University of Kent

Kent Academic Repository

Wang, Xiuli, Jiang, Bin, Wu, Shaomin, Lu, Ningyun and Ding, Steven (2021) ***Multivariate Relevance Vector Regression based Degradation Modeling and Remaining Useful Life Prediction.*** **IEEE Transactions on Industrial Electronics, 69 (9). pp. 9514-9523. ISSN 0278-0046.**

Downloaded from

https://kar.kent.ac.uk/91113/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1109/TIE.2021.3114724

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Multivariate Relevance Vector Regression based Degradation Modeling and Remaining Useful Life Prediction

Xiuli Wang, *Member, IEEE*, Bin Jiang, *Fellow, IEEE*, Shaomin Wu, Ningyun Lu, *Member, IEEE*, and Steven X. Ding,

***Abstract*—Relevance Vector Regression (RVR) is a useful tool for degradation modeling and Remaining Useful Life (RUL) prediction. However, most RVR models are for one-dimensional degradation processes and can only handle univariate observations. This paper proposes a degradation path based RUL prediction framework using a dynamic Multivariate Relevance Vector Regression (MRVR) model. Specifically, a multi-step regression model is established for describing the degradation dynamics and extends the classical RVR into a multivariate one with consideration of the multivariate environment. The paper introduces a matrix Gaussian distribution based RVR approach and then estimates the hyperparameters with Nesterov's accelerated gradient method to avoid the exhausting re-estimation phenomenon in seeking analytical solutions. It further forecasts the degradation path for monitoring the degradation status. Based on the forecasted path, the RUL is predicted by the First Hitting Time (FHT) method. Finally, the proposed methods are illustrated by two case studies, one is presented in the paper and the other in the supplement, both of which investigate the capacitors' performance degradation in the traction systems of high-speed trains.**

***Index Terms*—Degradation process, Multivariate relevance vector regression, Nesterov's accelerated gradient, Remaining useful life, First hitting time, Capacitors.**

I. Introduction

REMAINING Useful Life (RUL), an important concept in Prognostics and Health Management (PHM), has attracted a great deal of attention in recent years. RUL is the time between the current time instant and the end of the useful life. The prediction of RUL helps assess the health status of a system and obtain an estimation of time before a failure occurs, based on historical and on-going degradation evolution and system's operational and usage conditions [1]. Yang et al. proposed a RUL prediction method based on a double-convolutional neural networks (CNNs) model architecture [2]. Cheng et al. developed a novel data-driven framework to exploit the adoption of deep CNNs in predicting the RULs of bearings [3]. Chen et al. projected an attention-based deep learning framework for a machine's RUL prediction [4]. An accurate prediction of the RUL provides valuable information that enables the operator to anticipate the failure occurrence in advance and then plan maintenance accordingly to avoid system failures. It has been widely used for enhancing system safety in various sectors, such as the electronic industry, the chemical industry, the energy industry, etc [5–7].

Approaches to predicting RUL can be categorized into two main groups, depending on the type of condition monitoring data: direct and indirect [8]. The direct RUL prediction approaches mainly focus on the Health Indicator (HI) construction by extracting feature information from the acquired data that can identify and quantify a history and on-going degradation process [9]. The indirect approaches monitor and forecast the evolution of a degradation signal firstly and then predict the RUL [10]. As the quality of the constructed HI largely influences the efficacy of the RUL prediction, this paper proposes a degradation path based RUL method, which is able to monitor the degradation status and predict the RUL, without the need of HI construction.

As for the degradation process, kernel-based learning methods have been widely used for modeling the system degradation, among which the Support Vector Regression (SVR) and Relevance Vector Regression (RVR) are the most well known methods with competitive performance [11, 12]. The SVR is built by minimizing the generalization error bound to achieve generalized performance [13, 14]. The RVR is a Bayesian regression framework, in which the weights of each input are governed by a set of hyperparameters. These hyperparameters describe the posterior distribution of the weights and are estimated iteratively by maximizing the marginal likelihood over the hyperparameters [15, 16]. As mentioned in [17], the RVR offers some advantages over the SVR. The prediction of the RVR is a probabilistic regression model under the Bayesian framework. Moreover, compared to the SVR, the RVR results

This work was supported by the National Natural Science Foundation of China (62020106003, 61873122), Priority Academic Program Development of Jiangsu Higher Education Institutions, and the 111 Project (B20007).

Xiuli Wang, Bin Jiang, and Ningyun Lu are with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China(corresponding author: Bin Jiang; e-mail: binjiang@nuaa.edu.cn).

Shaomin Wu is with Kent Business School, University of Kent, Canterbury, Kent CT2 7FS, United Kingdom.

Steven X. Ding is with the Institute for Automatic Control and Complex Systems (AKS), University of Duisburg-Essen, Duisburg 47057, Germany.

Suggested citation: X. Wang, J. Bin, S. Wu, N. Lu and S. Ding, "Multivariate Relevance Vector Regression based Degradation Modeling and Remaining Useful Life Prediction," in *IEEE Transactions on Industrial Electronics*, doi: 10.1109/TIE.2021.3114724.

in a sparser model and facilitates utilizing arbitrary kernel functions.

Although the RVR provides a promising performance in terms of both accuracy and sparsity, it only allows regression from multivariate inputs to a univariate output variable. In other words, the RVR is not capable of identifying multi-features for the degradation model. However, in the real applications, there may be more than one degradation feature. To overcome this drawback, a number of papers have been devoted to extending RVR into a multivariate form. Thayananthan et al. proposed a RVR to learn a one-to-many mapping from image features to state space for the pose ambiguity problem [18, 19]. This method is just a mixture of RVRs with different parameters, and it is essentially a univariate RVR method. Mohsenzadeh et al. therefore proposed a relevance sample-feature machine to perform a joint feature selection and classifier design simultaneously [20]. Further, Mohsenzadeh et al. proposed an incremental relevance sample-feature machine for high computational cost of large training sets [21]. Nevertheless, there is also a strong limitation that the weight is designed to be separable with respect to the parameters that determines the relevance samples and a parameter and the relevance features in the data set, respectively.

Motivated by the aforementioned considerations, we propose a Multivariate RVR (MRVR) approach, in which the weight matrix is inducted by a matrix Gaussian distribution instead of separating into a vector distribution. To address the RUL prediction issue under the dynamic degradation path, a MRVR model is firstly constructed by extending the classical RVR approach to a multivariate one, which is also a multi-step model considering the dynamic characteristics of the degradation process. Then, the hyperparameters of the MRVR model are estimated by Nesterov's Accelerated Gradient (NAG) method to obtain numerical solutions. Afterwards, the degradation path is forecasted based on the estimated hyperparameters. Concerning the forecasted path, the RUL is predicted by the First Hitting Time (FHT) approach. The major contributions and novelty are summarized as follows.

1) A degradation path based RUL framework is constructed through a multi-step dynamic MRVR model for the dynamic and multivariate degradation process. Different from the existing RVR method, which is a static regression model, our proposed model can describe the relationship between time-related future features and historical samples of features. The degradation tendency can therefore be monitored in advance under this RUL framework; and
2) The existing RVR is extended into a multivariate model, in which the weight matrix obeys a matrix Gaussian distribution. Due to the computational complexity of the matrix distribution, the analytical solutions of the hyperparameters of MRVR cannot be obtained with commonly used methods. The evidence function is deduced step by step for hyperparameters estimation.

The remainder of this paper is arranged as follows. Section II establishes a dynamic MRVR. Section III estimates the hyperparameters of MRVR using the NAG method. Section IV provides the degradation path and RUL prediction methods. Section V uses a case study to demonstrate the feasibility and effectiveness of the proposed algorithm. Finally, Section VII concludes the paper and proposes future work.

