Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction From Noisy Periodic Signals

Li, Yongxiang; Zhang, Yunji; Wu, Jianguo; Xie, Min

Published in:
IEEE Transactions on Automation Science and Engineering

Published: 01/01/2025

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Published version (DOI):
10.1109/TASE.2024.3387833

Publication details:
Li, Y., Zhang, Y., Wu, J., & Xie, M. (2025). Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction From Noisy Periodic Signals. *IEEE Transactions on Automation Science and Engineering*, *22*, 3011-3020. https://doi.org/10.1109/TASE.2024.3387833

Li, Y., Zhang, Y., Wu, J., & Xie, M. (2024). Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction From Noisy Periodic Signals. *IEEE Transactions on Automation Science and Engineering*. Advance online publication. https://doi.org/10.1109/TASE.2024.3387833

IEEE Transactions on Automation Science and Engineering

Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction from Noisy Periodic Signals

Journal:	*IEEE Transactions on Automation Science and Engineering*
Manuscript ID	T-ASE-2023-236
Manuscript Categories:	Regular Paper
Date Submitted by the Author:	10-Feb-2023
Complete List of Authors:	Li, Yongxiang; Shanghai Jiao Tong University, Industrial Engineering and Management Zhang, Yunji; Shanghai Jiao Tong University, Department of Industrial Engineering and Management Wu, Jianguo; Peking University, Industrial Engineering and Management Xie, Min; City University of Hong Kong, SEEM
Key Words:	Periodic Gaussian Process, Sparse Feature Extraction, mixed effects model, Lasso, Circulant matrix

There May Be Code or a Dataset Associated with This Manuscript

During the submission process, authors are asked if their article has Code in Code Ocean or a dataset at IEEE DataPort. We provide their responses below.

Do you have CODE (software) associated with your article?

No

Do you have DATA associated with your article?

No

What is the IEEE DataPort DOI for your data?

CUST_DATA_DOI :No data available.

What is the title of your data as entered into IEEE DataPort?

CUST_DATA_TITLE :No data available.

Note: If the authors do not have Code or DATA, the responses will say “No data available.”

Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction from Noisy Periodic Signals

Yongxiang Li, *Member, IEEE,* Yunji Zhang, Jianguo Wu*, *Member, IEEE,* and Min Xie, *Fellow, IEEE*

***Abstract*—This study proposes a nonparametric sparse feature extraction approach based on a Periodic Gaussian process (PGP) for highly nonlinear sparse periodic signals, which may not be effectively modeled by conventional linear models based on some user-specified dictionaries. The PGP model is reformulated as a mixed effects model. Hence a regularization term is allowed to be imposed on the random effect of the PGP model, called regularized PGP (RPGP) in this study, for sparse feature extraction. Unlike conventional sparse models, the proposed RPGP can simultaneously model both fixed and random effects (global trend and local sparsity) of the signals. A computationally scalable algorithm based on the alternating direction method of multipliers (ADMM) is tailored for RPGP to iteratively optimize the fixed and random effects. The efficient computation of RPGP is achieved by utilizing the block matrix decomposition of the periodic covariance matrix, fast Fourier transform (FFT) of circulant matrices, and Cholesky factorization of Toeplitz matrices. The performance of RPGP is evaluated through a simulation study on synthetic signals and a case study on real vibration signals.**

***Note to Practitioners*—This work is motivated by the problem of sparse feature extraction from highly nonlinear periodic signals with a complex global trend. The key issues involved in this problem include: 1) how to model the highly nonlinear periodic signals; 2) how to separate the periodic signals and the global trend from noisy signals; 3) how to extract the sparse periodic feature. Existing approaches for sparse feature extraction assume that the signals have a constant trend and can be effectively modeled by some dictionaries. This work proposes a novel approach that leverages a regularized PGP to simultaneously extract the sparse periodic feature and the global trend. Moreover, the nonparametric nature of the PGP model allows automatic feature extraction, eliminating the need to manually select a dictionary. The proposed approach involves four main steps: 1) collecting the periodic signals from embedded sensors; 2) modeling the signals with PGP models following the fixed effects framework; 3) training the model parameters through the proposed optimization procedure; 4) extracting the global trend and sparse periodic feature based on the model parameters. This paper demonstrated the application of rolling bearing vibration signals, while the effectiveness of the proposed model is not limited to this topic. Other periodic signals with similar characteristics can also be analyzed by applying the proposed model. In the future, this approach will be extended to handle sparse feature extraction from pseudo-periodic signals in a more complex noise environment.**

***Index Terms*—Fault detection, Bearing fault diagnosis, Sparse nonparametric regression, Circulant matrix, Lasso.**

I. Introduction

Periodic signals, such as bearing vibration signals [1], [2] and ECG signals [3], [4], are widely observed in industrial and biomedical fields. Before applying periodic signals for beneficial applications such as fault detection, feature extraction is an essential step in signal processing due to the disturbances of background noises. Due to its ability to extract robust and explainable features, sparse feature extraction is of great significance to many applications. For example, the vibration signals of faulty bearings are commonly sparse transient signals because the transients would be produced in groups once the rolling elements pass over a defect [5], [6]. Thus, sparse feature extraction of the vibration signals is crucial for bearing fault detection.

Recent years have seen the development of sparse representation (SR [7]), a major category of sparse feature extraction models. By solving a regularized least square problem, SR-based models project the dense high-dimensional signals into a sparse low-dimensional feature space. Thus, by choosing a proper dictionary, the features buried in strong noises can be extracted. Several regularization terms, such as L_1 and $L_{1/2}$ penalties [8], [9], have been proposed to promote the sparsity of the extracted feature. Meanwhile, for the choice of the dictionary, quite a few bases have been proposed to construct the sparse domain. For example, linear dictionaries such as B-spline basis functions have been used in [6] to represent the signals in the sparse domain, and constraints on the derivatives of the B-spline have been imposed to guarantee the continuity of the sparse feature [10]. Other nonlinear

The work of Yongxiang Li was supported in part by the National Natural Science Foundation of China under Grant Grant NSFC-72101147 and by Shanghai Pujiang Program under Grant 21PJ1405500. The work of Jianguo Wu was supported in part by the National Natural Science Foundation of China under Grant Grant NSFC-72171003.

Yongxiang Li and Yunji Zhang are with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: yongxili-c@my.cityu.edu.hk).

Jianguo Wu *(Corresponding author)* is with the Department of Industrial Engineering and Management, Peking University, Beijing 100871, China, e-mail: j.wu@pku.edu.cn.

Min Xie is with the Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong, e-mail: minxie@cityu.edu.hk.

parametric dictionaries, such as the tunable Q-factor wavelets [11] and Morlet wavelets [12], have also been proposed to extract the sparse feature.

However, these SR methods may suffer from several difficulties because the periodic signals may have complex and nonlinear structures in many applications. First, conventional SR models based on linear dictionaries such as B-splines [10] may not effectively model highly nonlinear signals. Second, conventional SR models based on nonlinear parametric dictionaries depend heavily on the choice of the dictionary and its tuning parameters, which may not be universally optimal for different applications [13]. Specifying the tuning parameters manually, even with the help of prior knowledge, may still suffer from over-fitting or under-fitting problem because the parameters are not estimated from data. Third, periodic signals often have a global trend and a local feature. Conventional SR methods mainly focus on modeling the sparse periodic feature, and few works have been proposed to simultaneously model the global trend and the local sparsity of highly nonlinear periodic signals.

To fill the gap and address the difficulties, this paper proposes a nonlinear and nonparametric model based on PGP to extract the sparse periodic feature from highly nonlinear periodic signals. The PGP model is reformulated as a mixed effects model to simultaneously extract the global trend and sparse periodic feature, which are treated as the fixed and mixed effects, respectively. This reformulation allows a regularization term to be imposed on PGP models; hence a sparse penalty on the random effect is achieved in RPGP for sparse periodic feature extraction. The fixed and random effects are iteratively optimized through a tailored alternating direction method of multipliers algorithm (ADMM [14]). Block matrix decomposition of the periodic covariance matrix and fast Fourier transform (FFT) of circulant matrices are utilized in RPGP to achieve a computationally scalable complexity. Moreover, an adaptive model selection criterion is proposed to estimate the regularization parameter of RPGP [15], [16].

Unlike conventional linear parametric sparse models based on dictionaries [12], [11], [6], RPGP is a nonlinear and nonparametric model with all parameters estimated from signals, and the dictionary is not needed anymore. Another advantage of RPGP is its capability of modeling both fixed and random effects (global trend and local sparsity) of the periodic signals. Thus, RPGP can effectively extract the sparse periodic feature for highly nonlinear signals with a non-constant trend. It is worth mentioning that some regularization terms have been imposed on GP models for computationally efficient approximation, such as GPLasso [17] and GPHalf [18]. These regularized GPs are not designed for sparse prediction. Moreover, their computation convenience using an approximation approach sacrifices some model accuracy. Unlike these approximate accelerations for GPs [19], [20], the proposed scalable acceleration for RPGP is achieved without any approximation.

