Fault Diagnosis and Prognosis in multivariate complex systems

Junjie Yang

université PARIS-SACLAY

Fault diagnosis and prognosis in multivariate complex systems

Diagnostic et pronostic des défaillances dans les systèmes complexes multivariés

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Sciences du traitement du signal et des images
Graduate School : Sciences de l'ingénierie et des systèmes
Référent : CentraleSupélec

Thèse préparée dans la (ou les) unité(s) de recherche **Laboratoire des signaux et systèmes (Université Paris-Saclay, CNRS, CentraleSupélec)**, sous la direction de **Claude DELPHA**, Professeur.

Thèse soutenue à Paris-Saclay, le 5 Janvier 2023, par

Junjie YANG

Composition du jury

Membres du jury avec voix délibérative

Audine SUBIAS Professeur, INSA de Toulouse	Présidente
Mohamed BENBOUZID Professeur, Université de Bretagne Occidentale	Rapporteur & Examinateur
Vincent COCQUEMPOT Professeur, Université de Lille	Rapporteur & Examinateur
Demba DIALLO Professeur, Université Paris Saclay	Examinateur
Didier THEILLIOL Professeur, Université de Lorraine	Examinateur
Antoine PICOT Maitre de Conférence-HDR, Université de Toulouse	Examinateur

THESE DE DOCTORAT

NNT : 2023UPAST001

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Diagnostic et pronostic des défauts dans les systèmes complexes multivariés
Mots clés : Classification à une classe, défaut naissant, distance de Mahalanobis, Détection des défauts, isolement des variables défectueuses, estimation de la gravité des défauts

Résumé : Le diagnostic et le pronostic des défaillances ont suscité une attention considérable dans l'industrie et le monde universitaire en raison des exigences croissantes en matière de fiabilité, de disponibilité, de maintenabilité et de sécurité. Dans de nombreux scénarios industriels, tels que les processus chimiques, les systèmes d'entraînement des machines et la surveillance des structures, la détection opportune des défauts est d'une grande importance pour réduire les risques de sécurité, éviter les pannes de système et permettre aux systèmes de fonctionner dans des conditions optimales. De plus, la caractérisation d'une défaillance est également nécessaire pour offrir une aide à la décision pour la maintenance du système, le contrôle tolérant aux défaillances, l'estimation de la durée de vie restante, etc. Par conséquent, ce travail se concentre sur les trois tâches principales du problème de diagnostic des défauts : la détection des défauts, l'isolation des causes profondes (variables défectueuses) et l'évaluation de la gravité des défauts (estimation de l'amplitude des défauts). Au cours de la dernière décennie, de nombreuses approches de diagnostic de défauts ont été proposées à base de modèles ou de données de systèmes, ce qui a permis de réaliser des progrès significatifs dans ce domaine. Cependant, les méthodologies de diagnostic de défauts existantes souffrent toujours de leurs propres limitations, telles que le manque de données défectueuses suffisantes pour l'apprentissage, l'inefficacité face aux données distribuées complexes, la faible sensibilité aux défauts naissants, et l'interférence du bruit et des valeurs aberrantes. Parmi ces défis, le diagnostic des défauts naissants est l'un des plus délicats, ce qui suscite une attention croissante dans la littérature. Reconnues comme de minuscules déviations à évolution lente, les défauts naissants sont dangereux car ils risquent de se transformer en défaillances graves, mais ils sont difficiles à diagnostiquer en raison de la confusion entre les déviations minuscules et le bruit. Pour résoudre ce problème, ce travail propose une nouvelle méthode de classification à une classe mise en œuvre en générant des points d'ancrage spécifiques et en sélectionnant la marge de la région locale correspondante pour déterminer une région saine comme zone de décision. Ensuite, une mesure de distance particulière appelée distance de Mahalanobis locale est définie pour indiquer la distance entre un échantillon et la région saine. Sur la base de la méthode de classification à une classe proposée et de l'indice LMD, ce travail développe d'abord une approche de détection des défauts naissants en combinant l'indice LMD et la technique de somme cumulée de densité de probabilité empirique. Ce travail examine également l'efficacité de l'indice LMD en tant que caractéristique représentative pour la détection des défauts. Deuxièmement, ce travail propose une méthode d'isolation de la variable défectueuse pour les cas de défaut unique en combinant la technique LMD avec l'idée du diagramme de contribution. Troisièmement, une expression analytique du taux d'augmentation des défauts est dérivée de l'indice LMD pour la tâche d'estimation de la gravité des défauts. Enfin, nous développons une nouvelle approche basée sur la reconstruction en utilisant la distance de Mahalanobis locale comme indice de détection pour améliorer les performances d'isolation et d'estimation. La méthode améliorée peut isoler avec précision plusieurs variables défectueuses et estimer simultanément l'amplitude de leurs défauts. L'étude de cas basée sur les données de processus du réacteur à réservoir agité à flux continu montre que la technique LMD présente des avantages significatifs pour le problème de diagnostic des défauts, tels qu'une sensibilité élevée aux défauts naissants, une robustesse au bruit et aux valeurs aberrantes, et l'absence d'hypothèse de distribution. Les méthodes de diagnostic de défauts développées sur la base de la technique LMD sont nettement plus performantes que les solutions les plus récentes. L'étude comparative sur les données de roulement de la Case Western Reserve University indique que la technique LMD peut être utilisée comme approche d'extraction de caractéristiques et qu'elle est plus efficace et plus robuste que les autres techniques statistiques.

Title : Fault diagnosis and prognosis in multivariate complex systems

Keywords : One-class classification, Incipient fault, Mahalanobis distance, Fault detection, Faulty variable isolation, Fault severity estimation

Abstract : Fault diagnosis and prognosis have attracted huge attention in industry and academia for the increasing requirements on reliability, availability, maintainability, and safety. In many industrial scenarios, such as chemical processes, machine drive systems, and structure monitoring, timely fault detection is of great importance to reduce security risk, avoid system breakdown, and allow systems to operate in optimal conditions. Moreover, characterizing a fault is also necessary to offer decision support for system maintenance, fault tolerant control, remaining useful life estimation, and so forth. Therefore, this work focus on the three principal tasks : fault detection, root causes (faulty variables) isolation, and fault severity assessment (fault amplitude estimation). Note that this study on fault prognosis problem will not cover the discussion of remaining useful life estimation. In the past decade, abundant fault diagnosis approaches have been proposed based on either system models or data, making significant progress in this domain. However, the existing fault diagnosis methodologies still suffer from their own limitations, such as the lack of sufficient faulty data for training, ineffectiveness to complex distributed data, low sensitivity to incipient faults, and the interference of noise and outliers. Among these challenges, incipient fault diagnosis is one of the trickiest, thereby attracting increasing attention in the literature. Recognized as tiny deviations with slow-growing evolution, incipient faults are dangerous for the tremendous risk of developing to serious failures but are difficult to diagnose due to the confusion of tiny deviations and noise. To address this issue, this work proposes a new one-class classification method implemented by generating anchors and selecting the region margin to determine a healthy region as a decision area. Then a particular distance measurement called local Mahalanobis distance is then defined to indicate the distance between a sample and the healthy region. Based on the proposed one-class classification method and the LMD index, this work first develops an incipient fault detection approach by combining the LMD index and the empirical probability density cumulative sum technique. This work also discusses the efficiency of LMD as a representative feature for fault detection. Secondly, this work proposes the faulty variable isolation method for single fault cases by combining the LMD technique with the contribution plot idea. Thirdly, an analytical expression of fault increasing rate is derived from the LMD index for the fault severity estimation task. Finally, we further develop a new reconstruction-based approach using the local Mahalanobis distance as a detection index to improve the isolation and estimation performance. The improved method can accurately isolate multiple faulty variables and estimate their fault amplitudes simultaneously. The case study based on the Continuous-flow Stirred Tank Reactor process data shows that the LMD technique has significant benefits for the fault diagnosis problem, such as high sensitivity to incipient faults, robustness to noise and outliers, and no distribution assumption. The fault diagnosis methods developed on LMD significantly outperform state-of-the-art solutions. The comparative study on the Case Western Reserve University bearing data indicates that the LMD technique can be used as a feature extraction approach and is more effective and robust than the other statistical techniques.

Acknowledgements

Three years have passed since I came to France and started my Ph.D. study, which now becomes the most precious memory in my life. I could not grow into the person I am today without much help and support.

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Claude Delpha, for his invaluable guidance, support, and encouragement throughout my three years of research work. His insights and expertise have been invaluable in helping me to shape and focus my research. He is not only my guide on the road of scientific research but also a great friend in my life.

Secondly, I would like to thank my defense jury members and reviewers, Prof. Mohamed Benbouzid, Prof. Vincent Cocquempot, Prof. Demba Diallo, Prof. Didier Theilliol, Prof. Audine Subias, Prof. Antoine Picot, for their valuable feedback and support throughout the writing and defense processes. Their constructive comments and suggestions have greatly improved the quality of this manuscript.

Besides, I would like to acknowledge the China Scholarship Council's support for funding this research. Without their financial support, this work would not have been possible.

I also would like to thank my girlfriend, Dr. Yuting Feng, for her valuable suggestions on my research work and her company. She is the love of my life and also my best friend. I appreciate all the happy and upset moments we have experienced together in the past three years.

Last but not least, I would like to thank my colleagues and friends, Dr. Yifei Sun, Dr. Maria Veizaga, Dr. Xuewen Qian, Lu Zhang, etc., for their discussions on this work and their help during my Ph.D. study.

Finally, I would like to express my gratitude to my family, who always support and encourage me. I am proud of them.