II. Dynamic Modeling based on Multi-step MRVR

Given an observed degradation series $\boldsymbol{x}_{1:N} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ along with a series of time points $\{t_1, t_2, \cdots, t_N\}$, where $\boldsymbol{x}_n = [x_{1,n}, \cdots, x_{M,n}]^{\mathrm{T}} \in \mathbb{R}^M (n = 1, 2, \cdots, N)$ is a multivariate vector, M represents the dimension of the multivariate space, and N denotes the sample size. A dynamic model can be established via a multi-step MRVR as

$$\boldsymbol{x}_{n+l} = \boldsymbol{W}\boldsymbol{\phi}(\boldsymbol{x}_n) + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{x}_{n+l} = [x_{1,n+l}, \cdots, x_{M,n+l}]^{\mathrm{T}} \in \mathbb{R}^M$ represents the $l-$step prediction vector, and $1 < n + l \leq N$; $\boldsymbol{\phi}(\boldsymbol{x}_n) = [\mathbf{1}, \mathcal{K}(\boldsymbol{x}_n, \boldsymbol{x}_1), \cdots, \mathcal{K}(\boldsymbol{x}_n, \boldsymbol{x}_N)] \in \mathbb{R}^{N+1}$ denotes a design vector, in which $\mathcal{K}(\boldsymbol{x}_n, \boldsymbol{x}_j) \in \mathbb{R}$ is a kernel function between the vector $\boldsymbol{x}_n$ and $\boldsymbol{x}_j (j = 1, 2, \cdots, N)$, $\boldsymbol{W} \in \mathbb{R}^{M \times (N+1)}$ is a weight matrix of the design vector $\boldsymbol{\phi}(\boldsymbol{x}_n) \triangleq \boldsymbol{\phi}$, and $\boldsymbol{\epsilon}$ is assumed to be a Gaussian distributed random error vector with the zero mean and a diagonal covariance matrix $\boldsymbol{\Sigma}_0 = diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\} \in \mathbb{R}^{M \times M}$, and $diag(\cdot)$ denotes a diagonal matrix.

A kernel function projects the input features into a higher dimensional space, by which the model becomes a linear regression model. Commonly used kernels include the linear kernel, the polynomial kernel, the Gaussian kernel, and the S-type kernel [22]. As the Gaussian kernel owns a strong generalization ability, it is adopted to construct the basis function $\boldsymbol{\phi}(\boldsymbol{x}_n)$ in this study.

The classical RVR is inherently a static univariate algorithm, which may reduce model accuracy and cannot capture the evolution of degradation features. Additionally, a univariate RVR is often not powerful enough to describe the behavior of engineering systems. The operation of many engineering systems is influenced by multiple variables. For example, current and voltage are indispensable for electrical systems. Hence, this study develops a multivariate dynamic model that describes the relationship between time-related future features and historical samples of features. As shown in (1), it is an extension of the classical RVR from a univariate static model to a multivariate dynamic model.

The Probability Density Function (PDF) of $\boldsymbol{x}_{n+l}$ conditioned on $\boldsymbol{W}$ and $\boldsymbol{\Sigma}_0$ can be written by

$$p(\boldsymbol{x}_{n+l}|\boldsymbol{W}, \boldsymbol{\Sigma}_0) = (2\pi)^{-\frac{M}{2}} |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}(\boldsymbol{x}_{n+l} - \boldsymbol{W}\boldsymbol{\phi})^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{x}_{n+l} - \boldsymbol{W}\boldsymbol{\phi})\right), \tag{2}$$

where $|\cdot|$ is the determinant of a square matrix. To avoid the over-fitting problem of model (1), a prior matrix Gaussian distribution is assigned on the $M \times (N+1)$ dimension weight matrix $\boldsymbol{W}$, which is denoted as $\boldsymbol{W} \sim \mathcal{MN}_{M,N+1}(\mathbf{0}, \boldsymbol{\Psi}, \boldsymbol{\Gamma})$. It suggests the random matrix $\boldsymbol{W}$ is governed by the zero mean matrix and variance matrix $\boldsymbol{\Psi} = diag(\Psi_1, \Psi_2, \cdots, \Psi_M) \in$

$\mathbb{R}^{M\times M}$, $\boldsymbol{\Gamma} = diag(\Gamma_1, \Gamma_2, \cdots, \Gamma_{N+1}) \in \mathbb{R}^{(N+1)\times(N+1)}$. Then we have

$$p(\boldsymbol{W}|\boldsymbol{\Psi},\boldsymbol{\Gamma}) = (2\pi)^{-\frac{M(N+1)}{2}}|\boldsymbol{\Psi}|^{-\frac{N+1}{2}}|\boldsymbol{\Gamma}|^{-\frac{M}{2}} \times \operatorname{etr}\left(-\tfrac{1}{2}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{W}\right), \tag{3}$$

where $\operatorname{etr}(\cdot)$ is the exponential function of the trace of the matrix.

The matrix Gaussian can be converted into a multivariate Gaussian form by Lemma 1.

Lemma 1. *(Vectorizable [23]) $\boldsymbol{W}$ obeys a $M\times(N+1)$-matrix Gaussian distribution, i.e. $\boldsymbol{W} \sim \mathcal{MN}(0, \boldsymbol{\Psi}, \boldsymbol{\Gamma})$ if and only if $\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)$ obeys a $M(N+1)$-variate Gaussian distribution, i.e. $\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right) \sim \mathcal{N}_{M(N+1)}\left(0, \boldsymbol{\Psi}\otimes\boldsymbol{\Gamma}\right)$, where $\operatorname{vec}(\cdot)$ is the vector operator and $\otimes$ is the Kronecker product (or tensor product).*

On the basis of Lemma 1, the posterior PDF of $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ is given in (4) and its detailed derivation process is provided in Appendix A.

$$p\left(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+1},\boldsymbol{\Psi},\boldsymbol{\Gamma}\right) = (2\pi)^{-\frac{M(N+1)}{2}}|\boldsymbol{\Sigma}^{-1}| \times \exp\left(-\frac{1}{2}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}) - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}) - \boldsymbol{\mu})\right), \tag{4}$$

where the mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ are

$$\boldsymbol{\mu} = \operatorname{vec}\left(\boldsymbol{\Gamma}\boldsymbol{\phi}\boldsymbol{x}_{n+l}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}\right) + \operatorname{vec}\left((\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}})^{-1}\boldsymbol{\phi}\boldsymbol{x}_{n+l}^{\mathrm{T}}\right), \tag{5}$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi}\otimes\boldsymbol{\Gamma} + \boldsymbol{\Sigma}_0 \otimes (\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}})^{-1} \tag{6}$$

respectively. According to the mean of $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$, conditioned on $\boldsymbol{x}_{n+l}$ and shown in (5), the mean of $p(\boldsymbol{W}|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0)$ is obtained by,

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\Psi}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{x}_{n+l}\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Gamma} + \boldsymbol{x}_{n+l}\boldsymbol{\phi}^{\mathrm{T}}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}. \tag{7}$$

According to [24], the matrix $\boldsymbol{\Sigma}$ in (6) can be decomposed into a Kronecker product of two matrices, the covariance matrices of $\boldsymbol{W}$ under the condition of $\boldsymbol{x}_{n+l}$ are obtained, which is denoted by $\tilde{\boldsymbol{\Psi}} \in \mathbb{R}^{M\times M}$ and $\tilde{\boldsymbol{\Gamma}} \in \mathbb{R}^{(N+1)\times(N+1)}$. Then, the posterior distribution of weight matrix $\boldsymbol{W}$ is matrix Gaussian, and its PDF is formulated in the following form.

$$p(\boldsymbol{W}|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0) = (2\pi)^{-\frac{M(N+1)}{2}}|\tilde{\boldsymbol{\Psi}}|^{-\frac{N+1}{2}} \times|\tilde{\boldsymbol{\Gamma}}|^{-\frac{M}{2}}\operatorname{etr}\left(-\tfrac{1}{2}\tilde{\boldsymbol{\Gamma}}^{-1}(\boldsymbol{W}-\tilde{\boldsymbol{\mu}})^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}(\boldsymbol{W}-\tilde{\boldsymbol{\mu}})\right). \tag{8}$$

III. Parameter Estimation

The hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$ should be estimated to make the multi-step dynamic model (1) available. An evidence function is evaluated firstly by the Bayes' theorem. The hyperparameters are then estimated via the NAG method.

A. Evaluation of the Evidence Function

Theoretically, when a new input vector $\boldsymbol{x}_k$ ($k > N$) is available, the distribution of the predicted $\boldsymbol{x}_{k+l}$, based on the former prediction $\boldsymbol{x}_{n+l}$, can be obtained by

$$p\left(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l}\right) = \iiiint p\left(\boldsymbol{x}_{k+l}|\boldsymbol{W},\boldsymbol{\Sigma}_0\right)p\left(\boldsymbol{W}|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right) \times p\left(\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0|\boldsymbol{x}_{n+l}\right)\mathrm{d}\boldsymbol{W}\mathrm{d}\boldsymbol{\Psi}\mathrm{d}\boldsymbol{\Gamma}\mathrm{d}\boldsymbol{\Sigma}_0. \tag{9}$$

Although we can integrate analytically over either $\boldsymbol{W}$ or the hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$, the complete marginalization over all of these variables is analytically intractable. Here, the evidence approximation is applied in which the hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$ are set to specific values by maximizing the marginal likelihood function that integrates over the parameters $\boldsymbol{W}$.