The rest parts of this paper are organized as follows. Sec. II briefly reviews the related works and explains the research gaps. The formulation of RPGP is detailed in Sec. III, where an efficient optimization algorithm based on ADMM is applied to optimize the global trend and local sparsity iteratively, and some useful theoretical properties are also derived. A simulation study on synthetic signals is conducted in Sec. IV to evaluate the effectiveness of RPGP. It is also applied to a real case study for bearing fault detection in Sec. V to show its applicability and usefulness. Finally, we conclude this study in Sec. VI.

II. Literature Review

Conventional sparse feature extraction models for noisy signals $y(t) = x(t) + \epsilon(t)$ mainly utilize sparse representation (SR) on the true signal $x(t)$, where $\epsilon(t)$ is the background noise masking the true signal $x(t)$. SR-based models project the dense high-dimensional signals into a sparse low-dimensional feature space by solving a regularized least squares problem: $\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{\gamma}\|_2^2 + \lambda P(\boldsymbol{\gamma})$, where $\boldsymbol{y} = [y(t_1), ..., y(t_n)]^T$ is the noisy signal, $\boldsymbol{D} \in \mathbb{R}^{N \times M}$ is a representative dictionary constructed on time indexes $t_1, \cdots, t_n$, $\boldsymbol{\gamma} \in \mathbb{R}^M$ is a vector of sparse coefficients corresponding to the dictionary, λ is a tuning parameter trading off model sparsity against precision, and $P(\boldsymbol{\gamma})$ is a penalty function to promote sparsity. The L1 regularization term $\|\boldsymbol{\gamma}\|_1$ is widely used for $P(\boldsymbol{\gamma})$ [8], although several variations have been proposed, such as the constraint lasso [10] and group lasso [13]. Several bases, such as B-spline functions [6], [10], tunable Q-factor wavelets [11], and Morlet wavelets [12] have been proposed to construct the dictionary $\boldsymbol{D}$.

However, the performance of SR-based models depends heavily on the choice of the dictionary $\boldsymbol{D}$ because the choice of the basis functions needs to be drawn from prior knowledge, which is commonly hard to acquire in some real applications. Furthermore, the parametric basis may not be able to effectively model the highly nonlinear signals because the tuning parameters (e.g., the number of knots, the order of the spline, the scale and translation parameters of mother wavelets, etc.) require to be specified by users, instead of estimated from the signals.

In addition, conventional SR-based methods [12], [6], [10] may not be able to simultaneously model both the global trend and local sparsity of the signal, i.e., the fixed and random effects shown in Fig. 1 (a) and (b). To the best of our knowledge, most existing methods assume the fixed effect of signals is zero and only model the sparse random effect using $\boldsymbol{D}\boldsymbol{\gamma}$. This assumption may be inadequate for signals in some applications with a global trend, as shown in Fig. 1 (c) and (d).

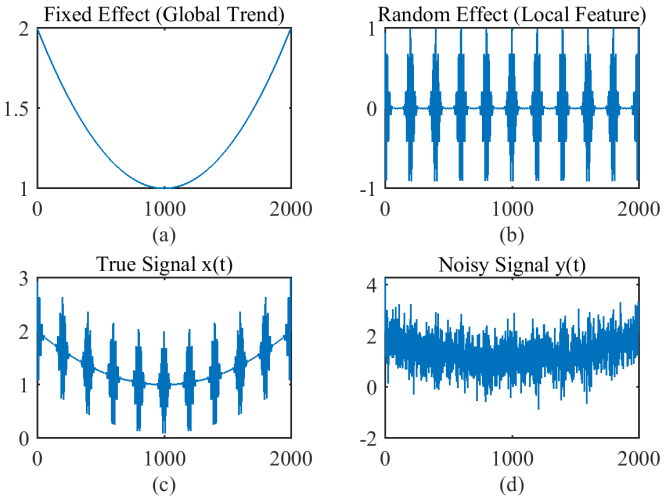

Fig. 1. Components of the noisy signal, (a) Quadratic fixed effect (Global Trend); (b) periodic random effect (sparse local feature); (c) the true signal $x(t)$ consisting of the fixed and random effects; (d) the noisy signal $y(t)$.

To address the challenges faced by the conventional linear parametric models, a GP-based nonparametric model is developed in this paper to extract the sparse feature from highly nonlinear signals with a global trend. GP is a powerful nonparametric model for nonlinear regression and classification. Due to its flexibility, fully-probability, and sound generalization capacity, GP has been a popular tool in statistical modeling and machine learning for the last decade.

Generally, GP models can be formulated as follows:

$$y(t) = \boldsymbol{f}^T(t)\boldsymbol{\beta} + z(t) + \epsilon(t) \tag{1}$$

where $\boldsymbol{f}(t)$ is a regression function, $\boldsymbol{\beta}$ is a $q \times 1$ regression coefficient vector, $z(t)$ is a zero mean GP with a variance σ^2 and a correlation function $\psi_{\boldsymbol{\phi}}(t,t')$, and $\epsilon(t) \sim \mathcal{N}(0, \sigma^2\delta^2)$ is an i.i.d. noise. It is worth noting that $x(t) = \boldsymbol{f}^T(t)\boldsymbol{\beta} + z(t)$ in GP modeling. To model the periodic signals, this study uses the periodic correlation function [21]

$$\psi_{\boldsymbol{\phi}}(t,t') = \exp\left(-\theta^2 \sin^2\left(\frac{\pi(t-t')}{T}\right)\right), \tag{2}$$

where $\boldsymbol{\phi} = \{\theta, T\}$, θ is the roughness parameter (length scale), and T is the period.

Given the data $\boldsymbol{y} = [y(t_1), ..., y(t_n)]^T$, the parameters $\boldsymbol{\beta}$ and σ^2 of the GP model can be estimated through the maximum likelihood estimation (MLE, [22]):

$$\begin{cases} \hat{\boldsymbol{\beta}} &= \left(\boldsymbol{F}^T\boldsymbol{K}_\delta^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\boldsymbol{K}_\delta^{-1}\boldsymbol{y} \\ \hat{\sigma}^2 &= \left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right)^T\boldsymbol{K}_\delta^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right)/n \end{cases}, \tag{3}$$

where $\boldsymbol{F} = [\boldsymbol{f}(t_1), \cdots, \boldsymbol{f}(t_n)]^T$ and $\boldsymbol{K}_\delta = \boldsymbol{K}_{\boldsymbol{\phi}} + \delta^2\boldsymbol{I}_n$. Here, the correlation matrix $\boldsymbol{K}_{\boldsymbol{\phi}}$ is calculated from the signal $\boldsymbol{y}$ according to the correlation function $\psi_{\boldsymbol{\phi}}(t,t)$. Other parameters can be optimized via some searching algorithms, that is

$$\hat{\boldsymbol{\phi}}, \hat{\delta} = \arg\min_{\boldsymbol{\phi},\delta}\left\{n\log\hat{\sigma}^2 + \log|\boldsymbol{K}_\delta|\right\}.$$

After the model parameters have been estimated, the model prediction can be made through BLUP [22]:

$$\hat{y}(t) = \boldsymbol{f}^T(t)\hat{\boldsymbol{\beta}} + \boldsymbol{r}_{\hat{\boldsymbol{\phi}}}^T(t)\boldsymbol{K}_{\hat{\delta}}^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right), \tag{4}$$

where $\boldsymbol{r}_{\hat{\boldsymbol{\phi}}}(t) = \left[\psi_{\hat{\boldsymbol{\phi}}}(t,t_1), \cdots, \psi_{\hat{\boldsymbol{\phi}}}(t,t_n)\right]^T$.

A Circulant PGP (CPGP) modeling has been proposed in [23] to accelerate the computation of MLE and BLUP without any approximation. CPGP is developed based on the assumption of grid (equally time spaced) observations, which is valid in the case when collecting signals at a given sampling frequency. Thus, CPGP will be adopted to estimate model parameters (e.g., $\hat{\theta}$, $\hat{T}$, and $\hat{\delta}$) in this study.

Thus, the extracted feature for $\boldsymbol{y}$ after removing the noises is

$$\hat{\boldsymbol{y}} = \boldsymbol{F}\hat{\boldsymbol{\beta}} + \boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\hat{\delta}}^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right), \tag{5}$$

where $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ is obtained by substituting the parameter $\boldsymbol{\phi}$ of the correlation matrix $\boldsymbol{K}_{\boldsymbol{\phi}}$ with its estimated value $\hat{\boldsymbol{\phi}}$, and $\boldsymbol{K}_{\hat{\delta}} = \boldsymbol{K}_{\hat{\boldsymbol{\phi}}} + \hat{\delta}^2\boldsymbol{I}_n$. However, conventional GP models [24], [22] cannot extract sparse features from the signals because $\hat{\boldsymbol{y}}$ produced by BLUP is not sparse.