Abstract

Fault diagnosis and prognosis have attracted huge attention in industry and academia for the increasing requirements on reliability, availability, maintainability, and safety. In many industrial scenarios, such as chemical processes, machine drive systems, and structure monitoring, timely fault detection is of great importance to reduce security risk, avoid system breakdown, and allow systems to operate in optimal conditions. Moreover, characterizing a fault is also necessary to offer decision support for system maintenance, fault tolerant control, remaining useful life estimation, and so forth. Therefore, this work focus on the three principal tasks : fault detection, root causes (faulty variables) isolation, and fault severity assessment (fault amplitude estimation). Note that this study on fault prognosis problem will not cover the discussion of remaining useful life estimation. In the past decade, abundant fault diagnosis approaches have been proposed based on either system models or data, making significant progress in this domain. However, the existing fault diagnosis methodologies still suffer from their own limitations, such as the lack of sufficient faulty data for training, ineffectiveness to complex distributed data, low sensitivity to incipient faults, and the interference of noise and outliers. Among these challenges, incipient fault diagnosis is one of the trickiest, thereby attracting increasing attention in the literature. Recognized as tiny deviations with slow-growing evolution, incipient faults are dangerous for the tremendous risk of developing to serious failures but are difficult to diagnose due to the confusion of tiny deviations and noise. To address this issue, this work proposes a new one-class classification method implemented by generating anchors and selecting the region margin to determine a healthy region as a decision area. Then a particular distance measurement called local Mahalanobis distance is then defined to indicate the distance between a sample and the healthy region. Based on the proposed one-class classification method and the LMD index, this work first develops an incipient fault detection approach by combining the LMD index and the empirical probability density cumulative sum technique. This work also discusses the efficiency of LMD as a representative feature for fault detection. Secondly, this work proposes the faulty variable isolation method for single fault cases by combining the LMD technique with the contribution plot idea. Thirdly, an analytical expression of fault increasing rate is derived from the LMD index for the fault severity estimation task. Finally, we further develop a new reconstruction-based approach using the local Mahalanobis distance as a detection index to improve the isolation and estimation performance. The improved method can accurately isolate multiple faulty variables and estimate their fault amplitudes simultaneously. The case study based on the Continuous-flow Stirred Tank Reactor process data shows that the LMD technique has significant benefits for the fault diagnosis problem, such as high sensitivity to incipient faults, robustness to noise and outliers, and no distribution assumption. The fault diagnosis methods developed on LMD significantly outperform state-of-the-art solutions. The comparative study on the Case Western Reserve University bearing data indicates that the LMD technique can be used as a feature extraction approach and is more effective and robust than the other statistical techniques.

List of publications

The work during PhD has led to 3 journal papers and 6 international conference papers.

International journal papers :

1. **J. Yang** and C. Delpha. An Incipient Fault Diagnosis Methodology Using Local Mahalanobis Distance : Detection Process Based on Empirical Probability Density Estimation. Signal Processing, 2022, vol. 190, p. 108308.
2. **J. Yang** and C. Delpha. An Incipient Fault Diagnosis Methodology Using Local Mahalanobis Distance : Fault Isolation and Fault Severity Estimation. Signal Processing, 2022, p. 108657.
3. **J. Yang** and C. Delpha. A New Reconstruction-based Method Using Local Mahalanobis Distance for Incipient Fault Isolation and Amplitude Estimation. , Under Review.

International conference papers :

1. **J. Yang** and C. Delpha. Empirical Probability Density Cumulative Sum for Incipient Fault Detection, Prognostics and Health Management Conference (PHM 2020), May 4 - 7, 2020, Besancon, France, IEEE, pp. 187-192.
2. **J. Yang** and C. Delpha. Open-Circuit Fault Diagnosis for Interleaved DC-DC Converters, IEEE International Conference on Industrial Electronics (IECON 2020), Oct 18 - 21, 2020, Singapore, Singapore, IEEE, pp. 3982-3987.
3. **J. Yang** and C. Delpha. A Local Mahalanobis Distance Analysis Based Methodology for Incipient Fault Diagnosis. IEEE International Conference on Prognostic and Health Management (ICPHM 2021), June 7-9, 2021, Detroit, Michigan, USA, IEEE.
4. **J. Yang** and C. Delpha. Local Mahalanobis Distance Envelope Using A Robust Healthy Domain Approximation For Incipient Fault Diagnosis. IEEE International Conference on Industrial Electronics (IECON 2021), Oct 13-16, 2021, Toronto, Canada, pp. 1-6.
5. **J. Yang** and C. Delpha. Incipient Fault Severity Estimation Using Local Mahalanobis Distance. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2022), May 7-13, 2022, Singapore, pp. 5977-5981.
6. **J. Yang** and C. Delpha. Bearing Faults Detection Approaches Using Statistical Feature Extraction and Probability Based Distance : A Comparative Study, IEEE International Conference on Industrial Electronics (IECON 2022), Oct 17 - 20, 2022, Brussels, Belgium.

Contents

List of figures

List of tables

Nomenclature

Acronyms

CCA	Canonical Correlation Analysis
CDF	Cumulative Density Function
CSTR	Continuous-flow Stirred Tank Reactor
CVA	Canonical Variate Analysis
CVDA	Canonical Variate Dissimilarity Analysis
DPCA	Dynamic Principal Component Analysis
EMD	Earth Mover Distance
EPD	Empirical Probability Density
EPD-CUSUM	Empirical Probability Density Cumulative Sum
EWMA	Exponentially Weighted Moving Average
FDA	Fisher Discriminant Analysis
FFT	Fast Fourier Transform
FNR	Fault to Noise Ratio
GCCA	Generalized Canonical Correlation Analysis
GMM	Gaussian Mixture Model
ICA	Independent Component Analysis
IF	Isolation Forest
JSD	Jensen-Shannon Divergence
KDE	Kernel Density Estimation
KLD	Kullback-Leibler Divergence
KPCA	Kernel Principal Component Analysis
KS	Kolmogorov Smirnov
LMD	Local Mahalanobis Distance
MCUSUM	Multivariate Cumulative Sum
MD	Mahalanobis Distance

MEWMA	Multivariate Exponentially Weighted Moving Average
MSE	Mean Squared Error
MSPM	Multivariate Statistical Process Monitoring
OC-SVM	One-Class Support Vector Machine
OCC	One-Class Classification
PCA	Principal Component Analysis
PCP	Principal Component Pursuit
PDF	Probability Density Function
PLS	Partial Least Squares
RBC	Reconstruction-Based Contribution
RBF	Radial Basis Function Kernel
RPCA	Robust Principal Component Analysis
SPE	Squared Prediction Error
STFT	Short Time Fourier Transform
TEP	Tennessee Eastman Process

Symbols

α	Significance level
$\bar{\mathcal{G}}$	Average position change vector
β	Scale parameter of GEV
$\boldsymbol{\Theta}$	Fault direction matrix
$\boldsymbol{\Xi}$	Candidate set of fault direction
$\boldsymbol{\zeta}$	Notation for fault estimation
Δ	Fault amplitude
$\mathcal{H}_0$	Healthy hypothesis
$\mathcal{H}_1$	Faulty hypothesis
$\mathcal{J}(\mathbb{F}, \mathbb{G})$	The collection of the joint distributions
$\mathcal{N}(m_1, \Sigma_1)$	Multivariate Gaussian distribution with mean m_1 and Covariance matrix Σ_1
Con	Contirbution of a variable being faulty
R_{in}	Cumulative sum of instantaneous log ratio
r_{in}	Instantaneous log ratio

$\mathrm{s}(x, n_e)$	Anomaly score of isolation forest
x^*	Healthy component
ω	EPD weight factor
$\overline{F}_F$	Frechet form of GEV for maxima
$\overline{F}_G$	Gumbel form of GEV for maxima
Φ_G	Cumulative density function of GEV
Ψ	Samples matrix for fault estimation
τ_a	Shape parameter of GEV
$\underline{F}_G$	Gumbel form of GEV for minima
$\underline{F}_W$	Weibull form of GEV for minima
ξ	Exponential smoothing index
ζ_s	Smoothing factor
b_o	Fault occurrence time
D_w	Wasserstein distance
$H_p(\cdot)$	Harmonic number
n_e	The number of external nodes
n_f	Number of faulty variables
P, Q	Unknown distributions
p	The order of Wasserstein distance
P_f	Probability of a sample being faulty
p_f	Power of faulty component.
p_n	Power of noise
$P_{\mathcal{H}_0}$	Probability of the health hypothesis
$P_{\mathcal{H}_1}$	Probability of the faulty hypothesis
T_e	Sampling time
th	Threshold
$\boldsymbol{\beta}_k$	Supporting objects
$\hat{\boldsymbol{X}}$	Reconstructed data
$\boldsymbol{\Lambda}$	Eigen (singular) values
$\mathbb{S}$	The set of independent source signal
$\boldsymbol{\mathcal{H}}$	Scaled Hankel matrix
$\boldsymbol{\mathcal{U}}$	Inverse of unmixing coefficients matrix
$\mathbf{p}_k$	Past data vector

$\boldsymbol{\mu}$	Mean vector of samples
$\boldsymbol{\omega}$	Weight matrix
$\boldsymbol{\Sigma}$	Covariance matrix
$\boldsymbol{\Sigma}_A$	Covariance matrix of augmented data
$\boldsymbol{\Sigma}_{ff}$	Covariance matrix of the future samples
$\boldsymbol{\Sigma}_{fp}$	Cross-covariance matrix of the future and past samples
$\boldsymbol{\Sigma}_{pp}$	Covariance matrix of the past samples
$\boldsymbol{b}$	Bias vector
$\boldsymbol{B}$	Transformation matrix from $\boldsymbol{H}$ to $\boldsymbol{J}$
$\boldsymbol{C}$	Anchor of healthy region
$\boldsymbol{E}$	Residual matrix between the original and projected data
$\boldsymbol{f}_k$	Future data vector
$\boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{J}$	Decomposition matrix of PLS
$\boldsymbol{K}$	Kernel matrix
$\boldsymbol{L}_0$	Low-rank component of RPCA
$\boldsymbol{M}$	Lagrange multiplier matrix
$\boldsymbol{P}$	Tramsformation matrix
$\boldsymbol{P}_{pc}$	Tramsformation matrix for principal components
$\boldsymbol{P}_{res}$	Tramsformation matrix for residual components
$\boldsymbol{r}, \boldsymbol{h}, \boldsymbol{q}$	Components of matrices $\boldsymbol{R}$, $\boldsymbol{H}$, $\boldsymbol{Q}$
$\boldsymbol{s}$	Independent source signals
$\boldsymbol{S}_0$	Sparse component of RPCA
$\boldsymbol{U}$	Left singular matrix
$\boldsymbol{V}$	Right singular matrix
$\boldsymbol{W}$	Unmixing coefficients matrix for ICA
$\boldsymbol{X}^{[N\times m]}$	Data matrix, where $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_N]^T$
$\boldsymbol{X}_A$	Augmented matrix
$\boldsymbol{X}_o$	Outliers
$\boldsymbol{Z}$	Score matrix or latent variable matrix
δ	Growing slope of a fault
ϵ	Reconstructed error
η	Scalar coefficient of RPCA
η_{an}	Minimum number parameter of local samples for anchor generation algorithm