From Bayes' theorem, the posterior distribution for $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$ is given by

$$p\left(\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0|\boldsymbol{x}_{n+l}\right) \propto p\left(\boldsymbol{x}_{n+l}|\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right)p\left(\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right). \tag{10}$$

If the prior is relatively flat, then the hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$ can be estimated by maximizing the marginal likelihood function $p\left(\boldsymbol{x}_{n+l}|\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right)$ in the evidence framework. The marginal likelihood function $p\left(\boldsymbol{x}_{n+l}|\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right)$ can be obtained by integrating over the weight parameters $\boldsymbol{W}$ as

$$\begin{aligned} p\left(\boldsymbol{x}_{n+l}|\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right) &= \int p\left(\boldsymbol{x}_{n+l}|\boldsymbol{W},\boldsymbol{\Sigma}_0\right)p(\boldsymbol{W}|\boldsymbol{\Psi},\boldsymbol{\Gamma})\mathrm{d}\boldsymbol{W} \\ &= \int p\left(\boldsymbol{x}_{n+l}|\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right),\boldsymbol{\Sigma}_0\right)p(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)|\boldsymbol{\Psi},\boldsymbol{\Gamma})\mathrm{d}\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right). \end{aligned} \tag{11}$$

The integral will then be evaluated by completing the square in the exponent and making use of the standard form for the normalization coefficient of the Gaussian distribution. From (2) and (3), the evidence function (11) can be written as

$$p\left(\boldsymbol{x}_{n+l}|\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right) = (2\pi)^{-\frac{M(N+2)}{2}}|\boldsymbol{\Sigma}_0|^{-\frac{1}{2}}|\boldsymbol{\Psi}|^{-\frac{N+1}{2}}|\boldsymbol{\Gamma}|^{-\frac{M}{2}} \times \int \exp[-E(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right))]\mathrm{d}\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right), \tag{12}$$

where $E(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right))$ is given by

$$E\left(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)\right) = \frac{1}{2}\Big((\boldsymbol{x}_{n+l}-\boldsymbol{W}\boldsymbol{\phi})^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}_{n+l}-\boldsymbol{W}\boldsymbol{\phi}) + \left(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)\right)^{\mathrm{T}}(\boldsymbol{\Psi}\otimes\boldsymbol{\Gamma})^{-1}\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)\Big). \tag{13}$$

Then, the quadratic form of $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ is given as

$$E(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})) = \frac{1}{2}\left(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{A}\left(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)-\boldsymbol{\mu}\right)+E(\boldsymbol{\mu}), \tag{14}$$

where $\boldsymbol{\mu}$ is the mean of the posterior distribution of the $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ and is provided in (5) and $\boldsymbol{A} = \left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right) + (\boldsymbol{\Psi}\otimes\boldsymbol{\Gamma})^{-1}$ happens to be the precision of $p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0)$. One can therefore have

$$E(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_0^{-1} - \left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}^{\mathrm{T}}\right)\boldsymbol{A}^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}\right)\right)\boldsymbol{x}_{n+l}. \tag{15}$$

The integral over $\boldsymbol{W}$ can be evaluated by converting it to the

standard multivariate Gaussian, giving

$$\begin{aligned}&\int \exp\left(-E(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}}))\right)\mathrm{d}\,\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}}) = \exp\left(-E\left(\boldsymbol{\mu}\right)\right)\\ &\times\int \exp\left(-\frac{1}{2}\left(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{A}\left(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})-\boldsymbol{\mu}\right)\right)\mathrm{d}\,\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})\\ &=\exp\left(-E(\boldsymbol{\mu})\right)(2\pi)^{\frac{M(N+1)}{2}}|\boldsymbol{A}|^{-\frac{1}{2}}.\end{aligned}\tag{16}$$

Using (12), the negative log of the marginal likelihood is acquired by

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0\right) &= \frac{1}{2}\ln|\boldsymbol{\Sigma}_0| + \frac{N+1}{2}\ln|\boldsymbol{\Psi}| + \frac{M}{2}\ln|\boldsymbol{\Gamma}|\\ &+\frac{1}{2}\ln|\boldsymbol{A}| + E(\boldsymbol{\mu}) + \frac{M}{2}\ln(2\pi),\end{aligned}\tag{17}$$

which is the required expression of the evidence function.

Algorithm 1: Hyperparameter $\Psi_i, (i = 1, 2, \cdots, M)$ Estimation by NAG Method

Input: initial learning rate $\alpha_0 > 0$, momentum constant $u \in [0, 1]$, initial hyperparameter Ψ_{i0}, and the total number of iterations S.
Output: optimal hyperparameter Ψ_i^*
for $s = 1, 2, \cdots, S$ **do**
  $\alpha_s = \alpha_0 10^{-s/S}$,
  $v_s = uv_{s-1} - \alpha_s \nabla\mathcal{L}\left(\Psi_{i,s-1} + uv_{s-1}\right)$,
  $\Psi_{is} = \Psi_{i,s-1} + v_s$,
  where $\nabla\mathcal{L}\left(\Psi_{i,s-1} + uv_{s-1}\right)$ is calculated via replacing Ψ_i with $\Psi_{i,s-1} + uv_{s-1}$ in the first formula of (19).
end

B. Hyperparameters' Estimation by the NAG Method

Using the chain derivation rule of the multivariate composite function, the partial derivatives of $\ln|\boldsymbol{A}|$ and $E(\boldsymbol{\mu})$ to Ψ_i, σ_i^2 $(i = 1, 2, \cdots, M)$, and $\Gamma_j, (j = 1, 2, \cdots, N+1)$ are calculated as

$$\begin{aligned}\frac{\partial\ln|\boldsymbol{A}|}{\partial\Psi_i} &= -\mathrm{tr}\left(\boldsymbol{A}^{-1}\left(\left(\boldsymbol{\Psi}^{-2}\boldsymbol{E}_{ii}\right)\otimes\boldsymbol{\Gamma}^{-1}\right)\right),\\ \frac{\partial\ln|\boldsymbol{A}|}{\partial\Gamma_j} &= -\mathrm{tr}\left(\boldsymbol{A}^{-1}\left(\boldsymbol{\Psi}^{-1}\otimes\left(\boldsymbol{\Gamma}^{-2}\boldsymbol{E}_{jj}\right)\right)\right),\\ \frac{\partial\ln|\boldsymbol{A}|}{\partial\sigma_i^2} &= -\sigma_i^{-4}\,\mathrm{tr}\left(\boldsymbol{A}^{-1}\left(\boldsymbol{E}_{ii}\otimes\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\right)\right),\\ \frac{\partial E(\boldsymbol{\mu})}{\partial\Psi_i} &= \frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_0^{-2}\otimes\left(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}\right)\right)\boldsymbol{x}_{n+l}\,\mathrm{tr}\left(\boldsymbol{E}_{ii}\otimes\boldsymbol{\Gamma}\right),\\ \frac{\partial E(\boldsymbol{\mu})}{\partial\Gamma_j} &= \frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_0^{-2}\otimes\left(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}\right)\right)\boldsymbol{x}_{n+l}\,\mathrm{tr}\left(\boldsymbol{\Psi}\otimes\boldsymbol{E}_{jj}\right),\\ \frac{\partial E(\boldsymbol{\mu})}{\partial\sigma_i^2} &= \frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}\left(-\sigma_i^{-4}\boldsymbol{E}_{ii}+2\sigma_i^{-4}\left(\boldsymbol{E}_{ii}\otimes\boldsymbol{\phi}^{\mathrm{T}}\right)A^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}\right)\right.\\ &\left.-\left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}^{\mathrm{T}}\right)\left(\boldsymbol{E}_{ii}\otimes\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\right)\left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}\right)\right)\boldsymbol{x}_{n+l},\end{aligned}\tag{18}$$

respectively, where $\boldsymbol{E}_{ii}$ is a matrix, in which the (i,i)th element is 1, and zeros elsewhere. The derivatives of the negative log marginal likelihood $\mathcal{L}$ with respect to hyperparameters Ψ_i, Γ_j and σ_i^2 are as follows:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\Psi_i} &= \frac{N+1}{2\Psi_i} + \frac{1}{2}\frac{\partial\ln|\boldsymbol{A}|}{\partial\Psi_i} + \frac{\partial E(\boldsymbol{\mu})}{\partial\Psi_i},\\ \frac{\partial\mathcal{L}}{\partial\Gamma_j} &= \frac{M}{2\Gamma_j} + \frac{1}{2}\frac{\partial\ln|\boldsymbol{A}|}{\partial\Gamma_j} + \frac{\partial E(\boldsymbol{\mu})}{\partial\Gamma_j},\\ \frac{\partial\mathcal{L}}{\partial\sigma_i^2} &= \frac{1}{2\sigma_i^2} + \frac{1}{2}\frac{\partial\ln|\boldsymbol{A}|}{\partial\sigma_i^2} + \frac{\partial E(\boldsymbol{\mu})}{\partial\sigma_i^2}.\end{aligned}\tag{19}$$

It is noteworthy that by setting the left side of each equation in (19) to 0, it is difficult to obtain explicit solutions of the hyperparameters Ψ_i, Γ_j and σ_i^2, and widely used re-estimation methods also seem to be unreliable. So, the NAG method is used to obtain numerical solutions of the hyperparameters in this study [25]. An algorithm, i.e., Algorithm 1, is shown in the following.