Some penalties have been imposed to conventional GP models to reduce the high computational cost, instead of producing sparse predictions. Its cost function has the following form:

$$Q(\boldsymbol{\alpha}) = -\boldsymbol{y}^T\boldsymbol{K}_\delta\boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^T\left(\sigma^2\boldsymbol{K}_\delta + \boldsymbol{K}_\delta^T\boldsymbol{K}_\delta\right)\boldsymbol{\alpha} + \lambda P(\boldsymbol{\alpha}),$$

after GP is written as generalized linear regression, where $\boldsymbol{\alpha}$ is the regression coefficient. $P(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}^*\|_1$ is proposed in [17] and $P(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}^*\|_{1/2}$ is proposed in [18]. Based on the optimized $\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha})$, the corresponding prediction is $\hat{\boldsymbol{y}} = \boldsymbol{K}_\delta\hat{\boldsymbol{\alpha}}$. However, a sparse $\hat{\boldsymbol{\alpha}}$ in these regularized GP models cannot ensure the prediction $\hat{\boldsymbol{y}}$ is sparse. To overcome this issue, the RPGP will be developed in this study.

III. Regularized Periodic Gaussian Process

To fill this gap, we reformulate the PGP model similar to a mixed effects model, where $\boldsymbol{F}\boldsymbol{\beta}$ is the fixed effect with user-specified bases, such as polynomial or B-spline bases, $\boldsymbol{z} = [z(t_1), \cdots, z(t_n)]^T$ is the random effect following a PGP with the kernel function given in (2), and $\boldsymbol{\epsilon} = [\epsilon(t_1), \cdots, \epsilon(t_n)]^T$ is noises. That is,

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{z} + \boldsymbol{\epsilon}, \tag{6}$$

where $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{K})$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\delta^2\boldsymbol{I})$. According to [25], the parameters of the mixed effects model can be derived

by maximizing the joint density of $\boldsymbol{y}$ and $\boldsymbol{\epsilon}$, expressed as

$$\Pr\left(\boldsymbol{y},\boldsymbol{z}\mid\boldsymbol{\beta},\hat{\sigma}^2,\hat{\delta}^2,\hat{T}\right)=\left|\hat{\sigma}^4\hat{\delta}^2\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\right|^{-1/2} \times\exp\left\{-\frac{\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right)^T\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right)}{2\hat{\sigma}^2\hat{\delta}^2}-\frac{\boldsymbol{z}^T\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}^{-1}\boldsymbol{z}}{2\hat{\sigma}^2}\right\}. \tag{7}$$

This new formulation allows writing the random effect $\boldsymbol{z}$ explicitly in the cost function, making the proposed RPGP possible by adding a sparse penalty to the random effect $\boldsymbol{z}$. It is worth noting that $\hat{\sigma}$, $\hat{\delta}$, and $\hat{T}$ are estimated by CPGP before RPGP is applied.

The fixed and random effects can be estimated by maximizing the joint density $\Pr\left(\boldsymbol{\epsilon},\boldsymbol{z}\mid\boldsymbol{\beta},\hat{\sigma}^2,\hat{\delta}^2,\hat{T}\right)$, that is,

$$\hat{\boldsymbol{\beta}},\hat{\boldsymbol{z}}=\arg\min_{\boldsymbol{\beta},\boldsymbol{z}}\frac{\left\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right\|_2^2}{2\hat{\delta}^2}+\frac{\boldsymbol{z}^T\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}^{-1}\boldsymbol{z}}{2}. \tag{8}$$

Note that the matrix $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ is singular; thus, $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}^{-1}$ in Eq. (8) does not exist. To address this singularity issue, we employ the analytic continuation on $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$:

$$\hat{\boldsymbol{\beta}}_\tau,\hat{\boldsymbol{z}}_\tau=\arg\min_{\boldsymbol{\beta},\boldsymbol{z}}\frac{\left\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right\|_2^2}{2\hat{\delta}^2}+\frac{\boldsymbol{z}^T\boldsymbol{K}_\tau^{-1}\boldsymbol{z}}{2}, \tag{9}$$

where $\boldsymbol{K}_\tau=\left(\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}+\tau\boldsymbol{I}\right)$.

By equating the derivatives with regard $\boldsymbol{\beta}$ and $\boldsymbol{z}$ to zero, the parameter estimates of $\boldsymbol{\beta}$ and $\boldsymbol{z}$ by optimizing Eqs. (8) and (9) are $\hat{\boldsymbol{\beta}}_\tau=\hat{\boldsymbol{\beta}}=\left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\left(\boldsymbol{y}-\hat{\boldsymbol{z}}\right)$,

$$\hat{\boldsymbol{z}}=\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\hat{\delta}}^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right), \tag{10}$$

and

$$\hat{\boldsymbol{z}}_\tau=\boldsymbol{K}_\tau\left(\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}+\left(\tau+\hat{\delta}^2\right)\boldsymbol{I}\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right).$$

As $\tau\to0$, we have $\boldsymbol{K}_\tau\to\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$, and hence $\hat{\boldsymbol{z}}_\tau\to\hat{\boldsymbol{z}}$. This analytic continuation on $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ ensures the existence of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{z}}$.

Theorem 1 shows that the optimization in Eq. (8) gives an identical solution to BLUP. This equivalence opens a new door to impose the sparse penalty to the optimization in Eq. (8). Compared with regularized GPs [17], [18], where imposing sparsity aims to accelerate computation, the sparsity will be imposed on the random effect $\boldsymbol{z}$ to promote sparse output for RPGP. The proposed nonparametric sparse feature extraction will be detailed in the next subsection.

Theorem 1. *$\hat{\boldsymbol{y}}=\boldsymbol{F}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{z}}$ is the BLUP of the PGP, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{z}}$ are optimized via Eq.* (8).

Proof. Plugging Eq. (10) into $\hat{\boldsymbol{\beta}}=\left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\left(\boldsymbol{y}-\hat{\boldsymbol{z}}\right)$ results in $\hat{\boldsymbol{\beta}}=\left(\boldsymbol{F}^T\boldsymbol{K}_{\hat{\delta}}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\boldsymbol{K}_{\hat{\delta}}^{-1}\boldsymbol{y}$, which is the same as the MLE of $\hat{\boldsymbol{\beta}}$ provided in Eq. (3). Thus, $\hat{\boldsymbol{y}}=\boldsymbol{F}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{z}}=\boldsymbol{F}\hat{\boldsymbol{\beta}}+\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\hat{\delta}}^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{\beta}}\right)$ is the same as the BLUP provided in Eq. (5). This completes the proof. □

A. Regularized Periodic Gaussian Process

We assume that the random effect $\boldsymbol{z}$ is sparse in this study by adding an L1 penalty to the optimization (9), which becomes

$$\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{z}}_\lambda=\arg\min_{\boldsymbol{\beta},\boldsymbol{z}}\frac{\frac{1}{\hat{\delta}^2}\left\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right\|_2^2+\boldsymbol{z}^T\boldsymbol{K}_\tau^{-1}\boldsymbol{z}}{2}+\lambda\left\|\boldsymbol{z}\right\|_1. \tag{11}$$

Unlike conventional sparse methods that impose the sparse penalty on the fixed effect $\boldsymbol{F}\boldsymbol{\beta}$, the proposed method introduces the sparse penalty to the random effect $\boldsymbol{z}$. This difference allows the proposed sparse feature more flexible to be applied to real signals. For example, the extracted sparse feature $\boldsymbol{F}\hat{\boldsymbol{\beta}}_\lambda+\hat{\boldsymbol{z}}_\lambda$ can be nonzero, even though the sparse fixed effect $\hat{\boldsymbol{z}}_\lambda$ is zero. As far as we know, there may exist a drift term in the periodic signals, which may not be well considered by conventional methods. This drift term can be regarded as the fixed effect by the proposed method. More discussions will be provided in the simulation study.

According to ADMM [14], this optimization is equivalent to

$$\hat{\boldsymbol{\beta}}_\lambda,\hat{\boldsymbol{z}}_\lambda=\arg\min_{\boldsymbol{\beta},\boldsymbol{z},\boldsymbol{v}}\frac{\frac{1}{\hat{\delta}^2}\left\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right\|_2^2+\boldsymbol{z}^T\boldsymbol{K}_\tau^{-1}\boldsymbol{z}}{2}+\lambda\left\|\boldsymbol{v}\right\|_1$$

subject to $\boldsymbol{z}-\boldsymbol{v}=0$. The corresponding augmented Lagrangian function is

$$h\left(\boldsymbol{\beta},\boldsymbol{z}\right)+g\left(\boldsymbol{v}\right)+\boldsymbol{u}\left(\boldsymbol{z}-\boldsymbol{v}\right)+\frac{\rho}{2\hat{\delta}^2}\left\|\boldsymbol{z}-\boldsymbol{v}\right\|_2^2,$$

where $g\left(\boldsymbol{v}\right)=\lambda\left\|\boldsymbol{v}\right\|_1$ and

$$h\left(\boldsymbol{\beta},\boldsymbol{z}\right)=\frac{1}{2}\left(\frac{\left\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}-\boldsymbol{z}\right\|_2^2}{\hat{\delta}^2}+\boldsymbol{z}^T\boldsymbol{K}_\tau^{-1}\boldsymbol{z}\right).$$