Γ	Estimated average path length
γ	Bandwidth parameter of the Gaussian kernel function
γ_{an}	Local radius parameter of anchor generation algorithm
γ_{an}^{opt}	Optimal value of the local radius parameter
$\hat{\iota}$	Index of the isolated faulty variable
ι	Index of the true faulty variable
$\kappa(x, x)$	Kernel function
$\mathbb{F}, \mathbb{G}$	Cumulative density function
$\mathbb{L}$	Log ratio
$\mathbb{R}$	Real number space
$\mathcal{A}$	Anchor of a sample
$\mathcal{B}$	Function of separating hyperplane
$\mathcal{D}_\tau$	Singular value thresholding operator
$\mathcal{G}$	Position change of a sample
$\mathcal{I}_A(\cdot)$	Indicator function
$\mathcal{L}$	Augmented lagrangian function
$\mathcal{S}_\tau$	Shrinkage operator
$\mathcal{W}$	Normalization weight
A, B, C, D, K	Coefficient matrices of a linear model
$\mathrm{E}(\cdot)$	Expectation
$\mathrm{U}(a_1, a_2)$	Uniform distribution in the interval $[a_1, a_2]$
u	Input vector
v	Process noises
x	State vector
x^f	Faulty component
y	Output vector
μ	Incoherence condition parameter for updating the Lagrange multiplier matrix
ν	Balancing parameter
Ω_F	Faulty region
Ω_H	Healthy region
$\overline{F}_W$	Weibull form of GEV for maxima
$\Phi(x)$	Projecting function

ρ	Location parameter of GEV
$\sigma(\cdot)$	Sigmoid function
τ	Time lag shift
P	Probability density
$\underline{F}_F$	Frechet form of GEV for minima
Υ	Coefficients vector
$\varrho_0, \varrho_1, \varrho_2$	Coefficients of LMD calculation eqaution
ς	Element of covariance matiex
$\widetilde{\boldsymbol{K}}$	Gram matrix
ζ_i	Slack variables
a	Mixing coefficients
c	The center of hypersphere
D_e	Maximum likelihood of the log ratio
D_L	Local Mahalanobis distance
d_m	Mahalanobis distance
D_F	f Divergence
D_{js}	JS Divergence
D_{kl}	KL Divergence
D_{ks}	KS distance
e	Residual score of a sample
E_X	Residual matrices of data $\boldsymbol{X}$
E_Y	Residual matrices of data $\boldsymbol{Y}$
f, g	Probability density function
G	Nonlinear function for encoder
G'	Nonlinear function for decoder
h_p	Path length
I	Relative entropy
K_{an}	Anchors number
l	The number of principal components or latent variables
L_f	The lags of the future data
L_p	The lags of the past data
m	The number of variables
m_u	Dimension number of input vectors

m_y	Dimension number of output vectors
N	The number of observations
r	The radius of hypersphere
r_L	Healthy region margin
S	Anchor set
U_G	Control limit for fault detection

1 General introduction

1.1 Background and motivation

Fault diagnosis and prognosis of complex systems is nowadays a hot topic in industry and academia. With the rapid development of industrial technology, the efficiency and safety of equipment have been greatly improved. However, the risk of accidents still exists and threatens the safety of human life and property. For example, a serious high-speed rail collision occurred in Wenzhou, China, in 2011, with 40 people dying and at least 192 injured [44, 108]. It is believed that one of the major causes of this disaster was the failure of the railway signaling systems due to a lightning strike. In 2010, a violent explosion occurred at Tesoro Anacortes oil refinery in Washington State, USA, resulting in the death of seven people [53, 156]. The disaster involved the catastrophic failure of a heat exchanger housing. The heat exchanger had been continuously exposed to high temperature and pressure hydrogen for a long time, leading to the crack of its carbon steel shell. These two serious accidents have raised significant concerns about equipment safety and also the attention to fault diagnosis. In addition to security risks, failures can also lead to the reduced efficiency of systems. For example, in the feedback control of the air-fuel ratio of automobile engines, abnormal oxygen sensors increase harmful emissions and reduce fuel economy [94, 165]. Although controllers can compensate for slight sensor faults to maintain satisfactory operations, the inaccurate feedback degrades systems' performance and even lead to an unexpected breakdown. Moreover, any deviation of sensor signals (considered a fault) in ultraprecision manufacturing systems is unacceptable and should be detected as soon as possible to achieve high levels of process productivity [46, 147].

With these growing demands for safety, reliability, and maintainability, the development of fault diagnosis technology is urgent. However, the increased complexity of modern systems raises new challenges for security surveillance and system maintenance. Modern industrial systems are usually integrated with a large number of components and driven by advanced techniques. The large number of components increase the risk of failure

and the difficulty of locating fault root causes [61, 174]. Monitoring the key components individually, such as the short circuit of input current and the pressure of a reactor, is still necessary to ensure systems security. But this strategy is time-consuming and expensive for complex systems. On the other hand, the individual components interact with each other, making it more challenging to locate root faulty components [54, 146].

The growing development of technologies involving big data, the internet of things, machine learning, artificial intelligence, etc., provides a real opportunity to deal with these challenges. In the last decade, a large number of fault diagnosis techniques have been proposed and used in real applications, such as the detection of aging components in high-speed rail vehicle suspension systems [186, 196], faulty sensors detection in nuclear power plants [55, 177], and crack detection in civil engineering structures [48, 191]. These techniques can save human lives and significantly increase productivity and efficiency. It is believed that more and more fault diagnosis techniques will be applied to different scenarios, which motivates us to keep improving fault diagnosis techniques.

1.2 Objectives

Conventionally, a fault is *"the unpermitted deviation of at least one process parameter from an acceptable condition"* [81]. In the signal processing domain, fault diagnosis aims to detect and characterize a faulty behavior of systems from their observed signals. There are different tasks of fault diagnosis in different application scenarios. More specifically, this work focuses on three common problems : fault detection, faulty variable isolation, and fault severity estimation.

Fault detection, as the foundation base of fault diagnosis, is dedicated to decide if a system operated in faulty condition. It has numerous industrial scenarios, such as chemical process monitoring [6, 169], defective product detection [15, 50], crack detection [76, 224], and abnormal behavior detection [168, 219]. Fault detection essentially is a classification problem trying to distinguish faulty patterns from healthy ones. However, fault detection problem faces particular challenges, especially in industrial applications. For example, faulty behaviors of complex systems are diverse and unpredictable, resulting in insufficient data on the faulty pattern. Without these faulty data for training, typical classification approaches are unavailable. Moreover, there is usually a higher requirement in detection delay time (time between fault occurrence and detection) since fault-tolerant or protection actions are necessary after a fault is detected [1, 5]. Recently, incipient fault detection has attracted increasing attention. It aims to recognize tiny deviations in monitoring signals to provide early warning. Incipient faults are easily confused with noise, making them more challenging to diagnose.

After the fault detection step, isolating faulty sources (variables) is also crucial for subsequent system maintenance. When a fault occurs in a system, sensors' monitoring signals may have different characteristics. This faulty behavior is usually caused by one or more malfunctioning components, which should be isolated for maintenance purposes. For instance, in the air brake control system of high-speed trains, the brake cylinder pressure is assigned by the control unit according to braking requirements [22, 85]. The sensors measuring cylinder pressure may suffer from tiny deviation due to the train's harsh operating environment. Although these incipient faults will not directly cause a serious

accident, the risk increases with time, threatening peoples' safety. Therefore, the faulty sensors should be isolated at the early stage of a fault for maintenance and replacement.

Fault severity estimation is an important part of the fault prognosis problem, dedicated to assessing how serious the fault is. Unlike most of the literature, this work does not directly discuss the remaining useful life estimation in the fault prognosis problem. Instead, we estimate the fault severity in this problem to indicate the systems' remaining useful life. Generally, fault severity can be evaluated by estimating either the increasing rate or the fault amplitude. Take the bearing fault diagnosis as an example. The rolling bearing elements are one of the essential components in rotating machine drives. Bearing faults, such as inner race wear, cage fracture, and outer race fracture, frequently occur due to harsh operating conditions and sometimes manufactory defects [109, 190]. After fault detection, it is necessary to assess the fault severity to decide on protection actions. In the event of a serious fracture in a component, the system should be stopped immediately, while a warning should be raised for minor wear. Therefore, this work tries to estimate the fault severity of the faulty signals.

Motivated by the above discussion, this work aims to propose fault diagnosis solutions for different tasks with the following goals.

1. Develop fault diagnosis approaches for complex distributed data, such as non-Gaussian cases.
2. Be able to offer fault diagnosis solutions when faulty information is weak or missing.
3. Improve the sensitivity of fault diagnosis approaches to incipient faults.
4. Achieve strong robustness to noise and outliers.

1.3 Contributions

The contributions associated with this work are composed of the following points.

1. **We propose the healthy region approximation approach as one-class classification method for fault diagnosis.**
   - By considering samples' spatial distribution, this method determines the healthy region of fault-free samples with multiple anchors and a region margin.
   - A robust anchor-generation algorithm with its optimization procedure is proposed to extract the spatial distribution information of the healthy region.
   - Based on the generalized extreme value distribution, we estimate the CDF of the LMD index in the healthy condition and select the optimal healthy region margin.
   - Based on the approximated healthy region, a key index named the local Mahalanobis distance (LMD) is defined, and its properties are also studied so as to provide a theoretical foundation for fault detection.
2. **We propose fault detection methods for time series signals and time-independent data, respectively.**
   - The first solution combines the LMD index with the Empirical Probability Density Cumulative Sum (EPD-CUSUM) method to increase sensitivity to incipient faults.
   - The benefit of using the EPD-CUSUM method is shown by comparing it with the original LMD index and typical detection methods, which again validates its excellent detection capability.