NAG prescribes a particular formula for the learning rate α and the momentum constant u. Usually, the learning rate is set large enough to ensure a fast convergence rate at the beginning, and then slowly decays to ensure that optimal stable points can be reached. So an exponential decay function $\alpha_s = \alpha_0 10^{-s/S}$ is chosen in this paper, where α_0 is an initial value of the learning rate, where s denotes an iteration variable and S represents the total number of iterations. The momentum constant u ($\in [0, 1]$) controls the "decay" of the velocity vector v. A higher value of u makes the gradient change in a quicker way. So u usually takes a value close to 1, for example, 0.9.

Correspondingly, the hyperparameters Γ_j $(j = 1, 2, \cdots, N+1)$ and σ_i^2 $(i = 1, 2, \cdots, M)$ can be estimated by following analogous procedures of the Algorithm 1. Due to their similarity, no further discussion is presented here.

IV. Degradation Path and RUL Prediction

A. The Predicted PDF of Degradation Path

With the estimated hyperparameters $\boldsymbol{\Psi}^*$, $\boldsymbol{\Gamma}^*$ and $\boldsymbol{\Sigma}_0^*$, when a new vector $\boldsymbol{x}_k$ is available, the PDF of $l-$step prediction $\boldsymbol{x}_{k+l}$ based on the historical data can be calculated by

$$\begin{aligned}p\left(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l}\right) &= \int p\left(\boldsymbol{x}_{k+l}|\boldsymbol{W},\boldsymbol{\Sigma}_0^*\right)p\left(\boldsymbol{W}|\boldsymbol{x}_{n+l},\boldsymbol{\Psi}^*,\boldsymbol{\Gamma}^*\right)\mathrm{d}\boldsymbol{W}\\ &= \int p\big(\boldsymbol{x}_{k+l}|\,\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}}),\boldsymbol{\Sigma}_0^*\big)\,p\big(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi}^*,\boldsymbol{\Gamma}^*\big)\,\mathrm{dvec}(\boldsymbol{W}^{\mathrm{T}})\end{aligned}\tag{20}$$

The conditional PDF $p\left(\boldsymbol{x}_{k+l}|\,\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}}),\boldsymbol{\Sigma}_0^*\right)$ and the posterior weight PDF are given by replacing the estimated hyperparameters $\boldsymbol{\Psi}^*$, $\boldsymbol{\Gamma}^*$ and $\boldsymbol{\Sigma}_0^*$ into (2) and (4), respectively. It is obvious that (20) involves a convolution of two Gaussian distributions, and it can be regarded as a marginal PDF by taking $(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l})$ and $\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})$ as random variables.

Following a similar procedure to that shown in Appendix A, the joint PDF can be obtained for the random variables $(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l})$ and $\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})$. Then, the marginal PDF is easily obtained from the partitioned mean and covariance matrices of the joint distribution. The PDF of the degradation prediction takes the form

$$p\left(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l}\right) = \mathcal{N}\left(\boldsymbol{x}_{k+l}|\boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_k))\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},\boldsymbol{\Sigma}_k\right),\tag{21}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are presented by (5) and (6), respectively. The covariance $\boldsymbol{\Sigma}_k$ of the predictive distribution is given by

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_k)\right) \boldsymbol{\Lambda} \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\phi}(\boldsymbol{x}_k)\right) \boldsymbol{\Sigma}_0, \tag{22}$$

in which

$$\boldsymbol{\Lambda} = \left(\boldsymbol{\Sigma}^{-1} + \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\phi}(\boldsymbol{x}_k)\boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_k)\right) - \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\phi}(\boldsymbol{x}_k)\right) \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\phi}^{\mathrm{T}}(\boldsymbol{x}_k)\right)\right)^{-1}.$$

The first term in (22) represents the noise on the data whereas the second term reflects the uncertainty associated with the parameters $\boldsymbol{W}$. Because the noise process and $\boldsymbol{W}$ are independent and both are Gaussian, their covariances are additive.

B. RUL Prediction

As the FHT describes the relationship between the time and the degradation path, it is used for RUL prediction with the estimated degradation path. Given the observed measurements $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$, the RUL of each variable at time t_k is defined by

$$L_{ik} \triangleq L_i(t_k) = \inf\{t_l : x_i(t_k + t_l) \equiv x_{i,k+l} \in \mathcal{B}_i | \boldsymbol{x}_{1:k}\}, \tag{23}$$

where $\inf\{\cdot\}$ denotes the infimum of a discrete set in this study; t_l represents the time length of the multi-step prediction; $x_i(t_k + t_l)$ represents the degradation path at time $t_k + t_l$, which is denoted by $x_{i,k+l}$ and forecasted by the mean of (21); $\boldsymbol{x}_{1:k}$ denotes the historical measurements from t_1 to t_k; and $\mathcal{B}_i$ refers to a boundary set (i.e., threshold set), containing a boundary, barrier, or failure threshold, which is usually determined by empirical knowledge [26].

The FHT is the time when the degradation path first hits the boundary set $\mathcal{B}_i$, which defines a stopping condition for the degradation process. With (23), the predicted PDF of the RUL can be completely derived according to [27], in which the mean of the RUL is obtained by

$$E_i(t_k) = \sum_{L_{ik}=0}^{+\infty} L_{ik} \cdot p_i(L_{ik}), \tag{24}$$

where

$$p_i(L_{ik}) = \frac{\phi(g_{i,k+l})\Delta g_{i,k+l}}{1 - \Phi(g_{i,k})}, \tag{25}$$

$\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and cumulative distribution function of a standard normal random variable, respectively; $g_{i,k+l} = (\tilde{\mu}_{i,k+l} - H_i)/\sqrt{\tilde{\sigma}^2_{i,k+l}}$, $\Delta g_{i,k+l}$ represents the deviation of $g_{i,k+l}$ on t_{k+l}, where $H_i \in \mathcal{B}_i$ is a failure threshold of the ith-variate; $\tilde{\mu}_{i,k+l}$ is the mean of the predicted $x_{i,k+l}$; and $\tilde{\sigma}^2_{k+l}$ denotes the variance of $x_{i,k+l}$. $\tilde{\mu}_{i,k+l}$ and $\tilde{\sigma}^2_{k+l}$ are obtained by extracting the ith term of the mean and variance from (21), respectively.

According to (24) and the relationship between variance and mean, the standard deviation of RUL is obtained by $\sigma_i(t_k) = \sqrt{E_i(t_k^2) - E_i^2(t_k)}$, where $E_i(t_k^2)$ denotes the mean of L_{ik}^2, which can be obtained by replacing L_{ik} with L_{ik}^2 in (24). Then, the lower and upper RUL uncertainty bounds at time t_k, $D(t_k)$ and $U(t_k)$, for each variable $x_{i,k}$, are estimated by the 3σ-criterion as follows,

$$D_i(t_k) = E_i(t_k) - 3\sigma_i(t_k) \tag{26}$$

$$U_i(t_k) = E_i(t_k) + 3\sigma_i(t_k) \tag{27}$$

respectively.

As the RUL predictions seriously vary with the uncertainties between different variables, the RUL values are different for each variable. In order to reduce the effect of different variables to the RUL prediction, the RUL, the lower and upper bounds of the components are derived by averaging the RUL predictions and the bounds of each variable, which are given as below.

$$\widehat{RUL}(t_k) = \sum_{i=1}^{M} E_i(t_k)/M, \tag{28}$$

$$D(t_k) = \sum_{i=1}^{M} D_i(t_k)/M, \tag{29}$$

and

$$U(t_k) = \sum_{i=1}^{M} U_i(t_k)/M, \tag{30}$$

where $\widehat{RUL}(t_k)$, $D(t_k)$, and $U(t_k)$ are the estimates of the RUL, the lower and upper bounds of components at time t_k, respectively.