According to ADMM, the updating algorithm for $\boldsymbol{z}$ is

$$\boldsymbol{z}^{(m+1)}=\frac{1}{1+\rho}\boldsymbol{K}_\tau\left(\boldsymbol{K}_\tau+\frac{\hat{\delta}^2}{1+\rho}\boldsymbol{I}\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}^{(m)}\right)+\frac{\rho}{1+\rho}\boldsymbol{K}_\tau\left(\boldsymbol{K}_\tau+\frac{\hat{\delta}^2}{1+\rho}\boldsymbol{I}\right)^{-1}\left(\boldsymbol{v}^{(m)}-\boldsymbol{u}^{(m)}\right),$$

and the updating algorithm of other parameters is provided in Eq. (12). By taking the limit, we have

$$\boldsymbol{z}^{(m+1)}\to\frac{\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\tilde{\delta}}^{-1}\left(\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}^{(m)}\right)+\rho\left(\boldsymbol{v}^{(m)}-\boldsymbol{u}^{(m)}\right)\right)}{1+\rho}$$

because $\boldsymbol{K}_\tau\left(\boldsymbol{K}_\tau+\frac{\hat{\delta}^2}{1+\rho}\boldsymbol{I}\right)^{-1}\to\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\tilde{\delta}}^{-1}$ as $\tau\to0$, where $\boldsymbol{K}_{\tilde{\delta}}=\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}+\tilde{\delta}^2\boldsymbol{I}$ with $\tilde{\delta}^2=\hat{\delta}^2/\left(1+p\right)$. Thus, the proposed updating algorithm can be simplified as

$$\begin{cases}\boldsymbol{z}^{(m+1)} &= \frac{\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\tilde{\delta}}^{-1}\left(\left(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{\beta}^{(m)}\right)+\rho\left(\boldsymbol{v}^{(m)}-\boldsymbol{u}^{(m)}\right)\right)}{1+\rho}\\ \boldsymbol{\beta}^{(m+1)} &= \left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\left(\boldsymbol{y}-\boldsymbol{z}^{(m+1)}\right)\\ \boldsymbol{v}^{(m+1)} &= \mathcal{S}_{\lambda/\rho}\left(\boldsymbol{z}^{(m+1)}+\boldsymbol{u}^{(m)}\right)\\ \boldsymbol{u}^{(m+1)} &= \boldsymbol{u}^{(m)}+\boldsymbol{z}^{(m+1)}-\boldsymbol{v}^{(m+1)}\end{cases}, \tag{12}$$

where $\rho = 1$ and $\mathcal{S}_b(a) = a\max(0, 1 - b/|a|)$. The initial values suggested in this study are $\boldsymbol{u}^{(0)} = \boldsymbol{0}$, $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$, and $\boldsymbol{v}^{(0)} = \hat{\boldsymbol{z}} = \boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{K}_{\tilde{\delta}}^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}\right)$. The loop continues until a pre-specified stopping criterion (e.g. $\left\|\boldsymbol{v}^{(k+1)} - \boldsymbol{v}^{(k)}\right\|_\infty < 10^{-6}$) is satisfied. Finally, $\hat{\boldsymbol{\beta}}_\lambda = \boldsymbol{\beta}^{(m+1)}$ and $\hat{\boldsymbol{z}}_\lambda = \boldsymbol{z}^{(m+1)}$.

The loop in Eq. (12) intensively involves the matrix inverse of $\boldsymbol{K}_{\tilde{\delta}}$. Generally, the complexity of matrix inverse is $\mathcal{O}(n^3)$, which is extremely time-consuming. As the signals are grid data, $\boldsymbol{K}_{\tilde{\delta}}$ is a Toeplitz matrix. Fast computing algorithms for Toeplitz matrices reduce the complexity of matrix inverse to $\mathcal{O}(n^2)$. Thus, the Toeplitz-based RPGP (TRPGP) has a $\mathcal{O}(n^2)$ time complexity.

However, the computing time of TRPGP is highly time-consuming because the loop in Eq. (12) needs to be run hundreds of times to estimate $\hat{\boldsymbol{z}}_\lambda$ and select the optimal λ. Based on the assumption of grid observations, an efficient approach utilizing fast computing algorithms for circulant matrices is proposed in the next subsection to alleviate the prohibitive computation.

B. Circulant-based Computational Acceleration

In this study, we assume the signals are collected with the sampling frequency f_s, that is, $t_i = i/f_s$. According to CPGP, the estimated period can be written as $\hat{T} = p/(df_s)$, where p and d are co-prime integers whose greatest common divisor is 1. In other words, there are p/d signal points in the one cycle of the signal. This indicates that the proposed RPGP allows decimal (instead of integer) period in the sparse periodic feature.

Aiming to accelerate the computation, the signal $\boldsymbol{y}$ is split into $k = \lfloor n/p \rfloor$ segments $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k\}$ each has p points and a remaining segment $\boldsymbol{y}_*$ if $n > kp$, that is, $\boldsymbol{y} = \left[\ \boldsymbol{y}_1;\ \ \cdots;\ \ \boldsymbol{y}_k;\ \ \boldsymbol{y}_*\ \right]$, where the semicolon notation denotes vector/matrix connection in the vertical direction. The regression matrix $\boldsymbol{F}$ is also split in a similar way to $\boldsymbol{y}$, which is denoted as $\boldsymbol{F} = \left[\ \boldsymbol{\Gamma}_1;\ \ \cdots;\ \ \boldsymbol{\Gamma}_k;\ \ \boldsymbol{\Gamma}_*\ \right]$. Similarly, we denote

$$\boldsymbol{K}_{\hat{\boldsymbol{\phi}}} = \begin{bmatrix} \boldsymbol{R} & \cdots & \boldsymbol{R} & \boldsymbol{R}_\bullet \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{R} & \cdots & \boldsymbol{R} & \boldsymbol{R}_\bullet \\ \boldsymbol{R}_\bullet^T & \cdots & \boldsymbol{R}_\bullet^T & \boldsymbol{R}_* \end{bmatrix}, \tag{13}$$

where $\boldsymbol{R}$ denotes the $p \times p$ covariance matrix of $\boldsymbol{y}_i$, $\boldsymbol{R}_\bullet$ denotes the $p \times (n - kp)$ covariance matrix between $\boldsymbol{y}_i$ and $\boldsymbol{y}_*$ for $i = 1, 2, \cdots, k$, and $\boldsymbol{R}_*$ denotes the $(n - kp) \times (n - kp)$ covariance matrix of $\boldsymbol{y}_*$. It is worth noting that $\boldsymbol{R}_\bullet$ and $\boldsymbol{R}_*$ are sub-matrices of $\boldsymbol{R}$ because $p/(df_s)$ is the period of the signal.

Theorem 2 reveals that $\boldsymbol{z}^{(m)}$ is always periodic, whereas $\boldsymbol{u}^{(m)}$ and $\boldsymbol{v}^{(m)}$ are periodic for $m > 0$ if $\boldsymbol{u}^{(0)}$ is a periodic sequence. The proof of Theorem 2 is based on Lemma 1. Provided the suggested initial values for the loop in Eq. (12), $\boldsymbol{z}^{(m)}$, $\boldsymbol{v}^{(m)}$, and $\boldsymbol{u}^{(m)}$ are always periodic sequences with the same period p in each iteration. Their period structure can be utilized to accelerate the computation of the loop.

Lemma 1. *The matrix $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ is a periodic matrix with the period p.*

Proof. Clearly, p is one (unnecessarily the smallest) period of the matrix $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ because it repeats the matrix $\boldsymbol{R}$. It is sufficient to show that p is the smallest period. As $\boldsymbol{R}$ is a circulant matrix, we only need to study the first row of $\boldsymbol{R}$, which is

$$\psi_{\hat{\boldsymbol{\phi}}}(t_1, t_1), \psi_{\hat{\boldsymbol{\phi}}}(t_1, t_2), \cdots, \psi_{\hat{\boldsymbol{\phi}}}(t_1, t_p),$$

where $t_i = i/f_s$. The period of $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ cannot be smaller than p because it can be proved that the first row elements are different, that is, $\psi_{\hat{\boldsymbol{\phi}}}(t_1, t_i) \neq \psi_{\hat{\boldsymbol{\phi}}}(t_1, t_j)$ if $i \neq j$, where $i, j \leq p$. This conclusion is true because $\psi_{\hat{\boldsymbol{\phi}}}(t_1, t_i) = \psi_{\hat{\boldsymbol{\phi}}}(t_1, t_j)$ implies that $i \mod p/d = j \mod p/d$, which is equivalent to $id \mod p = jd \mod p$, where $\mod$ denotes the modulo operation. As p and d are co-prime, $id \mod p = jd \mod p$ implies $i = j$ when $i, j \leq p$, and thus $\psi_{\hat{\boldsymbol{\phi}}}(t_1, t_i) = \psi_{\hat{\boldsymbol{\phi}}}(t_1, t_j)$ implies $i = j$. This completes the proof. □

Theorem 2. *$\boldsymbol{z}^{(m)}$ is a periodic sequence with the period p if $t_i = i/f_s$. So are $\boldsymbol{u}^{(m)}$ and $\boldsymbol{v}^{(m)}$ for $m > 0$ if $\boldsymbol{u}^{(0)}$ is a periodic sequence with the period p.*