- The second detection solution extracts the LMD feature and calculates the dissimilarity between the healthy reference data and the testing one using probability-based distance.
- Compared with other statistical feature extraction approaches, such as principal component analysis, kernel principal component analysis, and independent component analysis, the LMD technique can extract more representative features for fault detection.

3. **A faulty variable isolation method using the contribution-plot idea is developed to isolate the single faulty variable.**
   - The fault contribution value of each variable is obtained by analyzing the relative position of faulty data and its anchor.
   - The proposed method is simple to implement and drastically improves isolation accuracy compared to existing approaches, especially for incipient faults.
4. **We establish the analytical expression for fault severity estimation connecting the fault increasing rate and LMD index.**
   - This approach succeeds the advantages of LMD and achieves high accuracy in the estimation procedure.
5. **We further propose the improved fault isolation and fault amplitude estimation approaches based on the reconstruction-based contribution framework.**
   - The smooth LMD index is designed and used to inhibit noise disturbance with high frequency, which significantly improves the isolation and estimation performance of the proposed method.
   - To overcome the difficulty of optimizing the non-linear calculation of LMD, this work proposes an iterative method alternating between updating an anchor and solving the optimization problem. Then an approximation solution for RBC calculation is obtained.
   - The RBC-based method can isolate multiple faulty variables and estimate their fault amplitude simultaneously.
6. **Based on the industrial engineering applications, we comprehensively evaluate the performance of each fault diagnosis solution.**
   - In the case study of the Continuous-flow Stirred Tank Reactor (CSTR) process, the performance evaluation shows that the proposed fault diagnosis approaches are robust to noise, highly sensitive to incipient faults, and outperform the state-of-the-art techniques.
   - The comparative study on the Case Western Reserve University (CWRU) bearing data indicates that the LMD technique can extract more meaningful features for fault detection and achieves the best sensitivity performance in this study.

1.4 Outline

The remaining part of this thesis is organized as follows.

Chapter 2 briefly introduces the concept of fault and fault diagnosis, such as fault definition and category, fault diagnosis tasks, and diagnosis procedures. The development of fault diagnosis methodologies is reviewed, and then challenges are highlighted. Specifically,

we review and summarize the state-of-the-art techniques significantly contributing to the fault diagnosis problem. The limitation and advantages of each technique are also discussed by considering the particular case of incipient faults.

Chapter 3 first reveals the major limitations of the existing fault diagnosis methodologies and then discusses the benefit of using the One-Class Classification (OCC) idea to tackle these challenges. A new OCC approach composed of the anchor-generation algorithm and region margin selection is proposed to develop a healthy region for fault-free samples. The key index named local Mahalanobis distance (LMD) is then defined to indicate the distance between samples and the healthy region. The healthy region effectiveness of the proposed method is validated, and its benefit is highlighted using simulation data.

In chapter 4, solutions for fault detection, faulty variables isolation, and fault severity estimation are proposed based on the developed healthy region and the LMD index. Based on simulation signals, we give the step-by-step guide of each solution and show the diagnosis results.

The efficiency evaluation of the proposed methods is considered in chapter 5 using two industrial engineering application data, namely Continuous-flow Stirred Tank Reactor (CSTR) process and Case Western Reserve University (CWRU) bearing data. The advantages of our proposal are highlighted by comparing it with the state-of-the-art approaches.

Finally, chapter 6 concludes this work and presents the perspectives of our work.

2

Fault diagnosis approaches : Review

2.1 Introduction

Health monitoring and fault diagnosis play a key role in modern industrial systems for increasing requirements on reliability, availability, maintainability, and safety. With the development of sensor technology and computational science, fault diagnosis, including fault detection, faulty variables isolation, and fault severity estimation, has become a more and more hot topic in the last decade. Its applications cover abundant scenarios such as manufacturing industry, electrical systems, and chemical processes [8, 193, 214].

A fault is conventionally defined as *"the unpermitted deviation of at least one process parameter from an acceptable condition"* [81]. It can be classified according to different attributes. By considering the time dependency and severity, we introduce three kinds of faults, namely abrupt fault (stepwise), intermittent fault (pulsewise), and gradual fault (incipient fault) [13, 39, 81]. Figure 2.1 exhibits the evolution of three kinds of faults along with time.

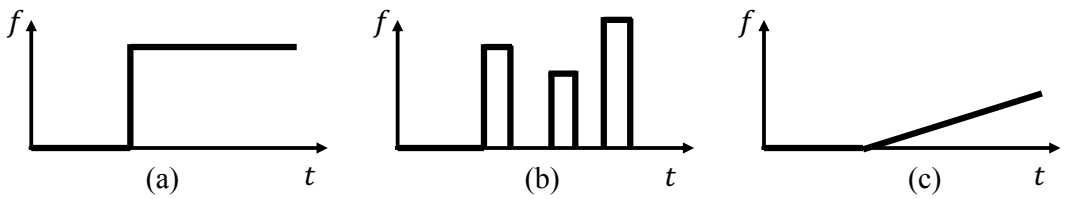

FIGURE 2.1 – Three kinds of faults : a) abrupt fault ; b) intermittent fault ; c) gradual fault.

- **Abrupt fault** is recognized as step-like deviation, as shown in Fig.2.1-(a). It usually links to serious damage of components, like the short circuit of power devices [162], the fracture of mechanical structures [24], and valve malfunction in a process [128].

The sudden occurrence of faults with high amplitude may immediately cause severe performance degeneration and even system breakdown. For distinct fault signatures, the detection of abrupt faults is less challenging. However, more attention should be paid to the detection delay time in detecting abrupt faults.

- **Intermittent fault** appears randomly with short time duration in a process, as shown in Fig.2.1-(b). In contrast to the persistent behavior of abrupt fault, intermittent fault results in systems switching between a faulty and healthy behavior, which is more tricky to detect. It is usually caused by the aging of equipment [2], poor soldering [221], and partial fault of a distributed network [137].
- **Gradual fault**, also called incipient fault, is defined as an increasing change with very low severity or slowly varying evolution (see Fig.2.1-(c)). It does not dramatically affect the system's performance at its early stage but may result in severe system failures if no protective action is taken. Therefore, early diagnosis of gradual fault in their incipient stage is crucial to discover the potential risk of system breakdown, which is also becoming a hot topic in the fault diagnosis domain. Examples of incipient fault are the slight crack of metal [102], mechanical wear and tear [134], and sensors drift [85].

All these faults can lead to system performance degradation and even an unexpected stop [81, 175] if neither protection action nor fault-tolerant control is taken. To avoid a severe accident and reduce economic impact, protection actions or necessary fault-tolerant control are required for malfunctional systems, which rely on quick and accurate fault detection. After a fault is detected, the fault isolation and fault severity estimation are also crucial to characterize a fault and to provide information for fault-tolerant control steps and system maintenance.

Generally, a fault diagnosis scheme includes three main tasks [39], fault detection, faulty sources isolation (sometimes treated as fault identification or classification), and fault severity assessment, as shown in Figure 2.2. They link to real applications like process state monitoring [26], anomaly detection [20], fault-tolerant control [212], and remaining useful life estimation [176].

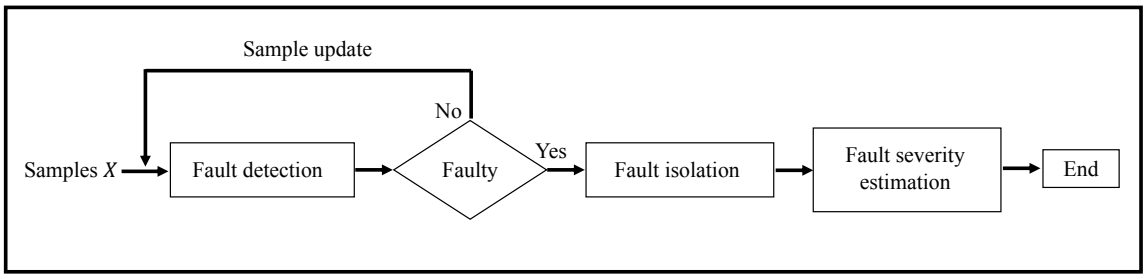

FIGURE 2.2 – General fault diagnosis framework

- **Fault detection** is the first step and the foundation base of the fault diagnosis framework. It aims to accurately recognize a faulty behavior for the input sample, which may face challenges like strong noise and small fault amplitude [66, 81]. For online applications, an additional requirement of the fault detection step is to detect a fault with a short delay time [142, 200].
- **Fault isolation**, also called fault root causes location in literature, proceed to locate the key factors affecting the system's performance after a fault is detected [83,85,213].

This step is usually realized based on some necessary assumptions, such as fixed candidate fault root causes and known fault types.

- **Fault severity estimation** continues to assess the fault severity when the fault root cause is isolated [68,135]. It is an important task of the fault prognosis problem, providing crucial information for remaining useful life prediction, fault-tolerant control, and so on [23,186]. More precisely, the physical meaning of fault severity is first specified according to application demand, such as increasing rate, the amplitude of additive fault, the gain of multiplicative fault, and the key parameter of a system component. The defined value is then estimated sometimes based on necessary assumptions, e.g., known fault type (abrupt, intermittent, incipient,...) and the fault's evolution model.

In the past decade, abundant fault diagnosis approaches have been proposed based on either system model or data. Model-based approaches require the knowledge of systems, such as mathematical process models [81,193]. The relationship between system inputs and outputs in the healthy operating condition is first established by using parametric or non-parametric methods. In the diagnosis procedure, the estimated outputs are calculated based on the obtained model and the new measured input signals. The residual error between the estimated and actual output signals is then generated and serves as a fault signature for fault detection and characterization. Model-based approaches are reliable when the basic model structure is known. However, fault diagnosis performance of these approaches is usually inadequate for complex systems due to the difficulty for developing an accurate process model, the large number of latent variables, and non-linear behavior.