Finally, an algorithm is proposed to show the entire procedure of estimating the model, the degradation path and the RUL prediction in Algorithm 2, respectively.

V. Case Study I

A. Platform Introduction and Feature Selection

In this section, an experimental platform, developed by the CRRC Zhuzhou Institute and the Central South University, China, is applied to validate the proposed MRVR method [28, 29]. As shown in Fig. 1, the hardware-in-the-loop platform chiefly includes a Traction Control Unit (TCU), a dSPACE real-time simulator, a signal conditioner, a host PC, and a power source. Briefly, fault injection algorithms, as well as control programs of the rectifier-side and the inverter-side, are loaded into the TCU. The whole methodology is integrated into the dSPACE simulator. The signal conditioner converts signals between the TCU and the dSPACE simulator. The host PC controls the running times of the system and monitors the sensor waveforms from the TCU and simulator.

DC-link capacitors are useful for maintaining the stability of voltages for the traction converter, whose electrical diagram is shown by Fig. 2. According to [30], a performance degradation is simulated by an exponential decay function shown in the following formula.

$$C_{\text{deg}} = \begin{cases} C & 0 \le t \le t_{\text{deg}} \\ C \cdot e^{-a(t - t_{\text{deg}})} & t \ge t_{\text{deg}} \end{cases} \tag{31}$$

where C_{degra} is the degraded capacitance value; C is the nominal capacitance value; t is the simulation time; t_{degra} is the start time of degradation; and $a \in [0, 1]$ determines

Algorithm 2: The MRVM based Degradation Path and RUL Prediction

for *Train Process* **do**

Input: The observed degradation series $\boldsymbol{x}_{1:N} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$

Output: The estimated hyperparameters $\boldsymbol{\Psi}^*$, $\boldsymbol{\Gamma}^*$, $\boldsymbol{\Sigma}_0^*$

1. Establish the dynamic MRVM model as (1);
2. Introduce a prior matrix Gaussian distribution over $\boldsymbol{W}$ as (3);
3. Evaluation the evidence function as (17);
4. Estimate the unknown hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Sigma}_0$ by the NAG method referring the Algorithm 1.

end

for *Degradation Path Prediction* **do**

Input: The estimated hyperparameters $\boldsymbol{\Psi}^*$, $\boldsymbol{\Gamma}^*$, $\boldsymbol{\Sigma}_0^*$

Output: The PDF of $l-$step prediction $\boldsymbol{x}_{k+l}$

1. Obtain the joint PDF for the random variables $(\boldsymbol{x}_{k+l}|\boldsymbol{x}_{n+l})$ and $\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})$;
2. Partition mean and covariance matrices of the joint distribution and get the PDF of the degradation prediction as (21).

end

for *RUL Prediction* **do**

Input: The PDF of the degradation path

Output: The mean, the low and high bounds of RUL

1. Induct the RUL mean, the low and high bounds of each variate as (24), (26) and (27);
2. Obtain the RUL mean, the low and high bounds of the component as (28), (29) and (30).

end

the degree of degradation. In this study, the value of normal capacitance C is 4250μF. The sampling frequency is 2500Hz. The total simulation time t is set to 1.3s. The degradation start time t_{degra} is 0.3s. The degradation coefficient a is set to 0.01 according to the engineering experience. Loading the exponential decay function into the TCU, the up and down terminal voltages of the DC-link, the sum of the three-phase current of the inverter, and the electromagnetic torque of the motor are collected from the PC and shown in Fig. 3.

As seen in Fig. 3, the curves varying with time go through three stages: the features fluctuate steadily during their normal stage; then the features degenerate with changing tendencies during the degradation stage; finally the tendencies deteriorate until the system's self-protection is triggered, which suggests the end of life. Considering the degradation path over time, we further extract the peaks of the terminal voltages, the sum of the three-phase current, and the electromagnetic torque from the original data as the degradation features. The peak is calculated by $x_p = max(x)$ for each variable. The whole selected peaks are depicted in Fig. 4.

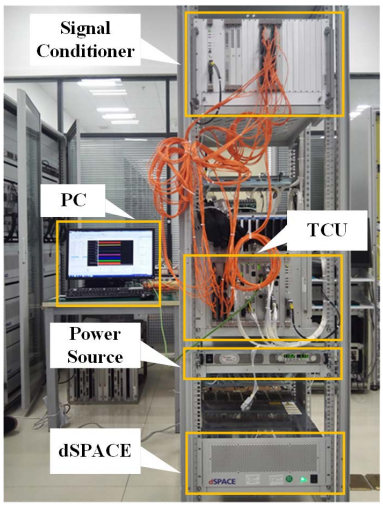


Fig. 1: The hardware-in-the-loop experimental platform.

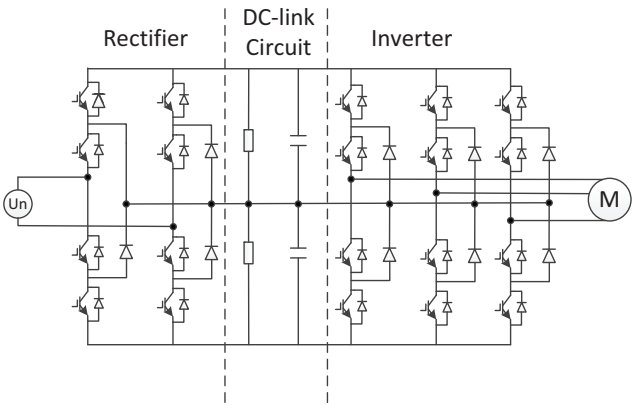
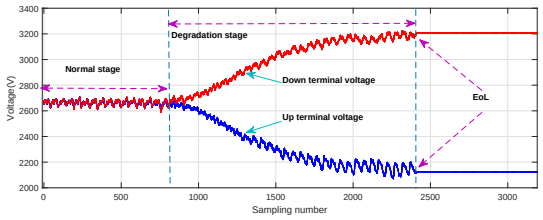


Fig. 2: The electrical diagram of the traction converter.

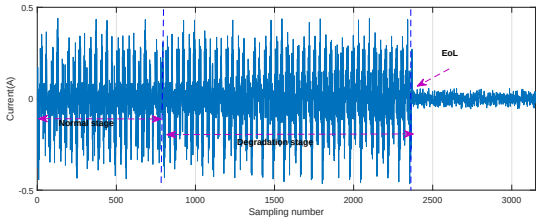
B. Degradation Modeling and Degradation Path Estimation

Starting from the 801st sampling point, the MRVR modeling is performed by identifying the unknown hyperparameters $\boldsymbol{\Psi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$, which are obtained by the NAG method proposed in III-B. Then, according to the method proposed in IV-A, the fitted model is extrapolated to estimate the propagation signals for monitoring the degradation status in advance.

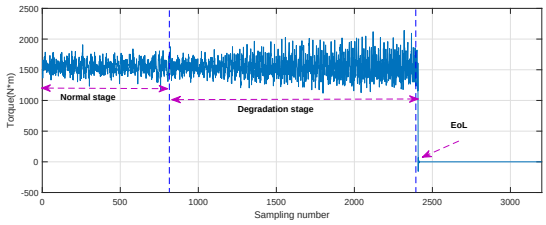
In order to thoroughly evaluate the performance of the proposed MRVR, more degradation prediction results are acquired from different training intervals (i.e. N's are different). The data measured from the 801st sampling point to the 1300th were selected as the training set of the proposed MRVR model, and then the 10-step (l in model (1)) ahead of the future degradation path is estimated according to (21). The data from the 801st sampling point to the 2000th of each feature is also used as a new training set for the comparison purpose of the MRVR method with different training intervals. The performance of MRVR in predicting the features' degradation paths is shown in Figs. 5-6. Moreover, performance indexes, the Mean Absolute Error (MAE) and Normalized Root Mean Relative Error (NRMSE), are adopted to evaluate the prediction performance with different training intervals under the



(a) The up and down terminal voltages of DC-link.

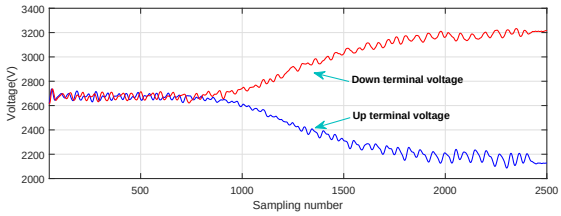


(b) The sum of three-phase currents of inverter.

