suggests that the selected atoms reflect, at least partially, the response-relevant structure of the slope function. Column (c) shows the predicted versus observed responses on the test set at $K = 6$.

Overall, this example provides a mechanism-level illustration of the proposed method. The results suggest that the covariance-driven updates reduce the response residual and that the selected atoms provide interpretable information about the structures exploited by the greedy approximation.

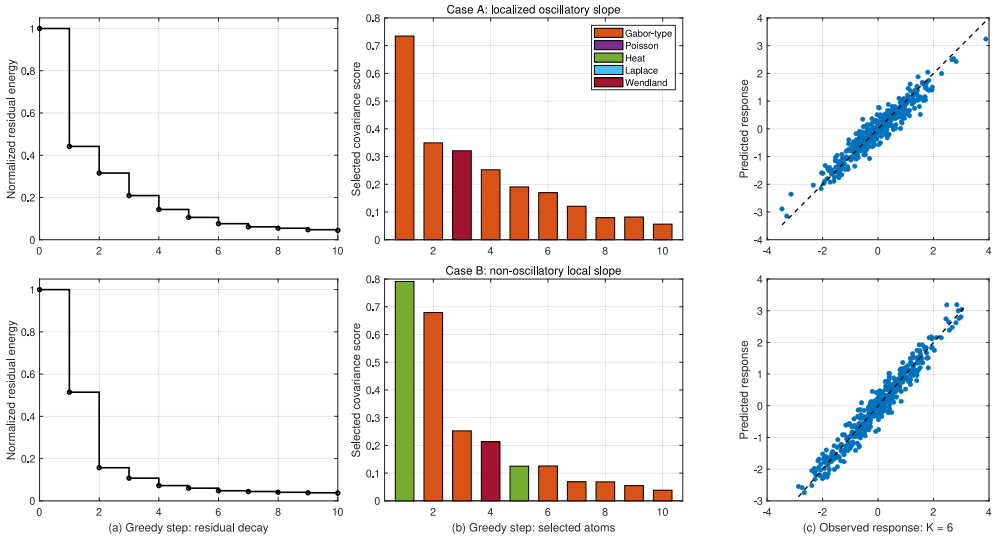


Fig. 4. Mechanism visualization of the proposed mixed-dictionary covariance-greedy algorithm. The upper row corresponds to Case A, the localized oscillatory slope, and the lower row corresponds to Case B, the non-oscillatory local slope. Column (a) shows the normalized residual energy. Column (b) shows the selected covariance scores, with colors indicating the selected atom types. Column (c) shows the predicted versus observed responses on the test set at $K = 6$.

5. Conclusion

We introduced CGA for scalar-on-function regression, which selects dictionary atoms according to their covariance with the current response residual and yields explicit slope updates. The residual-energy decomposition implies monotone decay of the population residual energy. Numerical examples demonstrate the potential of CGA for localized or oscillatory response-relevant structures, particularly when the dictionary is well matched to the underlying slope. Extensions to data-adaptive dictionaries and regularized implementations will be considered in future work.

Acknowledgments

This work was supported by the Prevention and Control of Emerging and Major Infectious Diseases-National Science and Technology Major Project (Grant No. 2027ZD01999504), the Major Project of Guangzhou National Laboratory (Grant No. GZNL2024A01004), the Science and Technology Development Fund of Macau SAR (Grant No. 0015/2025/AFJ), and the National Natural Science Foundation of China (Grant No. 12401127).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2026.110946.

Data availability

No data was used for the research described in the article.

References

Bühlmann, P., Hothorn, T., 2007. Boosting algorithms: Regularization, prediction and model fitting. Statist. Sci. 22, 477–505.
Li, X.Y., Wu, B.Y., 2022. A kernel regression approach for identification of first order differential equations based on functional data. Appl. Math. Lett. 127, 107832.
Morris, J.S., regression, Functional, 2015. Annu. Rev. Stat. Appl. 2, 321–359.
Preda, C., Saporta, G., 2005. Clusterwise PLS regression on a stochastic process. Comput. Statist. Data Anal. 49, 99–108.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Reiss, P.T., Goldsmith, J., Shang, H.L., Ogden, R.T., 2017. Methods for scalar-on-function regression. Int. Stat. Rev. 85, 228–249.
Temlyakov, V.N., 2011. Greedy Approximation. Cambridge University Press, Cambridge.
Wang, J.-L., Chiou, J.-M., Müller, H.-G., 2016. Functional data analysis. Annu. Rev. Stat. Appl. 3, 257–295.