With the development of computational science, data storage and management have become easier, which provides a real opportunity to analyze massive data. Therefore, data-driven approaches for fault diagnosis have attracted increasing attention. By directly extracting useful information from measured signals, data-driven approaches can allow to provide an effective decision and achieve promising diagnosis performance. Data-driven approaches may consist of different advanced techniques, which is generally summarized as four steps : pre-processing, feature extraction, feature analysis, and decision making [69].

- **Pre-processing step** processes raw signals in advance of the diagnosis procedure to ensure the correctness and availability of data. For example, in this step, the missing value of measured signals is handled using one of the estimation strategies. When raw signals contain a lot of noise, the denoising operation is performed on data. In addition, normalization and signal resampling could be also performed in this step.
- **Feature extraction** extracts meaningful features from pre-processed signals and removes irrelevant information. The ideal feature can reflect the change caused by faults but is insensitive to noise and other disturbances. As the key of data-driven approaches, feature extraction is usually performed based on expert knowledge or statistical properties.
- **Feature analysis** summarizes the dissimilarity between healthy reference data and testing data to support the decision-making. A diagnosis index is obtained for the fault detection task, which may be further used to isolate a fault's root causes and estimate fault severity.
- **Decision-making** is the final step of the fault detection task. The tradeoff among the detection rate, false alarm rate, miss detection rate, and the detection delay time is considered when one selects a threshold. The final decision is drawn in this

step by comparing the diagnosis index with the detection threshold.

In the following section, we summarize the challenges of the fault diagnosis problem and introduce the development of data-driven fault diagnosis methodologies. Then we look deeper into the state-of-the-art techniques that significantly contribute to the fault diagnosis problem. The definition of these techniques is reviewed, and their step-by-step usages are summarized.

2.2 Data-driven fault diagnosis methodologies : development and challenges

In the last two decades, lots of fault diagnosis methodologies have been proposed and applied to industrial systems, roughly divided into two main categories : model-based and data-driven approaches according to their prior knowledge. As discussed in the previous section, model-based fault diagnosis approaches are reliable and usually achieve excellent diagnosis performance when systems' knowledge is available, but are difficult to design for complex systems. Instead, data-driven approaches have more interesting advantages like flexibility and robustness. They explore data's characteristics and find the difference between healthy and faulty cases by describing the system's behavior in a more flexible way, such as statistic metric and distance in the feature space. Moreover, data from real systems may suffer from interference, distortions, and environmental noise. Facing these difficulties, using data-driven methods helps to develop fault diagnosis approaches with strong robustness. Therefore, they are wildly used and much more popular than model-based approaches. This section focuses on data-driven approaches by introducing their improvement to cope with different challenges.

2.2.1 Fault detection approaches

Fault detection is the foundation for fault diagnosis and prognosis, naturally attracting huge attention in the literature and yielding various approaches. Masses of studies were presented in the literature to characterize fault behavior and provide solutions for fault detection.

Typical methods like Shewhart charts, cumulative sum, and Exponentially Weighted Moving Average (EWMA) provided simple and effective univariate measurements to characterize process behavior [120, 140, 149, 161]. Then, parts of these methods were extended for multivariate analysis, such as Multivariate Exponentially Weighted Moving Average (MEWMA) and the Multivariate Cumulative Sum (MCUSUM) [59, 122]. For high dimensional data, multivariate statistical techniques are used to reduce the data dimension and further improve the detection capability by considering the relevant information in a low dimensional subspace. These kinds of approaches first apply multivariate statistical techniques, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Fisher Discriminant Analysis (FDA), to extract data's features. Then they compute Hotelling's T^2 and SPE(Q) statistics as discriminated indexes to determine if the observed data is different from the healthy one [199]. The work described in [198] has summarized the common multivariate statistical techniques and compared their performance based on the Tennessee Eastman Process (TEP) data. Later, multivariate statistical

techniques aiming to develop the relation between system input and output were applied to the fault detection problem. These techniques, like Partial Least Square (PLS) [125], Canonical Correlation Analysis (CCA) [28], and Canonical Variate Analysis (CVA) [139], usually learn latent variables of systems as fault features. Hotelling's T^2 and SPE statistics are also used for feature analysis in these approaches.

Multivariate statistical techniques are powerful tools for extracting features and usually achieve good performance in the fault detection task. However, the drawbacks of these methods lies in their ineffectiveness for non-Gaussian data, non-linear processes, and dynamic systems [198]. Using the kernel trick is a common way to cope with the non-linear nature of data. The extension of the conventional PCA technique using kernel function, called kernel PCA, was used in fault detection task and showed its superiority in non-linear cases. Similarly, non-linear ICA [104, 192] and non-linear PLS [145, 183] are good choices for analyzing the non-linear structure of data. Concerning the challenge of non-Gaussian data, methods based on the Gaussian Mixture Model (GMM) [90, 203] and Generalized Canonical Correlation Analysis (GCCA) [27] were proposed without the Gaussian distributed assumption of data. Fault detection for dynamic systems is also a tricky problem due to the changing behavior of systems. Several improvements of multivariate statistical techniques for this concern have been made by considering the time-varying information of data, such as the extension of PCA for dynamic systems (dynamic PCA) [45], and the Canonical Variate Dissimilarity Analysis (CVDA) approach [142]. Despite the significant progress of multivariate statistical techniques contributing to the fault diagnosis problem, they usually fail to handle complex systems involving a high dimensionality, the mass of data, and unusual data structure [105, 106]. The extracted features under statistical optimization goals may not relate to a system's faulty information. [106].

Recently, artificial intelligent approaches based on machine learning theories have been widely used in fault detection problems for their powerful feature representation capability. In particular, neural network-based methods, which can automatically extract faulty features and recognize different kinds of samples, are attracting much interest from academia [30, 106, 179]. However, one challenge of applying artificial intelligent approaches to industrial applications is the insufficient number of high-quality training samples. Most artificial intelligent approaches are supervised and require a large number of labeled data for both healthy and faulty scenarios. However, for most industrial applications, faulty data is usually scarce or unavailable. The limited number of training samples can lead to the poor performance of supervised and even unsupervised machine learning approaches. Although it is possible to develop fault diagnosis methods based on the knowledge or the data of known faulty behaviors, applying these fault diagnosis methods to unknown faulty behaviors is sometimes unacceptable. Therefore, the one-class classification (OCC) problem is raised and discussed to cater to the practical necessity of industrial cases. OCC techniques, such as k-centers [201], One-Class Support Vector Machine (OC-SVM) [93, 138, 157], auto-encoder [126, 180], and isolation forest [65, 114, 115], tend to become promising solutions for fault diagnosis in industrial applications. However, the OCC problem still faces several challenges, like the outlier issue, inadequate sensitivity for small deviation, and the non-linear nature of data.

Another challenge of the fault detection problem is that subtle deviation is usually buried in noise, leading to the detection difficulties [142, 153, 216]. The subtle deviation, also called incipient fault, is described as a slowly varying change of parameters, or

disturbance with a tiny fault amplitude compared to the noise power [68,81]. Unfortunately, most of the aforementioned approaches are reported as not sensitive enough to detect incipient faults except the CVDA method [142]. To improve detection sensitivity, the works [38, 67, 208, 216] introduced a new detection framework that first uses multivariate statistical techniques to extract fault features and then employs divergences to evaluate the dissimilarity between the observed data and healthy data. Divergence is highly sensitive statistical metric evaluating the closeness between two probabilities. In practice, the probability distributions of data's feature is estimated by using Kernel Density Estimation (KDE) method. Although these methods show the benefit of using the Kullback–Leibler divergence (KLD) and Jensen–Shannon divergence (JSD) in incipient fault detection, they seem unsuitable for online applications since they need a large number of samples to accurately estimate the probability distribution.

2.2.2 Fault isolation and estimation approaches

Compared to the considerable progress of fault detection, studies on incipient fault isolation and severity assessment are more tedious. Fault isolation aims at finding out the parameters or signal sources affected by a fault, and fault severity assessment is usually achieved by estimating the characteristics of a fault (amplitude, occurrence time, evolution function). Accurate solutions to both problems are necessary for fault-tolerant control [83,212], components' remaining useful life prediction [110], and the risk assessment of system breakdown [204].

Numerous model-based methods specific to particular systems have been proposed for fault isolation and severity estimation problems, but they usually suffer from weak generalization capability for other systems. General methods for fault isolation and fault severity estimation problems are therefore attracting researchers' attention. Additionally, to take advantage of the excellent characteristics of the existing detection approaches, such as high sensitivity, reliability, and robustness, researchers tend to develop fault isolation and fault severity estimation methods based on the fault monitoring index. Based on this motivation, several important works have been proposed [23, 69, 84, 213, 220], where divergence-based and reconstruction-based contribution techniques deserve particular attention for their high sensitivity or reliable performance.

Divergence-based methods are highly sensitive to incipient faults and robust to noise. Therefore, KLD and JSD serve as fault detection indexes and subsequently are used for fault isolation and severity estimation [41, 68, 216]. KLD combined with PCA feature was first proposed for incipient fault detection in works [66,69,200], as we introduced in the last section. Subsequently, the theoretical model estimating the fault severity coefficient was then developed based on KLD with Gaussian distributed assumption [67, 68]. To isolate the single faulty variable, C. Delpha et al. further proposed the Z-decomposition to linearly combine variables with binary coefficient and calculated the KLD of the combined signals between reference and testing samples [41]. The detection result of all combined signals is decoded from binary code to uniquely determine a single faulty variable (e.g., the result 010 indicates that the third variable is faulty). Similar to KLD-based approach, JSD was proposed with the same strategy to detect incipient faults, isolate faulty variables, and estimate fault severity. The main benefit of using JSD is its higher sensitivity for incipient faults facing noise environments [216, 217]. However, both KLD and JSD based approaches

have limitations. For example, they require a mass of samples to accurately estimate the probability density function.