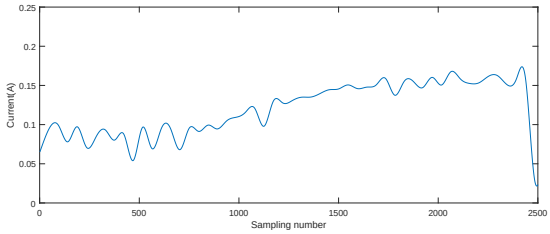


(c) The electromagnetic torque of motor.

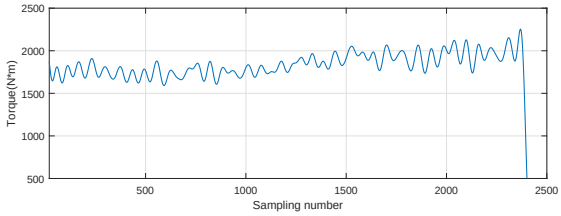
Fig. 3: Collected data from the experimental platform.



(a) The peaks of up and down terminal voltages.



(b) The peak of sum of three-phase currents.



(c) The peak of electromagnetic torque.

Fig. 4: The peaks selected from the collected data.

MRVR method [5]. As illustrated by the Table I, the accuracy of prediction varies with the amount of the training data in the sense that a large amount of training data lead to a higher prediction accuracy.

Further, the estimation results of degradation path based on MRVR are compared with the classical RVR to highlight the superiority of the proposed method. The RVR based degradation path estimation is performed on the same training set. The forecast horizon as the MRVR method and the estimated results are exhibited in Figs. 5-6. It is observed from Fig. 5 that the RVR based trained values cannot fit the actual ones well when the amount of training data is small. With the increasing of the training data, as shown in Fig. 6, the predicted degradation path based on the RVR becomes as accurate as the MRVR.

C. RUL Prediction and Prognostic Performance Evaluation

Starting from the 1701st sampling point, inspections of the degradation path of all features are made at predefined inspection times. The interval between two successive inspections is equal to 30 sampling points, to avoid too frequent and costly inspections of the component. At each inspection, MRVR regression is performed on the training data. The fitted model is then extrapolated to predict the times at which the degradations reach their thresholds. The threshold for the up and down terminal voltage, the sum of three-phase currents, and the torque are $2087V$, $3230V$, $0.1766A$, and $2254N*m$, respectively. The RUL, the low and high bounds of the RUL are then calculated by (28)-(30). The predicted results are shown in Fig. 7, with the comparative result based on the RVR method.

As illustrated by Fig. 7, the RUL prediction accuracy for the MRVR is apparently greater than the RVR, especially there is lag prediction for the RVR before the 2080th sampling point. This phenomenon is caused by the amount of the training data for training the degradation model, and the accuracy of RUL prediction extremely relies on the estimated degradation path. With the increasing of the training data, the predicted RUL is almost consistent with the actual one, both for the proposed MRVR and the classical RVR method.

Performance indexes, MAE and NRMSE, are adopted to evaluate the prediction performance between the MRVR and the RVR method. The performance indexes are presented in Table II. The performance indexes from Table II once again demonstrate that the prediction accuracy of the MRVR is better than that of the RVR.

VI. Case Study II-Public Prognostic Bearing Datasets

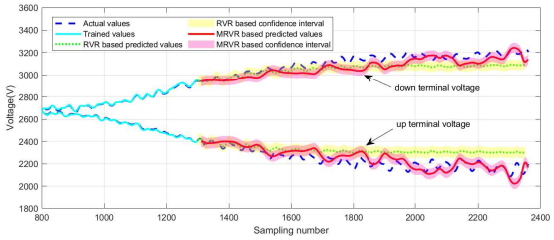
A. Data Description

A group of public bearing datasets, i.e., XJTU-SY bearing datasets, are used to demonstrate our proposed approach [31].

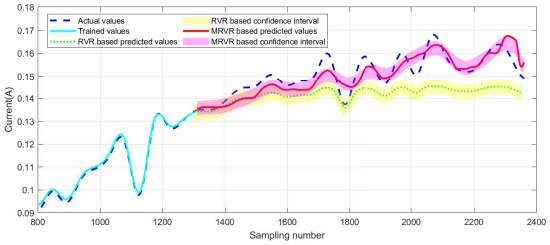
TABLE I: The performance metric comparisons with different training intervals under the MRVR method

Performance Metric	$801st-1300th$				$801st-2000th$			
	$Uu(V)$	$Ud(V)$	$I(A)$	$T(N*m)$	$Uu(V)$	$Ud(V)$	$I(A)$	$T(N*m)$
MAE	4.3207	0.7780	4.6721×10^{-6}	1.6472	6.7970×10^{-4}	1.1628×10^{-4}	1.7182×10^{-9}	2.6296×10^{-4}
$NRMSE(\%)$	0.1967	0.0325	0.0051	0.1113	3.7127×10^{-5}	5.3427×10^{-6}	1.4818×10^{-6}	1.8809×10^{-5}

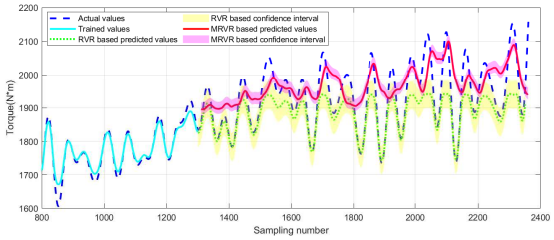
* Uu: Up terminal voltage; Ud: Down terminal voltage; I: the sum of three-phase currents; T: Electromagnetic torque.



(a) The estimated voltages of RVR and MRVR method.

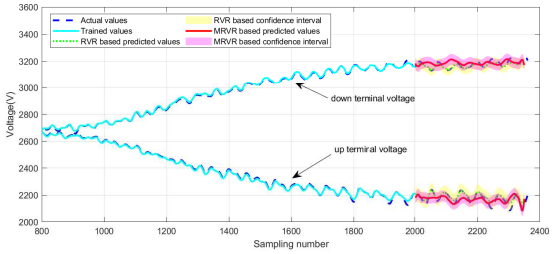


(b) The estimated current of RVR and MRVR method.

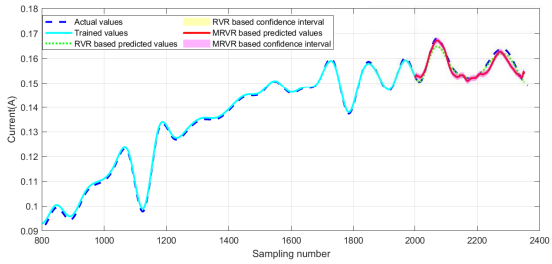


(c) The estimated electromagnetic torque of RVR and MRVR method.

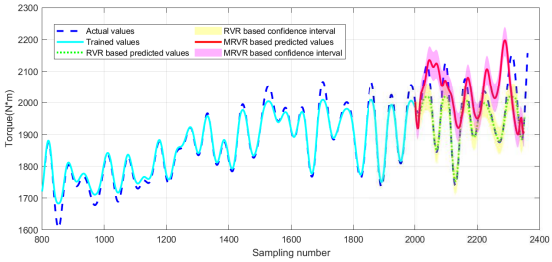
Fig. 5: The estimated degradation path of the RVR and the MRVR method under the 500 training data.



(a) The estimated voltages of RVR and MRVR method.



(b) The estimated current of RVR and MRVR method.



(c) The estimated electromagnetic torque of RVR and MRVR method.

Fig. 6: The estimated degradation path of the RVR and the MRVR method under the 1200 training data.

TABLE II: The performance metric comparisons between the MRVR and the RVR

Performance metric	MRVR	RVR
MAE	13.8804	24.6257
NRMSE(%)	3.4113	5.8745

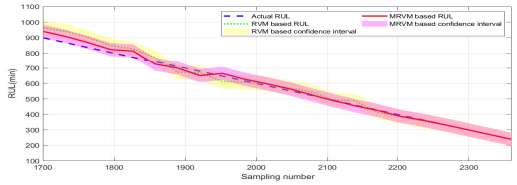
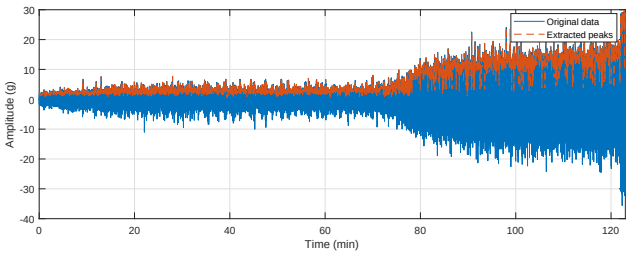
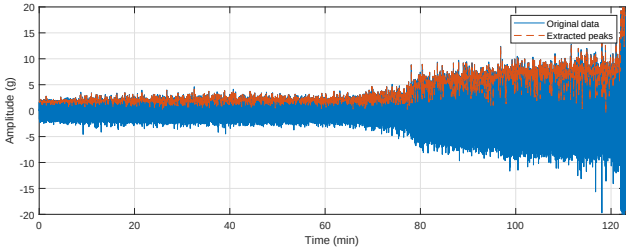


Fig. 7: The comparison of predicted RUL between the RVR and the MRVR method.