Proof. According to ADMM updating algorithm provided in Eq. (12), we have $\boldsymbol{z}^{(m+1)} = \boldsymbol{K}_{\hat{\boldsymbol{\phi}}}\boldsymbol{\alpha}^{(m)}$, where $\boldsymbol{\alpha}^{(m)} = \frac{1}{1+\rho}\boldsymbol{K}_{\tilde{\delta}}^{-1}\left(\left(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}^{(m)}\right) + \rho\left(\boldsymbol{v}^{(m)} - \boldsymbol{u}^{(m)}\right)\right)$ is a column vector, so each element of $\boldsymbol{z}^{(m+1)}$ is a fixed weighted sum of the corresponding row of $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$. Lemma 1 shows that $\boldsymbol{K}_{\hat{\boldsymbol{\phi}}}$ is a periodic matrix with a period of p, and thus $\boldsymbol{z}^{(m)}$ is periodic with the same period p for all $m > 0$. Meanwhile, the two equations

$$\begin{cases} \boldsymbol{v}^{(m+1)} & = \mathcal{S}_{\lambda/\rho}\left(\boldsymbol{z}^{(m+1)} + \boldsymbol{u}^{(m)}\right) \\ \boldsymbol{u}^{(m+1)} & = \boldsymbol{u}^{(m)} + \boldsymbol{z}^{(m+1)} - \boldsymbol{v}^{(m+1)} \end{cases},$$

reveal that $\boldsymbol{v}^{(m+1)}$ and $\boldsymbol{u}^{(m+1)}$ are periodic sequences similar to $\boldsymbol{z}^{(m+1)}$ if $\boldsymbol{u}^{(m)}$ is a periodic sequence with the same period as $\boldsymbol{z}^{(m+1)}$. As $\boldsymbol{u}^{(0)}$ is a periodic sequence with the period p, we can conclude that $\boldsymbol{v}^{(m)}$, and $\boldsymbol{u}^{(m)}$ are all periodic sequences with the same period p for $m > 0$ by the mathematical induction. This completes the proof. □

As $\boldsymbol{z}^{(m)}$ is a periodic sequence, it can be written as $\boldsymbol{z}^{(m)} = [\bar{\boldsymbol{z}}^{(m)}; \cdots; \bar{\boldsymbol{z}}^{(m)}; \boldsymbol{z}_*^{(m)}]$, where $\bar{\boldsymbol{z}}^{(m)}$ is a $p \times 1$ vector consisting of the first p elements of $\boldsymbol{z}^{(m)}$, and $\boldsymbol{z}_*^{(m)}$ is an $(n - kp) \times 1$ vector consisting of the first $n - kp$ elements of $\boldsymbol{z}^{(m)}$. Thus, Eq. (12) implies that

$$\bar{\boldsymbol{z}}^{(m+1)} = \frac{\bar{\boldsymbol{K}}_{\hat{\boldsymbol{\phi}}}^T\boldsymbol{K}_{\tilde{\delta}}^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}^{(m)} + \rho\left(\boldsymbol{v}^{(m)} - \boldsymbol{u}^{(m)}\right)\right)}{1 + \rho}, \tag{14}$$

where $\bar{\boldsymbol{K}}_{\hat{\boldsymbol{\phi}}} = \left[\ \boldsymbol{R},\ \cdots,\ \boldsymbol{R},\ \boldsymbol{R}_\bullet\ \right]^T$. Similarly, we denote $\boldsymbol{u}^{(m)} = [\bar{\boldsymbol{u}}^{(m)}; \cdots; \bar{\boldsymbol{u}}^{(m)}; \bar{\boldsymbol{u}}_*^{(m)}]$ and $\boldsymbol{v}^{(m)} =$

$[\bar{\boldsymbol{v}}^{(m)};\cdots;\bar{\boldsymbol{v}}^{(m)};\bar{\boldsymbol{v}}_*^{(m)}]$. To reduce the computation of $\bar{\boldsymbol{z}}^{(m)}$, the covariance matrix $\boldsymbol{K}_{\tilde{\delta}}$ is decomposed into a block matrix as the following:

$$\boldsymbol{K}_{\tilde{\delta}} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Xi} \\ \boldsymbol{\Xi}^T & \boldsymbol{\Sigma}_* \end{bmatrix},$$

where $\boldsymbol{\Sigma}_* = \boldsymbol{R}_* + \tilde{\delta}^2\boldsymbol{I}_{n-kp}$, $\boldsymbol{\Xi} = \begin{bmatrix}\boldsymbol{R}_\bullet^T \cdots \boldsymbol{R}_\bullet^T\end{bmatrix}^T$, and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{R}+\tilde{\delta}^2\boldsymbol{I}_p & \cdots & \boldsymbol{R} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R} & \cdots & \boldsymbol{R}+\tilde{\delta}^2\boldsymbol{I}_p \end{bmatrix}.$$

According to the block matrix inversion formula, the inverse of the matrix $\boldsymbol{K}_{\tilde{\delta}}$ can be written as

$$\boldsymbol{K}_{\tilde{\delta}}^{-1} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{Q} \\ \boldsymbol{Q}^T & \boldsymbol{\Pi}^{-1} \end{bmatrix}, \tag{15}$$

where

$$\begin{cases} \boldsymbol{P} &= \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}\boldsymbol{\Pi}^{-1}\boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1} \\ \boldsymbol{Q} &= -\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}\boldsymbol{\Pi}^{-1} \\ \boldsymbol{\Pi} &= \boldsymbol{\Sigma}_* - \boldsymbol{\Xi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi} \end{cases}. \tag{16}$$

By applying the matrix inversion lemma [26] twice, as is provided in CPGP [23], we have

$$\boldsymbol{\Sigma}^{-1} = \frac{k\boldsymbol{I}_{kp} - \boldsymbol{V}\left(\boldsymbol{I}_p - \boldsymbol{R}_{\tilde{\delta}}^{-1}\right)\boldsymbol{V}^T}{\tilde{\delta}^2 k}, \tag{17}$$

where $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{I}_p, & \cdots, & \boldsymbol{I}_p\end{bmatrix}^T$ and $\boldsymbol{R}_{\tilde{\delta}} = \boldsymbol{I}_p + k\boldsymbol{R}/\tilde{\delta}^2$. By plugging Eq. (17) into Eq. (16), we have

$$\begin{cases} \boldsymbol{P} &= \boldsymbol{\Sigma}^{-1} + \frac{1}{\tilde{\delta}^2}\boldsymbol{V}\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{V}^T\frac{1}{\tilde{\delta}^2} \\ \boldsymbol{Q} &= -\frac{1}{\tilde{\delta}^2}\boldsymbol{V}\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1} \\ \boldsymbol{\Pi} &= \boldsymbol{\Sigma}_* - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}_\bullet \end{cases}. \tag{18}$$

Plugging Eqs. (15) and (18) into Eq. (14), we have

$$\begin{aligned} \bar{\boldsymbol{z}}^{(m+1)} = &\frac{k\boldsymbol{R}\boldsymbol{R}_{\tilde{\delta}}^{-1}\left(\bar{\boldsymbol{y}} - \bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}^{(m)} + \rho\left(\bar{\boldsymbol{v}}^{(m)} - \bar{\boldsymbol{u}}^{(m)}\right)\right)}{\tilde{\delta}^2\left(1+\rho\right)} \\ &+ \frac{\boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\left(\boldsymbol{y}_\bullet - \boldsymbol{\Gamma}_\bullet\boldsymbol{\beta}^{(m)} + \rho\left(\boldsymbol{v}_\bullet^{(m)} - \boldsymbol{u}_\bullet^{(m)}\right)\right)}{\left(1+\rho\right)}, \end{aligned} \tag{19}$$

because

$$\begin{cases} \bar{\boldsymbol{K}}_{\tilde{\phi}}^T\boldsymbol{K}_{\tilde{\delta}}^{-1}\boldsymbol{y} &= \frac{k}{\tilde{\delta}^2}\boldsymbol{R}\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{y}} + \boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\boldsymbol{y}_\bullet \\ \bar{\boldsymbol{K}}_{\tilde{\phi}}^T\boldsymbol{K}_{\tilde{\delta}}^{-1}\boldsymbol{F} &= \frac{k}{\tilde{\delta}^2}\boldsymbol{R}\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}^{(m)} + \boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\boldsymbol{\Gamma}_\bullet \\ \bar{\boldsymbol{K}}_{\tilde{\phi}}^T\boldsymbol{K}_{\tilde{\delta}}^{-1}\boldsymbol{v}^{(m)} &= \frac{k}{\tilde{\delta}^2}\boldsymbol{R}\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{v}}^{(m)} + \boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\boldsymbol{v}_\bullet^{(m)} \\ \bar{\boldsymbol{K}}_{\tilde{\phi}}^T\boldsymbol{K}_{\tilde{\delta}}^{-1}\boldsymbol{u}^{(m)} &= \frac{k}{\tilde{\delta}^2}\boldsymbol{R}\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{u}}^{(m)} + \boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\boldsymbol{u}_\bullet^{(m)} \end{cases}$$

where $\bar{\boldsymbol{y}} = \left(\boldsymbol{y}_1 + \cdots + \boldsymbol{y}_k\right)/k$, $\bar{\boldsymbol{\Gamma}} = \left(\boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\Gamma}_k\right)/k$, and

$$\begin{cases} \boldsymbol{\Gamma}_\bullet &= \boldsymbol{\Gamma}_* - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{\Gamma}} \\ \boldsymbol{y}_\bullet &= \boldsymbol{y}_* - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{y}} \\ \boldsymbol{v}_\bullet^{(m)} &= \boldsymbol{v}_*^{(m)} - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{v}}^{(m)} \\ \boldsymbol{u}_\bullet^{(m)} &= \boldsymbol{u}_*^{(m)} - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\bar{\boldsymbol{u}}^{(m)} \\ \boldsymbol{I}_\bullet &= \boldsymbol{I}_p - \frac{k}{\tilde{\delta}^2}\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R} \end{cases}.$$