The contribution plots method is also commonly used for sensor fault isolation [72, 181], which assumes that faulty variables make a higher contribution to a fault detection index. This method is simple but only suitable for faults with large amplitude. Later, the reconstruction-based contribution (RBC), combining the fault reconstruction idea and contribution plot technique, was proposed with a solid theoretical foundation to improve the isolation performance and enable the fault amplitude estimation [3, 83, 84]. RBC methods are developed on the existing monitoring indexes, such as PCA-based T^2 and SPE statistics [3], Mahalanobis Distance (MD), and Exponentially Weighted Moving Average (EWMA) [83, 84]. More precisely, they reconstruct signals by removing the faulty component in a specify fault direction to recover healthy signals. Subsequently, one of the monitoring indexes, such as Hotelling's T^2 statistics, SPE statistics [3], the combined statistics of T^2 and SPE [207], and Mahalanobis distance [37], is calculated to decide if the reconstructed signal is correctly recovered. Once the monitoring index of the reconstructed signal is lower than the detection threshold, the faulty direction indicating the faulty variables is determined, and the faulty component is obtained simultaneously. RBC approaches have a solid theoretical foundation and can achieve a reliable fault diagnosis performance for serious faults with a distinct deviation from healthy conditions. However, the aforementioned monitoring indexes are not sensitive enough for tiny faults, leading to the low accuracy performance of the existing RBC-based approaches for incipient faults isolation and fault severity estimation.

2.3 State-of-the-art techniques for fault diagnosis

Various techniques can be used to develop fault diagnosis approaches. Conventionally, signal processing approaches based on either time-domain, frequency-domain, or time frequency-domain are used to extract features and analyze signals' characteristics [134]. For example, in time-domain analysis, signals' temporal features are extracted, such as mean value, root-mean-square, peak value, skewness, kurtosis, etc. [62, 162]. They are reliable and simple for computation. Similarly, fast Fourier transform (FFT) is frequently used to extract frequency components of raw signals, which helps to observe the change of frequency components when faults occur [62]. However, the frequency analysis is not suitable for non-stationary signals. Therefore, time-frequency analysis approaches, such as short-time Fourier transform (STFT) [33] and wavelet transform [217], were then proposed. Although typical signal processing techniques are effective as feature extraction tools, they have limitations, like the requirement of prior knowledge of faulty signals and high complexity for calculation [134]. The technical review in this section does not include typical signal processing techniques but focuses on other state-of-the-art techniques.

Statistic-based techniques are one of the promising methodology branches among data-driven fault diagnosis approaches. They extract data's statistical characteristics as features that are usually sensitive to the change of signals when faults occur. Moreover, multivariate statistic techniques can decrease the number of features to improve detection capability and reduce the computational burden. Therefore, statistic-based techniques have been dominant tools for feature extraction in fault diagnosis tasks and deserve particular reviews in this section.

Machine learning and deep learning techniques (artificial intelligence-based approaches) for fault diagnosis are prevalent since they can automatically extract features and effectively handle non-linear data [111, 205]. Both the detection capabilities of supervised and unsupervised learning approaches are well explored in the literature [105, 189]. However, the major challenge of applying artificial intelligence approaches to industrial scenarios is the lack of faulty samples for training. Unsupervised learning approaches also require faulty samples for training though they don't need samples' labels. The special artificial intelligence approaches only training on positive samples is referred to as the one-class classification (OCC) technique. The use of this technique for fault diagnosis has attracted increasing attention. Accordingly, this section introduces the state-of-the-art OCC approaches to show more details.

In the feature analysis step, two statistics derived from Euclidean distance and Mahalanobis distance, namely SPE(Q) and Hotelling's T^2, are commonly used as indexes to determine if the observed data is different from the healthy one [199]. Alternatively, probability density based distances are increasingly used to analyze samples' features for their high detection sensitivity. To pursue this development, we introduce state-of-the-art statistical metrics significantly contributing to the incipient fault diagnosis problem.

2.3.1 Multivariate statistic techniques

Multivariate statistic techniques refine useful information from raw data by using the different dimension reduction and space partition approaches, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Fisher Discriminant Analysis (FDA), Partial Least Squares (PLS), Canonical Variate Analysis (CVA), Generalized Canonical Correlation Analysis (GCCA), etc. [27, 88, 125, 154, 198]. Then, the Hotelling T^2 index and SPE statistics are calculated on different subspaces for fault detection purposes [198]. In the following introduction, we consider a multivariate sample matrix $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_N]^T \in \mathbb{R}^{N \times m}$ with zeros mean, where N is the number of observations, and m is the number of variables.

2.3.1.1 Principal component analysis and its extensions

- **Principal component analysis (PCA)** is a widely used technique for dimensionality reduction and feature extraction. Due to its simple operation and the efficiency of handling massive and high-dimensional data, PCA is popularly used for fault diagnosis [56, 164]. By projecting the original data to a lower-dimensional space, PCA preserves data's correlation structure and removes redundancy information. Therefore, the features extracted by PCA operation can effectively characterize the change of measured signals.
  The PCA operation divides the original space into two subspaces : principal and residual components. Both are derived as a linear combination of original variables with the purpose of maximizing the variance of projected samples. The principal components space is spanned by the principal components vectors that contain most of the data's variation. For example, the first principal component explains the major variance of data, while the second one explains the most variance when the first one is removed. Let us consider the projected sample $\boldsymbol{x}\boldsymbol{p}_i$, where $\boldsymbol{p}_i \in \mathbb{R}^m$

is called loading vector. The objective of PCA is associated with the following optimal problem.

$$\begin{aligned} &\text{maximize } \mathrm{Var}(\boldsymbol{x}\boldsymbol{p}_i) = \boldsymbol{p}_i^T \mathrm{Var}(\boldsymbol{x})\boldsymbol{p}_i \\ &\text{subject to } \|\boldsymbol{p}_i\| = 1 \end{aligned} \tag{2.1}$$

The m orthogonal loading vectors $\boldsymbol{p}_1 \cdots \boldsymbol{p}_m$ can be calculated as the eigenvectors by eigen-decomposing the data's covariance matrix $\boldsymbol{\Sigma}$.By ordering the loading vectors according to their corresponding eigenvalue, the transform matrix $\boldsymbol{P} = [\boldsymbol{p}_1 \cdots \boldsymbol{p}_m]$ is then obtained. We can select the first l vectors $\boldsymbol{P}_{pc} \in \mathbb{R}^{m \times l}$ as the principal components and obtain the score matrix $\boldsymbol{Z}$.

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{P}_{pc} \tag{2.2}$$

In this step, selecting the proper number of principal components is crucial to represent the data in an optimal way, which becomes an important issue in using the PCA technique. Therefore, various studies proposed their criteria for selecting the optimal principal components number, such as eigenvalues limits, cumulative percent variance, cross-validation method, the variance of reconstruction error [130, 166]. The reconstruction of original data using principal components is obtained as

$$\hat{\boldsymbol{X}} = \boldsymbol{Z}\boldsymbol{P}_{pc}^T \tag{2.3}$$

The residual matrix $\boldsymbol{E}$ is derived as the difference between $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$

$$\boldsymbol{E} = \boldsymbol{X} - \hat{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{pc}\boldsymbol{P}_{pc}^T) \tag{2.4}$$

The subspace spanned by $\boldsymbol{E}$ is called the residual space, which represents the error information caused by the PCA reconstruction. Based on the above introduction, we briefly formulate the PCA calculation procedure as follows

1. Step 1 : Calculate the samples' covariance matrix

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T \tag{2.5}$$

2. Step 2 : Perform the eigen-decomposition on the covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T \tag{2.6}$$

   where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_m)$ and $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$.

3. Step 3 : Calculate the first l principal component score of the sample $\boldsymbol{x}_i$

$$\boldsymbol{z} = \boldsymbol{x}_i\boldsymbol{P}_{pc} \tag{2.7}$$

4. Step 4 : Calculate the residual score of the sample $\boldsymbol{x}_i$

$$\boldsymbol{e_i} = \boldsymbol{x}_i(\boldsymbol{I} - \boldsymbol{P}_{pc}\boldsymbol{P}_{pc}^T) \tag{2.8}$$

Typical PCA-based fault detection approaches calculate the Hotelling T^2 statistic of principal components and the SPE statistic of residual components as the two fault diagnosis indexes. It is believed that the SPE statistic is more sensitive to

a fault than the T^2 statistic since the former relates to the change of signals, while the latter relates to the steady state of healthy operation [47, 198]. To take advantage of the two statistics, their combined diagnosis index has been proposed (called PCA-MIX). It was demonstrated that the combined index outperforms each index [207]. However, this typical approach with the above arbitrary indexes is ineffective for incipient faults [66, 142]. As a common feature extraction tool for fault diagnosis, the PCA technique is widely used in conjunction with other advanced techniques, like KL divergence [40, 66, 68, 69], JS divergence [216, 217], empirical mode decomposition [47] and the K-Nearest Neighbors [211]. Despite such progress these approaches make in different aspects, PCA is still unsuitable for handling complex data due to its linear nature.

- **Kernel principal component analysis (KPCA)** is an extension of the conventional PCA, improving the ability to handle non-linear data [74]. The conventional PCA can extract useful features by reducing data's dimension when variables are jointly normally distributed and highly correlated with each other. However, the above assumption is not satisfying when data is collected from a non-linear system, such chemical process [32, 124]. The use of conventional PCA in this case is inadequate, which motivates to propose the KPCA technique.
  Theoretically, N sample points can almost always be linearly separated by constructing a hyperplane in a d-dimensional space ($d > N$) [158]. Based on this theory, KPCA uses a kernel function to project data onto a kernel Hilbert space whose dimensions are higher than the original ones. Then, the native linear operations of PCA are performed to extract principal components. Let $\Phi(x)$ be a function projecting data onto a high dimensional space. Instead of directly working on the high-dimensional Φ-space, the kernel trick allows non-explicitly calculating high-dimensional projected data with a specified arbitrary kernel function. More specifically, an N-by-N kernel representing the inner product of the feature space is created as

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}) = \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x}) \tag{2.9}$$

  where $\boldsymbol{K}$ is called the kernel function. The common kernel functions include Radial Basis Function (RBF) or Gaussian function, polynomial function, Laplacian function, and sigmoid function. The conventional PCA operation is performed on the dual form of the projected data, which avoids the direct eigen-decomposition of the covariance matrix in the Φ-space. Similar to the conventional PCA operation, the zero-mean condition of the projected data should be guaranteed to perform the principal component analysis. Therefore, the centering operation to the kernel $\boldsymbol{K}$ is also necessary before the PCA operation. We briefly summarize the KPCA calculation procedure as follows.