The XJTU-SY bearing datasets are provided by the Xi'an Jiaotong University (XJTU) and the Changxing Sumyoung Technology Co., Ltd. (SY), Zhejiang, China. Two accelerometers are placed on the bearings and positioned at $90°$ to each other: one is placed on the vertical axis and the other one on the horizontal axis. Fig. 8 shows the horizontal and vertical vibration signals of one bearing during the whole operating life, in which the blue solid line illustrates the raw data and the red dotted line illustrates the peaks extracted from the raw data.



(a) The horizontal vibration signals.



(b) The vertical vibration signals.

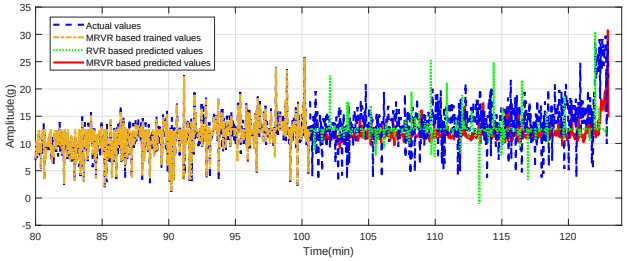
Fig. 8: The data extracted from the bearing.

It can be seen from Fig. 8 that the complete bearing degradation process is comprised of two different stages, i.e., the normal operating stage and the degradation stage. The vibration signals in the normal operating stage only present random fluctuations at a low level whereas in the degradation stage they show an increasing trend over operating time. In this paper, for the MRVR modeling, the degradation path and RUL prediction, the horizontal and vertical data are collected after the system has operated for 78 minutes. The bearings are stopped when the amplitudes of the horizontal and vertical vibration signals are higher than $25g$ and $15g$, respectively.

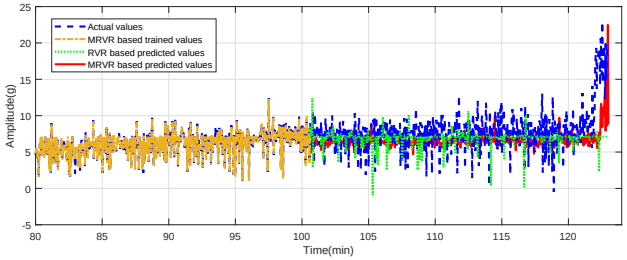
B. Degradation Path Prediction

The proposed MRVR model (1) is trained by selecting the extracted peaks during $80 - 100min$ as inputs and $80.6 - 100.6min$ data as outputs and following the training process of Algorithm 2. Then, with this trained MRVR model, degradation tendencies of the horizontal and vertical signals are predicted based on the mean of the random variable following the distribution shown in Eq (21), which are shown in Fig. 9.

Moreover, the predicted results of the degradation path based on MRVR are compared with the one based on RVR, which is performed on the same training set. The results from the RVR are also exhibited in Fig. 9. As illustrated by Fig. 9(a), the predicted degradation path based on the RVR exceeds its failure threshold $25g$ before that based on the MRVR. As shown in Fig. 9(b), however, the predicted degradation path based on the RVR cannot follow the actual vertical amplitude so well as that based on the MRVR. This confirms that the MRVR outperforms the RVR.



(a) The predicted amplitudes of the horizontal signal based on the RVR and MRVR.



(b) The predicted amplitudes of the vertical signal based on the RVR and MRVR.

Fig. 9: The predicted degradation path based on the RVR and the MRVR.

C. RUL Prediction

With the predicted amplitudes of the horizontal and vertical signals, the RUL of the bearing is predicted by the FHT method proposed in Section IV.B in this paper. It is assumed that the bearing's RUL is inspected every $2min$. Then, the RUL is predicted by Eq. (28) and compared with that based the RVR method, which are shown in Fig 10.

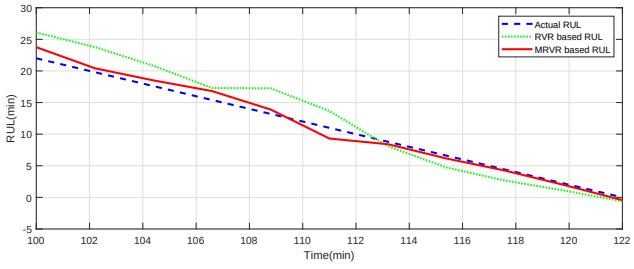


Fig. 10: The comparison of predicted RUL between the RVR and the MRVR methods for bearing.

The performance metrics, MAE and NRMSE, are adopted to evaluate the prediction performance between the MRVR and the RVR. The comparison results are presented in Table 1.

TABLE III: The performance metric comparisons between the MRVR and the RVR for bearing

Performance metric	MRVR	RVR
MAE(min)	0.7960	2.3594
NRMSE(%)	8.7793	24.3193

Both the predicted values in Fig. 10 and the performance indexes in Table III demonstrate that the prediction results based on the MRVR outperform those based on the RVR.

VII. Conclusion

In this study, a degradation path based RUL framework was constructed by a dynamic MRVR model. First of all, a multi-step regression model was established for describing the degradation dynamics. Then, the regression model was extended into a MRVR one by introducing a matrix Gaussian distribution into the classical RVR approach, wherein the hyperparameters were estimated by the Nesterov's accelerated gradient method to avoid tricky analytical solutions. Based on the degradation path forecasted by the MRVR approach, the RUL was predicted through the FHT method. Finally, the proposed schemes were demonstrated by a case study, which investigated the capacitors' performance degradation in the traction systems of the high-speed trains.

The proposed MRVR approach is intrinsically a top-down approach, which begins with all of the training samples and then prunes out the irrelevance samples in each iterative. Our future work will focus on an incremental version of MRVR in order to reduce the computational complexity caused by a large number of data. Moreover, robustness is an important property for regression problems. The proposed MRVR is under a Gaussian distributed framework, which owns a poor robust property. The MRVR method will be extended to a robustness form for the outlier regression problem in the future.

Appendix A
The Posterior Distribution of the Vectorized Weight

Firstly, the distribution of the joint of $\boldsymbol{x}_{n+l}$ and $\boldsymbol{W}$ needs establishing. To do this, we define

$$\boldsymbol{z} = \begin{pmatrix} \operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right) \\ \boldsymbol{x}_{n+l} \end{pmatrix}. \tag{32}$$

Then the log PDF of z is given by

$$\begin{aligned}
\ln p(\boldsymbol{z}) =& \ln p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{\Psi},\boldsymbol{\Gamma}) + \ln p(\boldsymbol{x}_{n+l}|\boldsymbol{W},\boldsymbol{\Sigma}_0) \\
=& -\frac{1}{2}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}))^{\mathrm{T}}(\boldsymbol{\Psi}\otimes\boldsymbol{\Gamma})^{-1}\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}) \\
& -\frac{1}{2}(\boldsymbol{x}_{n+l}-\boldsymbol{W}\boldsymbol{\phi})^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{x}_{n+l}-\boldsymbol{W}\boldsymbol{\phi}) + \text{const} \\
=& -\frac{1}{2}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}))^{\mathrm{T}}(\boldsymbol{\Psi}\otimes\boldsymbol{\Gamma})^{-1}\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right) \\
& -\frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{x}_{n+l} + \frac{1}{2}\boldsymbol{x}_{n+l}^{\mathrm{T}}(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}^{\mathrm{T}})\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}) \\
& +\frac{1}{2}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}))^{\mathrm{T}}(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi})\boldsymbol{x}_{n+l} \\
& -\frac{1}{2}\left(\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right)\right)^{\mathrm{T}}(\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}))\operatorname{vec}\left(\boldsymbol{W}^{\mathrm{T}}\right) + C
\end{aligned} \tag{33}$$

where 'C' denotes terms independent of $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ and $\boldsymbol{x}_{n+l}$. Since (33) is a quadratic function of the components of $\boldsymbol{z}$, the variable $\boldsymbol{z}$ is Gaussian distributed.