Recall that $\boldsymbol{z}^{(m)}$, $\boldsymbol{u}^{(m)}$, and $\boldsymbol{v}^{(m)}$ are periodic sequences, so only one of their period (i.e., $\bar{\boldsymbol{z}}^{(m)}$, $\bar{\boldsymbol{u}}^{(m)}$, and $\bar{\boldsymbol{v}}^{(m)}$) are required to be computed in each step of the loop, that is,

$$\begin{cases} \bar{\boldsymbol{z}}^{(m+1)} &= \frac{k\boldsymbol{R}^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\left(\bar{\boldsymbol{y}} - \bar{\boldsymbol{\Gamma}}\boldsymbol{\beta}^{(m)} + \rho\left(\bar{\boldsymbol{v}}^{(m)} - \bar{\boldsymbol{u}}^{(m)}\right)\right)}{\tilde{\delta}^2(1+\rho)} \\ &\quad + \frac{\boldsymbol{I}_\bullet^T\boldsymbol{R}_\bullet\boldsymbol{\Pi}^{-1}\left(\boldsymbol{y}_\bullet - \boldsymbol{\Gamma}_\bullet\boldsymbol{\beta}^{(m)} + \rho\left(\boldsymbol{v}_\bullet^{(m)} - \boldsymbol{u}_\bullet^{(m)}\right)\right)}{(1+\rho)} \\ \boldsymbol{z}^{(m+1)} &= \left[\bar{\boldsymbol{z}}^{(m+1)};\cdots;\bar{\boldsymbol{z}}^{(m+1)};\boldsymbol{z}_*^{(m+1)}\right] \\ \boldsymbol{\beta}^{(m+1)} &= \left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\left(\boldsymbol{y} - \boldsymbol{z}^{(m+1)}\right) \\ \bar{\boldsymbol{v}}^{(m+1)} &= \mathcal{S}_{\lambda/\rho}\left(\bar{\boldsymbol{z}}^{(m+1)} + \bar{\boldsymbol{u}}^{(m)}\right) \\ \bar{\boldsymbol{u}}^{(m+1)} &= \bar{\boldsymbol{u}}^{(m)} + \bar{\boldsymbol{z}}^{(m+1)} - \bar{\boldsymbol{v}}^{(m+1)} \end{cases}.$$

Computing the above equations is computationally efficient because FFT can be used for circulant matrices. $\boldsymbol{R}^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}$ is a product of three circulant matrices and thus is still circulant and apparently symmetric. So $\boldsymbol{R}_\bullet^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}_\bullet$, which is a sub-matrix of $\boldsymbol{R}^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}$, is a symmetric Toeplitz matrix. Thus, $\boldsymbol{\Pi}$ is also symmetric Toeplitz, and it is a sub-matrix of $\boldsymbol{C} = \boldsymbol{R} + \tilde{\delta}^2\boldsymbol{I}_p - k\boldsymbol{R}^T\boldsymbol{R}_{\tilde{\delta}}^{-1}\boldsymbol{R}/\tilde{\delta}^2$. Hence, computing $\boldsymbol{\Pi}$ using FFT only requires $\mathcal{O}\left(p\log p\right)$ complexity because only the first column of $\boldsymbol{\Pi}$ needs to be computed. Moreover, circulant embedding [27], [28] can be used to solve the linear equation $\boldsymbol{\Pi}\boldsymbol{a} = \boldsymbol{b}$ for any vector $\boldsymbol{a}$ and $\boldsymbol{b}$ via linear conjugate gradients (LCG), which has a $\mathcal{O}\left(p\log p\right)$ complexity. Specifically, the matrix $\boldsymbol{\Pi}$ is embedded into the matrix $\boldsymbol{C}$, and $\boldsymbol{\Pi}\boldsymbol{a}$ is a sub-vector of $\boldsymbol{C}\left[\boldsymbol{a}^T, \boldsymbol{0}\right]^T$, which can be efficiently evaluated via FFT. Thus, $\boldsymbol{\Pi}^{-1}\left(\boldsymbol{y}_\bullet - \boldsymbol{\Gamma}_\bullet\boldsymbol{\beta}^{(m)} + \rho\left(\boldsymbol{v}_\bullet^{(m)} - \boldsymbol{u}_\bullet^{(m)}\right)\right)$ in Eq. (19) can be computed more efficiently.

C. Adaptive Model Selection

Selection of the penalty parameter λ is fundamental to lasso problems because the degree of sparsity depends on the value of λ. One method to select the optimal λ is cross-validation, which may be highly time-consuming and inaccurate. Another model selection criterion is using the Akaike information criterion (AIC) or Bayesian information criterion (BIC) [29]. Specifically, this study employs Stein's unbiased risk estimate [15], [16], which is based on the degrees of freedom of the Lasso fit, to construct AIC and BIC.

The penalty parameter λ can be optimized via AIC, i.e., $\hat{\lambda} = \arg\min_\lambda \mathrm{AIC}\left(\lambda\right)$, where

$$\mathrm{AIC}\left(\lambda\right) = \frac{1}{n\,\hat{\sigma}^2\hat{\delta}^2}\left\|\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}_\lambda - \hat{\boldsymbol{z}}_\lambda\right\|_2^2 + \frac{2}{n}\widehat{df}\left(\lambda\right), \tag{20}$$

or it can be optimized via BIC, i.e., $\hat{\lambda} = \arg\min_{\lambda} \mathrm{BIC}(\lambda)$, where

$$\mathrm{BIC}(\lambda) = \frac{1}{n\hat{\sigma}^2\hat{\delta}^2}\left\|\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}_{\lambda} - \hat{\boldsymbol{z}}_{\lambda}\right\|_2^2 + \frac{\log n}{n}\widehat{df}(\lambda). \quad (21)$$

Here, $\widehat{df}(\lambda)$ denotes the degree of freedom for the random effect $\boldsymbol{z}$ [15], [30]. Theorem 2 indicates that $\hat{\boldsymbol{z}}_{\lambda}$ is a periodic sequence with the period p, so the degree freedom $\widehat{df}(\lambda)$ is the number of nonzero elements in one period of $\hat{\boldsymbol{z}}_{\lambda}$, i.e., the number of nonzero elements in $\bar{\boldsymbol{z}}^{(m+1)}$.

AIC tends to select the regularization parameter λ with the optimal prediction performance, whereas BIC tends to identify the true sparse model. As the sparsity feature extraction is the aim of RPGP, BIC is suggested in this study as the model selection criterion.

IV. Simulation Study

In this section, the performance of the proposed RPGP method for sparse feature extraction is studied on synthetic signals. We first compare the computing time of RPGP and TRPGP for varying signal lengths and then apply RPGP to extract sparse features from synthetic signals with a global trend. The synthetic signals have a fixed effect and a random effect in the form of Eq. (1). The quadratic polynomial basic functions are chosen to construct the fixed effect, that is, $\boldsymbol{f}(t) = \left[1, t/\tau, (t/\tau)^2\right]^T$, where $\tau = 2000$ is the length of synthetic signals, and the regression coefficient vector is set to be $\boldsymbol{\beta} = [2, -4, 4]^T$. Signals with transient periodic structures that mimic rotary mechanical vibration signals are utilized to generate the random effect $z(t)$ according to the following equation:

$$z(t) = \sum_{k=0}^{\left\lfloor \frac{\tau}{T_0} \right\rfloor} e^{\frac{-\zeta 2\pi\omega(t-kT_0)^2}{\sqrt{1-\zeta^2}}} \sin 2\pi\omega(t-kT_0), \qquad t \le \tau$$

where $\zeta = 0.01$ is the damping ratio, $\omega = 11$ Hz, $T_0 = 1$ s is the true period, and the sampling frequency of signals is $f_s = 200$ Hz. Gaussian white noises with a standard deviation of 0.5 (i.e., $\sigma = 0.5$) are added to generate the synthetic signals, and a typical waveform is shown in Fig. 1 (d).

The first experiment compares the average computing time (of 10 trials) for TRPGP and RPGP dealing with various signal lengths. The results are reported in Tab. I. As shown in Tab. I, the computation time for TRPGP is much higher than RPGP for all signal lengths tested. Meanwhile, the average computing time for TRPGP increases quadratically with the signal lengthening. The computing time of TRPGP exceeds three hours for signal length greater than 6000, indicating that applying TRPGP to real applications is prohibitive. In contrast, the average computing time of RPGP is substantially faster and remains almost constant at around 30 seconds in each trial for all signal lengths. Tab. I reveals that the computation complexity of RPGP is scalable.