  1. Step 1 : Construct the kernel matrix $\boldsymbol{K}$ using the Gaussian Kernel as an example

$$\boldsymbol{K}_{ij} = \exp(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2) \tag{2.10}$$

     where γ is the bandwidth parameter of the Gaussian kernel function.

  2. Step 2 : Compute the Gram matrix $\widetilde{\boldsymbol{K}}$

$$\widetilde{\boldsymbol{K}} = \boldsymbol{K} - 2(\mathbf{1}_{1/N}\boldsymbol{K}) + \mathbf{1}_{1/N}\boldsymbol{K}\mathbf{1}_{1/N} \tag{2.11}$$

     where $\mathbf{1}_{1/N}$ is a matrix whose elements are all $1/N$.

3. Step 3 : Perform the eigen-decomposition on the Gram matrix

$$\widetilde{\boldsymbol{K}}\boldsymbol{p}_i = \lambda_i \boldsymbol{p}_i, \quad i = 1, \cdots, N \tag{2.12}$$

4. Step 4 : Calculate the lth principal component as the signal's feature

$$\boldsymbol{z} = \sum_{i=1}^{N} p_{li} \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) \tag{2.13}$$

KPCA is a useful tool to deal with the challenge of non-linear variables in fault diagnosis tasks. The conventional fault detection approach first performs KPCA to obtain principal components in high-dimensional space and then computes the T^2 and SPE statistics for fault detection. To monitor the behavior of a non-linear process, the upper control limit of KPCA analysis is estimated using kernel density estimation techniques [155]. More advanced KPCA-based fault diagnosis approaches have been proposed to improve detection capability via combining it with generalized likelihood ratio test [124], FDA [34], and sliding median filter [218]. However, one major concern about this technique is the high computation complexity when the number of training samples is large [34].

- **Robust principal component analysis (RPCA)** is a modification of the conventional PCA to improve the robustness against outliers in the observed samples. As mentioned before, the conventional PCA performs well with respect to jointly normally distributed data. However, this assumption is not always satisfied in practical applications, such as fault diagnosis for industrial systems [123, 172, 188], video surveillance [10, 19, 43], face recognition [19, 71], and latent semantic indexing [11, 159]. Except for the non-linear nature of data, another potential issue of using the conventional PCA is the effect of outliers contained in observations (also called grossly corrupted observations [19]). Once the observations are corrupted, the variance of measured variables significantly increases, rendering the extracted principal component using the conventional PCA far from the true situation. In the case considering outliers, the data matrix $\boldsymbol{X}$ is assumed as the superposition of two components, namely low-rank component $\boldsymbol{L}_0$ (principal components) and sparse component $\boldsymbol{S}_0$ (outliers part), such that

$$\boldsymbol{X} = \boldsymbol{L}_0 + \boldsymbol{S}_0 \tag{2.14}$$

The RPCA technique is dedicated to recovering the two components individually even though the dimension of the low-dimensional space of $\boldsymbol{L}_0$ and the locations of the nonzero column of $\boldsymbol{S}_0$ are unknown. RPCA can be performed via different techniques like Principal Component Pursuit method (PCP) [19], stable PCP [185], quantized PCP [14], block-based PCP [167], and local PCP [182], where the PCP method can obtain the idealized version of RPCA by solving the following convex problem

$$\begin{aligned} &\text{minimize } \|\boldsymbol{L}\|_* + \eta \|\boldsymbol{S}\|_1 \\ &\text{subject to } \boldsymbol{L} + \boldsymbol{S} = \boldsymbol{X} \end{aligned} \tag{2.15}$$

where $\|\boldsymbol{L}\|_* := \sum_i \sigma_i(\boldsymbol{L})$ denotes the sum of the singular values of the matrix $\boldsymbol{L}$, $\|\boldsymbol{S}\|_1 = \sum_{ij} |\boldsymbol{S}_{ij}|$ denotes the ℓ_1-norm of $\boldsymbol{S}$, and η is a scalar coefficient. There are

also several algorithms that can be used for performing PCP, such as Augmented Lagrange Multiplier Method (ALM) [112], Alternating Direction Method (ADM) [206], Fast Alternating Minimization (FAM) [150], Iteratively Reweighted Least Squares (IRLS) [63, 64] or Alternating Projections (AP) [17, 18, 133], where ALM achieves higher accuracy in fewer iterations and works stably. ALM minimizes the augmented Lagrangian

$$\mathcal{L}(\boldsymbol{L}, \boldsymbol{S}, \boldsymbol{M}) = \|\boldsymbol{L}_* + \eta\|\boldsymbol{S}\|_1 + \langle \boldsymbol{M}, \boldsymbol{X} - \boldsymbol{L} - \boldsymbol{S}\rangle + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{L} - \boldsymbol{S}\|_F^2 \tag{2.16}$$

where $\boldsymbol{M}$ is the Lagrange multiplier matrix updated via

$$\boldsymbol{M}_{k+1} = \boldsymbol{M}_k + \mu(\boldsymbol{X} - \boldsymbol{L}_k - \boldsymbol{S}_k) \tag{2.17}$$

$\|\cdot\|_F$ denotes the Frobenius norm, and μ is the incoherence condition parameter. Solving Eq.(2.16) is simplified by alternately solving the following two problems

$$\arg\min_S \mathcal{L}(\boldsymbol{L}, \boldsymbol{S}, \boldsymbol{M}) = \mathcal{S}_{\eta\mu}(\boldsymbol{X} - \boldsymbol{L} + \mu^{-1}\boldsymbol{M}) \tag{2.18}$$

$$\arg\min_L \mathcal{L}(\boldsymbol{L}, \boldsymbol{S}, \boldsymbol{M}) = \mathcal{D}_{\mu}(\boldsymbol{X} - \boldsymbol{S} + \mu^{-1}\boldsymbol{M}) \tag{2.19}$$

where $\mathcal{S}_\tau(x) = sgn(x)\max(|x| - \tau, 0)$ denotes the shrinkage operator, and $\mathcal{D}_\tau(X) = \boldsymbol{U}\mathcal{S}_\tau(\Sigma)\boldsymbol{V}^T$ denotes the singular value thresholding operator with the singular value decomposition $\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{V}s^T$. We summarize the RPCA calculation procedure as follows.

1. Step 1 : Initialize values $\boldsymbol{S}_0 = 0$, $\boldsymbol{M}_0 = 0$ and set parameter $\mu > 0$
2. Step 2 : Update the low-rank component

$$\boldsymbol{L}_{k+1} = \mathcal{D}_{\mu}(\boldsymbol{X} - \boldsymbol{S}_k + \mu^{-1}\boldsymbol{M}_k) \tag{2.20}$$

3. Step 3 : Update the sparse component

$$\boldsymbol{S}_{k+1} = \mathcal{S}_{\lambda\mu}(\boldsymbol{X} - \boldsymbol{L}_K + \mu^{-1}\boldsymbol{M}_K) \tag{2.21}$$

4. Step 4 : Update the Lagrange multiplier matrix

$$\boldsymbol{M}_{k+1} = \boldsymbol{M}_k + \mu(\boldsymbol{X} - \boldsymbol{L}_{k+1} - \boldsymbol{S}_{k+1}) \tag{2.22}$$

5. Step 5 : Repeat Step 2 to Step 4 until $\boldsymbol{L}_{k+1} + \boldsymbol{S}_{k+1}$ converges to $\boldsymbol{X}$.

RPCA-based fault diagnosis approaches were developed early to cope with outliers. This approach decomposes the data into three parts : low-rank component, sparse component, and dense noise, where the low-rank component mainly represents systematic variation information. By computing singular value decomposition on the low-rank component, this approach obtains the principal components and computes T^2 and SPE statistics for fault detection. Based on this idea, many improvements have been proposed, such as using the Frobenius norm in the optimizing process to accelerate convergence [117] and adopting moments-based criterion in outliers selection to increase fault detection capability [121]. Despite of the significant progress of the RPCA technique in handling outliers, it still suffers from the same limitations as the traditional PCA technique, like insensitive to incipient faults and ineffective to non-linear data.

- **Dynamic Principal Component Analysis (DPCA)** is an extended PCA model for analyzing the dynamic behavior of systems [45]. The statistical process monitoring approach using the conventional PCA technique is usually designed for discrete manufacturing processes with the assumption that observed variables are independent and normally distributed. However, industrial systems do not always operate in steady states but sometimes show a dynamic behavior in which variables are correlated over time [77, 97, 148]. When applying PCA to analyze the static characteristics of dynamic systems, fault diagnosis methods' effectiveness can not be guaranteed since dynamic information is ignored and the time-independence assumption is not satisfied. To take into account the serial correlations of variables, the dynamic version of PCA uses the 'time lag shift' method to augment each sample vector with the previous τ samples. The augmented matrix $\boldsymbol{X}_A$ is generated as

$$\boldsymbol{X}_A(\tau) = \left[\begin{array}{c|c|c|c} \boldsymbol{x}^T(\tau+1) & \boldsymbol{x}^T(\tau) & \cdots & \boldsymbol{x}^T(1) \\ \boldsymbol{x}^T(\tau+2) & \boldsymbol{x}^T(\tau+1) & \cdots & \boldsymbol{x}^T(2) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}^T(N) & \boldsymbol{x}^T(N-1) & \cdots & \boldsymbol{x}^T(N-\tau) \end{array}\right] \tag{2.23}$$

Then the conventional PCA operation is applied to the augmented matrix. We summarize the DPCA calculation procedure as follows.

1. Step 1 : Construct augmented matrix $\boldsymbol{X}_A$ via Eq.(2.23).
2. Step 2 : Calculate the covariance matrix of $\boldsymbol{X}_A$

$$\boldsymbol{\Sigma}_A = \frac{1}{N-\tau-1}\boldsymbol{X}_A^T\boldsymbol{X}_A \tag{2.24}$$

3. Step 3 : Perform the singular value decomposition on the covariance matrix $\boldsymbol{\Sigma}_A$

$$\boldsymbol{\Sigma}_A = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T \tag{2.25}$$

   where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_{m(\tau+1)})$ and $\lambda_1 \geq \cdots \geq \lambda_{m(\tau+1)} \geq 0$.