Next, an explicit expression for the conditional PDF $p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma})$ should be sought. An important property of the multivariate Gaussian distribution is that if two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is also Gaussian [22]. Since Gaussian distribution is completely characterized by its mean and its covariance, our goal will be to identify expressions for the mean and covariance of $p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma})$. Such problems can be solved straightforwardly by regarding the $\boldsymbol{x}_{n+l}$ in (33) as a constant and setting the coefficient of the second order term in $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ to the precision (inverse covariance) matrix $\boldsymbol{\Sigma}^{-1}$ and the coefficient of the linear term in $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ to $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, from which we can obtain mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. This method is called "completing the square".

So, consider the functional dependence of (33) on $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$ in which $\boldsymbol{x}_{n+l}$ is regarded as a constant. If all terms that are second order are picked out from $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$, there is

$$-\frac{1}{2}(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}))^{\mathrm{T}}((\boldsymbol{\Psi}\otimes\boldsymbol{\Gamma})^{-1}+\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}))\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}}) \tag{34}$$

from which, it can immediately conclude that the covariance (inverse precision) of $p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma},\boldsymbol{\Sigma}_0)$ is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi}\otimes\boldsymbol{\Gamma} + \boldsymbol{\Sigma}_0\otimes(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}})^{-1}. \tag{35}$$

Now consider all of the terms in (33) that are linear in $\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})$

$$\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi})\boldsymbol{x}_{n+l} \tag{36}$$

Then, the mean of $p(\operatorname{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+l},\boldsymbol{\Psi},\boldsymbol{\Gamma})$ is obtained as

$$\begin{aligned}
\boldsymbol{\mu} &= \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\phi}\right)\boldsymbol{x}_{n+l} \\
&= \left(\boldsymbol{\Psi}\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Gamma}\boldsymbol{\phi}\right)\boldsymbol{x}_{n+l} + \left(\boldsymbol{I}_M\otimes\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{\phi}\right)\boldsymbol{x}_{n+l} \\
&= \operatorname{vec}\left(\boldsymbol{\Gamma}\boldsymbol{\phi}\boldsymbol{x}_{n+l}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Psi}\right) + \operatorname{vec}\left((\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}})^{-1}\boldsymbol{\phi}\boldsymbol{x}_{n+l}^{\mathrm{T}}\right)
\end{aligned} \tag{37}$$

where $\boldsymbol{I}_M$ denotes the identity matrix of order M. And the

PDF of $\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})$ conditioned on $\boldsymbol{x}_{n+l}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Gamma}$ are

$$p\left(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})|\boldsymbol{x}_{n+1},\boldsymbol{\Psi},\boldsymbol{\Gamma}\right)=(2\pi)^{-\frac{M(N+1)}{2}}|\boldsymbol{\Sigma}^{-1}| \times \exp\left(-\frac{1}{2}(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathrm{vec}(\boldsymbol{W}^{\mathrm{T}})-\boldsymbol{\mu})\right) \quad (38)$$

REFERENCES

[1] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Hoboken, NJ, USA: Wiley, 2006.

[2] B. Y. Yang, R. N. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9521–9530, 2019.

[3] C. Cheng, G. J. Ma, Y. Zhang, M. Y. Sun, F. Teng, H. Ding, and Y. Yuan, "A deep learning-based remaining useful life prediction approach for bearings," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 3, pp. 1243–1254, 2020.

[4] Z. H. Chen, M. Wu, R. Zhao, F. Guretno, R. Q. Yan, and X. L. Li, "Machine remaining useful life prediction via an attention-based deep learning approach," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2521–2531, 2020.

[5] X. L. Wang, B. Jiang, S. X. Ding, N. Y. Lu, and Y. Li, "Extended relevance vector machine-based remaining useful life prediction for dc-link capacitor in high-speed train," *IEEE Transactions on Cybernetics*, DOI 10.1109/TCYB.2020.3035796, 2020, to be published.

[6] S. Meraghni, L. S. Terrissa, M. Yue, J. Ma, S. Jemei, and N. Zerhouni, "A data-driven digital-twin prognostics method for proton exchange membrane fuel cell remaining useful life prediction," *International Journal of Hydrogen Energy*, DOI https://doi.org/10.1016/j.ijhydene.2020.10.108, 2020, to be published.

[7] Y. G. Lei, N. P. Li, L. Guo, N. B. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems & Signal Processing*, vol. 104, pp. 799–834, 2018.

[8] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni, "Direct remaining useful life estimation based on support vector regression." *IEEE Trans. Industrial Electronics*, vol. 64, no. 3, pp. 2276–2285, 2017.

[9] L. Ma, J. Dong, K. X. Peng, and C. F. Zhang, "Hierarchical monitoring and root-cause diagnosis framework for key performance indicator-related multiple faults in process industries," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2091–2100, 2019.

[10] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 2, pp. 291–296, 2009.

[11] S. Yin, X. C. Xie, and W. Sun, "A nonlinear process monitoring approach with locally weighted learning of available data," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1507–1516, 2017.

[12] G. Wang, J. F. Jiao, and S. Yin, "Efficient nonlinear fault diagnosis based on kernel sample equivalent replacement," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2682–2690, 2019.

[13] X. L. Wang, B. Jiang, N. Y. Lu, and C. Y. Zhang, "Dynamic fault prognosis for multivariate degradation process," *Neurocomputing*, vol. 275, pp. 1112 – 1120, 2018.

[14] S. Yin, Y. C. Jiang, Y. Tian, and O. Kaynak, "A data-driven fuzzy information granulation approach for freight volume forecasting," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1447–1456, 2017.

[15] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.

[16] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proceedings of the Ninth International Workshop on Artificial Intelligence & Statistics*, pp. 1–13, 2003.

[17] M. E. Tipping, "The relevance vector machine," in *Advances in neural information processing systems*, pp. 652–658, 2000.

[18] A. Thayananthan, R. Navaratnam, B. Stenger, P. H. S. Torr, and R. Cipolla, "Multivariate relevance vector machines for tracking," in *9th European Conference on Computer Vision (ECCV 2006), Part III, Graz, Austria, May 7-13*, pp. 124–138, 2006.

[19] A. Thayananthan, R. Navaratnam, B. Stenger, P. H. S. Torr, and R. Cipolla, "Pose estimation and tracking using multivariate regression," *Pattern Recognition Letters*, vol. 29, pp. 1302–1310, 2008.

[20] Y. Mohsenzadeh, H. Sheikhzadeh, A. M. Reza, and N. Bathaee, "The relevance sample-feature machine: a sparse bayesian learning approach to joint feature-sample selection," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 2241–2254, 2013.

[21] Y. Mohsenzadeh, H. Sheikhzadeh, and S. Nazari, "Incremental relevance sample-feature machine: A fast marginal likelihood maximization approach for joint feature selection and classification." *Pattern Recognition*, vol. 60, pp. 835–848, 2016.

[22] C. M. Bishop, *Pattern recognition and machine learning*. Springer-Verlag New York, Inc., 2006.

[23] Z. X. Chen, B. Wang, and A. N. Gorban, "Multivariate Gaussian and student-t process regression for multi-output prediction," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3005–3028, 2020.

[24] D. H. Lin, "The approach to get the decomposition of the kronecker product of matrix," *Journal of Minjiang University*, vol. 28, no. 5, pp. 7–9, 2007.

[25] I. Sutskever, *Training recurrent neural networks*. University of Toronto, 2013.

[26] M.-L. T. Lee and G. A. Whitmore, "Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary," *Statistical Science*, vol. 21, no. 4, pp. 501–513, 2006.

[27] X. L. Wang, B. Jiang, and N. Y. Lu, "Adaptive relevant vector machine based RUL prediction under uncertain conditions," *ISA Transactions*, vol. 87, pp. 217–224, 2019.

[28] J. R. Zhang, T. Peng, C. Yang, Z. W. Chen, H. W. Tao, and C. H. Yang, "A voltage-based hierarchical diagnosis approach for open-circuit fault of two-level traction converters," *Electronics*, vol. 8, no. 992, pp. 1–15, 2019.

[29] C. Yang, W. H. Gui, Z. W. Chen, J. R. Zhang, T. Peng, C. H. Yang, H. R. Karimi, and S. X. Ding, "Voltage difference residual-based open-circuit fault diagnosis approach for three-level converters in electric traction systems," *IEEE Transactions on Power Electronics*, vol. 35, no. 3, pp. 3012–3028, 2020.

[30] C. Y. Zhang, C. S. Wang, N. Y. Lu, and B. Jiang, "An RBMs-BN method to rul prediction of traction converter of crh2 trains," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 46–56, 2019.

[31] B. Wang, Y. G. Lei, N. P. Li, and N. B. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Cybernetics*, vol. 69, no. 1, pp. 401–412, 2020.