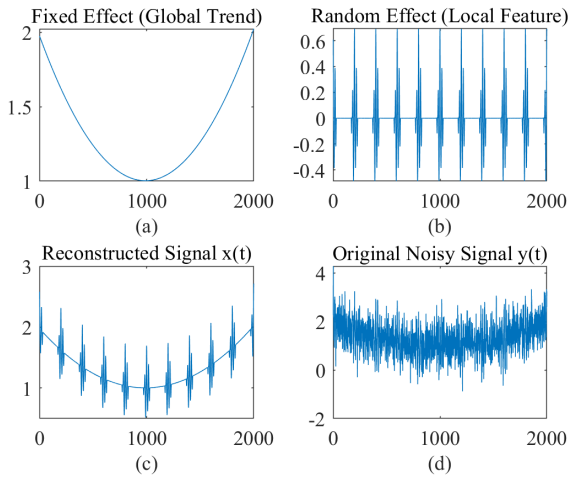

Fig. 2. The reconstructed signal by RPGP, (a) the extracted fixed effect (the global trend), (b) the extracted random effect (the local feature), (c) the reconstructed signal , (d) the original noisy signal.

To validate the effectiveness of RPGP in extracting sparse features from periodic signals with a global trend, we perform another numerical experiment. As shown in Fig. 1, the observed raw signal has a quadratic fixed effect, and the true signal is buried in strong background noises, which makes it hard to extract the sparse periodic feature. During adaptive model selection, a grid search of 30 candidates λ in the domain $[0, 2\hat{\sigma}\hat{\delta}/k]$ is applied to minimize $\mathrm{BIC}(\lambda)$ in Eq. (21). TRPGP and RPGP will produce the same prediction result, so they are not compared here.

Fig. 2 shows the reconstruction results, and the original noisy signal is also shown here for comparison. Compared with Fig. 1, the reconstructed signal by RPGP is a close estimation. The reconstructed fixed effect has the following formulation:

$$\boldsymbol{f}^T(t)\,\hat{\boldsymbol{\beta}} = 1.9762 - 3.9391 \times t/\tau + 3.9855 \times (t/\tau)^2 \quad (22)$$

Eq. (22) indicates that $\hat{\boldsymbol{\beta}}$ has a high approximation quality to the true $\boldsymbol{\beta}$. Moreover, the random effect successfully captures the transient periodic structure of the original random effect. In contrast, simulations show that conventional SR models [6], [10] cannot effectively extract the sparse periodic features from the signals shown in Fig. 1 (d). Thus, one can conclude that the RPGP model is capable of extracting the periodic sparse feature from noisy signals composed of a fixed and random effect.

TABLE I
AVERAGE COMPUTING TIME OF 10 TRIALS (UNIT: SECOND).

n	2000	3000	4000	5000	6000
TRPGP	1585.3	3675.7	6617.0	9654.3	12050.8
RPGP	32.2	33.2	34.2	32.2	29.2

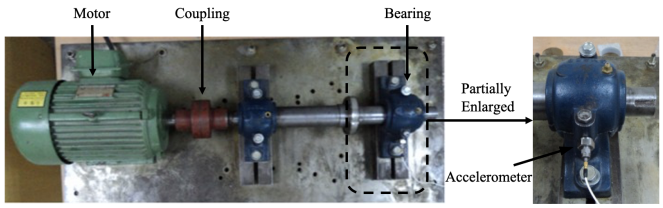

Fig. 3. Bearing test rig.

V. Real Case Study

Rolling bearings are crucial components in motors and rotating machine systems. The defects of the rolling bearings would eventually lead to the failure of the rotary machinery, which would trigger substantial economic losses. Since the rolling bearings usually work in harsh conditions such as high speed and heavy load, they are easily damaged. Thus, the fault diagnosis of the rolling bearings as early as possible is vital to prevent severe consequences. When the rolling elements of bearings pass over a defect, a quasi-periodic impulsive signal is produced. The impulsive transients are not isolated but likely to form groups, so they are often sparse signals [6]. Thus, applying sparse methods to extract fault features is in demand for bearing fault diagnosis.

In this section, the proposed RPGP model is applied to real vibration signals of rolling bearings to further demonstrate its performance in sparse feature extraction. B-spline basis functions with a degree of 3 and 50 equally-spaced knots are used to model the highly nonlinear trend of real vibration signals, instead of simply using polynomial basis functions to construct the fixed effect. The vibration signals are measured from a rotating bearing test rig shown in Fig. 3 with a sampling frequency of 51.2 kHz. In this experiment, the test bearing is a cylindrical roller bearing (NJ208 TMB). And the bearing fault, which is set in the outer raceway through wire cutting, is 0.2 mm in width and 0.2 mm in depth. The parameters of the test bearing are provided in [10]. The collected vibration signals with a length of $N = 10000$ sample points ($0.195\ s$) are analyzed in this study. A segment of the signals is shown in Fig. 4 (a).

CPGP is first applied to fit a PGP model before RPGP is applied. The estimated period $\hat{p}$ is 359, coinciding with the fault characteristic period. The reconstructed signal, as well as the extracted global trend and periodic sparse feature, are shown in Fig. 4 (b), (c), and (d). Compared with the measured vibration signal, the recovered sparse signal in Fig. 4 (c) is considerably clear. It shows a periodic impulsive feature, which suggests that the proposed mixed effects model can filter out meaningless background noises while extracting sparse periodic transient characteristics of the vibration signal.

RPGP also extracts a global trend, shown in Fig. 4 (b). In practice, there may be a complex trend in the signal when bearing vibrations are affected by complex operating conditions. For example, mechanical looseness under harsh conditions may trigger an increasing or decreasing trend in the vibration signals. Unfortunately, conventional methods assume the signals to be zero-centered with no global trend, which may cease to work properly under such a circumstance. These methods have to do mean subtraction in advance to analyze the signals, which is not necessary for RPGP. Moreover, the fixed effect of the periodic signals may not be exactly zero even after mean subtraction due to the nonlinear nature of the trend, and ignoring the global trend would inevitably lead to information loss. Thus, conventional methods may over/under-estimate the amplitude of the sparse periodic feature by simply using mean subtraction. In addition, for signals that are supposed to be zero-centered, the extracted global trend can indicate whether the test rig or sensors need further calibration. Therefore, extracting the global trend and the local sparse feature separately can provide guidance for further bearing diagnoses.

In this case study, no dictionary is used for extracting sparse feature in RPGP because it is a nonparametric method. The difficulty of choosing the appropriate dictionaries and their hyper-parameters no longer exists. Therefore, RPGP could be a handy tool for feature extraction in real applications. Furthermore, the extracted sparse feature in Fig. 4 (c) is quite complex, exhibiting its flexibility to extract sparse features for highly nonlinear signals. Moreover, the RPGP model also extracts a linear trend, although the slope is quite small. This means that RPGP is able to process periodic signals with a non-constant trend, rather than assuming the global trend is zero. The extracted sparse feature and the global trend can be further applied for fault diagnosis.

VI. Conclusions

This paper proposed a GP-based nonparametric model for sparse feature extraction. The nonlinear and nonparametric nature of RPGP enables the processing of highly nonlinear signals without suffering from the difficulty of choosing a dictionary. Another advantage of RPGP over conventional methods is that it utilizes the mixed effects structure to simultaneously extract the global trend and the local sparse feature of the signals, making it effective for highly nonlinear signals with a non-constant trend. An efficient approach using fast computing algorithms for circulant matrices is proposed to increase the computational efficiency, making RPGP a handy tool for the sparse feature extraction of periodic signals. A model selection criterion is proposed to automatically identify a reliable model for given signals.

The proposed method still has some room for improvement. First, we assume that the sparse feature is strictly periodic, which may not be the case for real applications. Therefore, our future research will develop a more robust model for sparse feature extraction of pseudo-periodic signals [31] in a complex

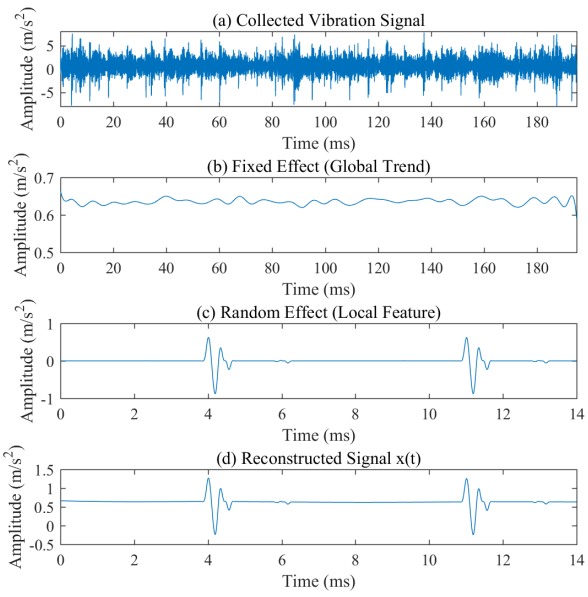

Fig. 4. The original signal and a segment of extracted feature of the vibration signal, (a) the collected vibration signal, (b) the extracted fixed effect (the global trend), (c) the extracted random effect (the periodic sparse feature), (d) the reconstructed vibration signal.

noise environment. Second, the penalty utilized in this paper is a conventional L1 penalty, which may underestimate the amplitude of impulses. Therefore, a new modeling framework using non-convex penalties [13] will be investigated in our future work to address this issue.

REFERENCES

[1] T. Wang, Z. Liu, G. Lu, and J. Liu, "Temporal-spatio graph based spectrum analysis for bearing fault detection and diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2598–2607, 2021.

[2] W. Fan, Z. Chen, Y. Li, F. Zhu, and M. Xie, "A reinforced noise resistant correlation method for bearing condition monitoring," *IEEE Transactions on Automation Science and Engineering*, 2022.

[3] X. Chen, J. Lin, C. Huang, and L. He, "A novel method based on adaptive periodic segment matrix and singular value decomposition for removing emg artifact in ecg signal," *Biomedical Signal Processing and Control*, vol. 62, p. 102060, 2020.

[4] B. Sun, X. Zhao, H. Zhang, R. Bai, and T. Li, "Eeg motor imagery classification with sparse spectrotemporal decomposition and deep learning," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 541–551, 2021.

[5] Y. Qin, "A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2716–2726, 2017.

[6] W. Fan, Y. Li, L. Chen, and Z. Xu, "Bearing fault detection via b-spline constructed sparse method," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[7] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[8] H. Liu, C. Liu, and Y. Huang, "Adaptive feature extraction using sparse coding for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 558–574, 2011.

[9] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Transactions on neural networks and learning systems*, vol. 23, no. 7, pp. 1013–1027, 2012.

[10] Y. Li, Y. Pu, W. Fan, and J. Wu, "Constraint linear model for period estimation and sparse feature extraction based on iterative likelihood ratio test," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 3, pp. 1–10, 2023.

[11] J. Shi and M. Liang, "Intelligent bearing fault signature extraction via iterative oscillatory behavior based signal decomposition (iobsd)," *Expert Systems with Applications*, vol. 45, pp. 40–55, 2016.

[12] W. Fan, G. Cai, Z. Zhu, C. Shen, W. Huang, and L. Shang, "Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction," *Mechanical Systems and Signal Processing*, vol. 56, pp. 230–245, 2015.

[13] Z. Zhao, S. Wu, B. Qiao, S. Wang, and X. Chen, "Enhanced sparse period-group lasso for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 2143–2153, 2019.

[14] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2011.

[15] H. Zou, T. Hastie, and R. Tibshirani, "On the "degrees of freedom" of the lasso," *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.

[16] R. J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *The annals of statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.

[17] F. Yan and Y. A. Qi, "Sparse gaussian process regression via l1 penalization," in *ICML*, 2010.

[18] P. Kou and F. Gao, "Sparse gaussian process regression model based on $\ell_{1/2}$ regularization," *Applied intelligence*, vol. 40, no. 4, pp. 669–681, 2014.

[19] Y. Li and Q. Zhou, "Pairwise meta-modeling of multivariate output computer models using nonseparable covariance function," *Technometrics*, vol. 58, no. 4, pp. 483–494, 2016.

[20] S. Xiong, "The reconstruction approach: From interpolation to regression," *Technometrics*, vol. 63, no. 2, pp. 225–235, 2021.

[21] D. J. MacKay, "Introduction to gaussian processes," *NATO ASI Series F Computer and Systems Sciences*, vol. 168, pp. 133–166, 1998.

[22] T. J. Santner, B. J. Williams, and W. I. Notz, *The design and analysis of computer experiments*. Springer Science & Business Media, 2013.

[23] Y. Li, Y. Pu, C. Cheng, and Q. Xiao, "A scalable gaussian process for large-scale periodic data," *Technometrics*, no. just-accepted, pp. 1–14, 2023. [Online]. Available: http://arxiv.org/abs/2301.01412

[24] C. E. Rasmussen and C. K. I. Williams, "Gaussian processes for machine learning," *the MIT Press*, 2006.

[25] C. R. Henderson, "Estimation of genetic parameters," *The Annals of Mathematical Statistics*, vol. 21, pp. 309–310, 1950.

[26] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.

[27] A. T. Wood and G. Chan, "Simulation of stationary gaussian processes in [0, 1] d," *Journal of computational and graphical statistics*, vol. 3, no. 4, pp. 409–432, 1994.

[28] J.-F. Coeurjolly and E. Porcu, "Fast and exact simulation of complex-valued stationary gaussian processes through embedding circulant matrix," *Journal of Computational and Graphical Statistics*, vol. 27, no. 2, pp. 278–290, 2018.

[29] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[30] R. J. Tibshirani and J. Taylor, "Degrees of freedom in lasso problems," *The Annals of Statistics*, vol. 40, no. 2, pp. 1198–1232, 2012.

[31] Y. Li, H. Zhao, W. Fan, and C. Shen, "Extended noise resistant correlation method for period estimation of pseudoperiodic signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

Yongxiang Li received his Ph.D. degree in data science from City University of Hong Kong in 2019. Currently, he is an Assistant Professor in the Department of Industrial Engineering and Management and the Chinese Institute for Quality Research at Shanghai Jiao Tong University, Shanghai, China. His research focuses on both the theoretical and applied aspects of data science integrated with domain knowledge for quality and reliability engineering using methodologies from statistics, machine learning, and signal processing. He has been working on the research areas such as computer experiments, statistical quality control, statistical anomaly detection, and intelligent fault diagnostics.

Yunji Zhang received a B.S. degree in the Department of Industrial Engineering and Management from Shanghai Jiao Tong University, Shanghai, China in 2021, where he is currently working toward his Master degree. His research interests include signal processing and statistical learning.

Jianguo Wu received his B.S. degree in Mechanical Engineering in 2009 from Tsinghua University, Beijing, China, M.S. degree in Mechanical Engineering in 2011 from Purdue University, and M.S. degree in Statistics in 2014, and Ph.D. degree in Industrial and Systems Engineering in 2015, both from University of Wisconsin-Madison. Currently, he is an Assistant Professor in the Dept. of Industrial Engineering and Management at Peking University, Beijing, China. He was an Assistant Professor at the Dept. of IMSE at UTEP, TX, USA, from 2015 to 2017. His research interests are focused on data-driven modeling, monitoring, and analysis of advanced manufacturing processes and complex systems for quality control and reliability improvement.

Min Xie (Fellow, IEEE) received the Ph.D. degree in quality technology from Linkoping University, Linkoping, Sweden, in 1987. He is currently the Chair Professor with the Department of Advanced Design and Systems Engineering, City University of Hong Kong, Hong Kong. His research interests include reliability engineering, quality management, software reliability, and applied statistics. Dr. Xie was elected as a member of the European Academy of Sciences and Arts in 2022. He was a recipient of the Prestigious Lee Kuan Yew (LKY) Research Fellowship in 1991. He serves as an editor and an associate editor and on the editorial board of many established international journals. He has chaired many international conferences and given keynote speeches.

Dear editors

We would like to submit the enclosed manuscript entitled "**Regularized Periodic Gaussian Process for Nonparametric Sparse Feature Extraction from Noisy Periodic Signals**", which we wish to be considered for publication in the journal **IEEE Transactions on Automation Science and Engineering**.

This paper proposes a nonparametric sparse feature extraction method for highly nonlinear sparse periodic signals based on a Periodic Gaussian process (PGP). In this study, the PGP model is reformulated as a mixed-effects model. Hence a regularization term is allowed to be imposed on the random effects of the PGP model for sparse feature extraction, called regularized PGP (RPGP). Unlike conventional sparse models, RPGP can simultaneously model both fixed and random effects of the signals. The extracted global trend and local sparse feature can provide valuable insights for the diagnosis of the bearing fault. Furthermore, an efficient approach utilizing fast computing algorithms for circulant matrices is proposed to speed up the computation, making RPGP a handy tool for the sparse feature extraction of periodic signals.

We believe the topic of our research falls within the scope of the journal and our findings will be of interest to its readership. In particular, rather than assuming the signals is zero-centered without trend in conventional methods, the proposed RPGP model will inspire readers to simultaneously extract the global trend and the local sparse feature of the signals. Furthermore, unlike conventional sparse representation methods, the nonlinear and nonparametric characteristics of RPGP exempt readers from the difficulty of selecting a dictionary for highly nonlinear periodic signals. RPGP will shed light on the nonparametric sparse feature extraction of bearing fault signals.

This manuscript has not been published previously and is not under consideration elsewhere. The authors have no conflicts of interest to disclose and have all approved this submission. Your consideration of our manuscript is deeply appreciated, and we are looking forward to receiving comments from the reviewers.

Sincerely yours,

Yongxiang Li, Yunji Zhang, Jianguo Wu, and Min Xie
E-mail: j.wu@pku.edu.cn