4. Step 4 : Calculate the first l principal component score of the augmented sample at time index k, such as $\boldsymbol{x}_A(k) = [\boldsymbol{x}(k)\,\boldsymbol{x}(k-1)\,\cdots\,\boldsymbol{x}(k-\tau)]$

$$\boldsymbol{z} = \boldsymbol{x}_A(k)\boldsymbol{P}_{pc} \tag{2.26}$$

   where $\boldsymbol{P}_{pc}$ consists of the first l loading vectors.

The fault detection procedure of the DPCA-based approach is the same as the traditional PCA after the augmented samples are generated. DPCA-based approaches are effective for monitoring processes with an autocorrelation nature [97]. To deal with non-linear and multimodal characteristics, a fault detection method based on DPCA associated with K-nearest neighbors was proposed and outperformed the classical DPCA-based fault detection approach [210]. However, because of the "time lag shift" in constructing augmented samples, the detection delay time is usually long [198]. Moreover, selecting time lag for the DPCA technique is important and tricky, which needs further study.

2.3.1.2 Independent Component Analysis

Derived from the blind source separation problem, Independent Component Analysis (ICA) is a wildly used multivariate statistical tool aimed to decompose a mixed signal into independent source signals. It assumes that at most one source signal is non-Gaussian distributed and the source signals are statistically independent of each other. The famous application of blind source separation is the "cocktail party problem" which focuses on separating persons' speech from a mixed sound signal recorded in a noisy room. Let us consider m independent source signals denoted as $\mathbb{S} = [\boldsymbol{s}_1\, \boldsymbol{s}_1\, \cdots\, \boldsymbol{s}_m]^T$. They are recorded by m sensors, e.g.,. microphones, and mixed as follows,

$$\boldsymbol{X}^T = [\mathrm{x}_1\, \mathrm{x}_2\, \cdots\, \mathrm{x}_m]^T = \boldsymbol{\mathcal{U}}\mathbb{S}, \quad \boldsymbol{\mathcal{U}} \in \mathbb{R}^{m\times m} \tag{2.27}$$

Each observed signal is the weighted sum of the source signals, such as

$$\mathrm{x}_1 = \sum_{j}^{m} a_{1j}\boldsymbol{s}_j \tag{2.28}$$

where a_{ij} is the mixing coefficients (weights). Therefore, the ICA technique is a process searching the unmixing coefficients matrix $\boldsymbol{W}$, where $\boldsymbol{W} = \boldsymbol{\mathcal{U}}^{-1}$ and $\mathbb{S} = \boldsymbol{W}\boldsymbol{X}^T$ ideally. This process is achieved by changing the length or orientation of the weight vector of $\boldsymbol{w}$ till each weight vector is orthogonal to other unmixed signals, according to the independent and non-Gaussian assumptions. In practice, the ICA technique is performed by using different approaches, such as FastICA [79], projection pursuit [80], and Infomax [80]. To maximize the non-Gaussianity, FastICA is commonly used among these approaches for its high efficiency. Hence, we summarize FastICA computational procedures as follows.

1. Step 1 : Whiten samples
$$\widetilde{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^T\boldsymbol{X} \tag{2.29}$$
where $\boldsymbol{P}$ is the orthogonal matrix of eigenvectors of the covariance matrix, and $\boldsymbol{\Lambda}$ is the diagonal matrix of its eigenvalues.
2. Step 2 : Calculate the weight vector $\boldsymbol{w}$ using the Fast ICA algorithm. Iterate the following calculations until they converge
$$\boldsymbol{w}^+ = \mathrm{E}\{\widetilde{\boldsymbol{X}}^T\eth(\boldsymbol{w}^T\widetilde{\boldsymbol{X}}^T)^T\} - \mathrm{E}\{\eth'(\boldsymbol{w}^T\widetilde{\boldsymbol{X}}^T)\}\boldsymbol{w} \tag{2.30}$$
$$\boldsymbol{w} = \boldsymbol{w}^+/\|\boldsymbol{w}^+\| \tag{2.31}$$
where
$$\eth(x) = \tanh(x) \tag{2.32}$$
$$\eth'(x) = 1 - \tanh^2(x) \tag{2.33}$$
the weight vector $\boldsymbol{w}$ is initialized randomly.
3. Step 3 : Repeat Step 2 for m times to calculate all the weight vector $\boldsymbol{w}_j$. Then $\boldsymbol{W} = [\boldsymbol{w}_1\, \cdots\, \boldsymbol{w}_m]$.
4. Step 4 : Calculate the jth source signal as the feature
$$\boldsymbol{z} = \boldsymbol{w}_j^T\widetilde{\boldsymbol{X}}^T \tag{2.34}$$

The typical ICA-based fault detection approaches assume the fault signal mixes healthy independent signals. They extract source signals and the residual error using the ICA technique and calculate their T^2 and SPE statistics as the diagnosis indexes, respectively [198]. The thresholds for decision-making are usually determined by the KDE technique [127]. Lots of fault diagnosis approaches use ICA to extract features. For example, the approach [217] performs the wavelet transform on independent components and calculates JS divergence for fault detection, significantly improving the sensitivity to incipient faults. Fault classification methods using independent components as the input features and using SVM as classifiers can achieve considerable classification accuracy [29]. Compared to the PCA technique, ICA is more suitable for non-Gaussian data but usually suffers from high computational complexity [198].

2.3.1.3 Partial Least Squares

Partial Least Squares (PLS) approach attempts to model a linear relationship between two multivariate data matrices $\boldsymbol{X} \in \mathbb{R}^{N \times m_1}$ and $\boldsymbol{Y} \in \mathbb{R}^{N \times m_2}$ [75, 184]. In contrast to traditional regression approaches, such as the least-squares regression, PLS also models the structure of $\boldsymbol{X}$ and $\boldsymbol{Y}$ by projecting them to a new space. This procedure is to cope with the case that data matrices have a large number of variables that are highly correlated. PLS reduces the dimension of both data matrices and then develops the linear relation of the latent variables. Benefiting from this operation, PLS is wildly used to analyze the data whose observations number is smaller than the number of independent variables [92, 187]. In fact, PLS is highly related to the PCA technique. In the PLS idea, the data matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are first decomposed like in PCA operation, such that

$$\boldsymbol{X} = \boldsymbol{H}\boldsymbol{R}^T \tag{2.35}$$

$$\boldsymbol{Y} = \boldsymbol{J}\boldsymbol{Q}^T \tag{2.36}$$

Then the regression is performed between $\boldsymbol{H}$ and $\boldsymbol{J}$, i.e.

$$\boldsymbol{J} = \boldsymbol{H}\boldsymbol{B} \tag{2.37}$$

Therefore, the relation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is obtained as

$$\boldsymbol{Y} = \boldsymbol{J}\boldsymbol{Q}^T = \boldsymbol{H}\boldsymbol{B}\boldsymbol{Q}^T = \boldsymbol{X}\boldsymbol{R}\boldsymbol{B}\boldsymbol{Q}^T \tag{2.38}$$

The statistical interpretation of PLS is then given in [144, 223], saying that $\boldsymbol{R} = [\boldsymbol{r}_1\, \boldsymbol{r}_2 \cdots \boldsymbol{r}_m]^T$ is the eigenvectors of the covariance matrix of $\boldsymbol{Y}^T\boldsymbol{X}$, which is calculated by solving the following optimization problem

$$\begin{aligned} &\text{maximize } \boldsymbol{r}^T\boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{r} \\ &\text{subject to } \|\boldsymbol{r}\| = 1 \end{aligned} \tag{2.39}$$

The calculation procedure of PLS is summarized as follows.

1. Step 1 : Take the j column of $\boldsymbol{Y}$ as $\boldsymbol{h}$
2. Step 2 : Calculate the vector $\boldsymbol{r}$ and the latent variable $\boldsymbol{h}_j$

$$\boldsymbol{r} = \frac{\boldsymbol{X}^T\boldsymbol{h}}{\|\boldsymbol{X}^T\boldsymbol{h}\|} \tag{2.40}$$

$$\boldsymbol{h} = \boldsymbol{X}\boldsymbol{r} \tag{2.41}$$

3. Step 3 : Calculate the vector q, and update the latent variable $\boldsymbol{h}_j$

$$\boldsymbol{q} = \frac{\boldsymbol{Y}^T\boldsymbol{h}}{\|\boldsymbol{Y}^T\boldsymbol{h}_j\|} \tag{2.42}$$

$$\boldsymbol{h}_j = \boldsymbol{Y}\boldsymbol{q} \tag{2.43}$$

4. Step 4 : Repeat Step 2 to Step 3 until converge ($\boldsymbol{h}_j$ stop changing).
5. Step 5 : Calculate the residual matrices for $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$\boldsymbol{E}_X = \boldsymbol{X} - \boldsymbol{h}_j\boldsymbol{p}^T \tag{2.44}$$

$$\boldsymbol{E}_Y = \boldsymbol{Y} - \boldsymbol{h}_j\boldsymbol{q}^T \tag{2.45}$$

6. Step 6 : Repeat Step 1 to Step 5 l times by setting

$$\boldsymbol{X} := \boldsymbol{E}_X \tag{2.46}$$

$$\boldsymbol{Y} := \boldsymbol{E}_Y \tag{2.47}$$

The l latent variables are obtained as $\boldsymbol{H}_l = [\boldsymbol{h}_1\, \boldsymbol{h}_2\, \cdots\, \boldsymbol{h}_l]$.

Typically, the PLS-based fault detection approach estimates the output $\hat{\boldsymbol{Y}}$ using the developed relation model for the given input data $\boldsymbol{X}$. Then, the error between the estimated output and the real one is computed as a diagnosis index [198]. To allow the use of the information on initial conditions, the work [96] proposed the multiblock multiway PLS for fault diagnosis. On the other side, the dynamic PLS approach was also proposed to improve fault diagnosis capability for dynamic processes [103]. Among those approaches, a system's input and output relation greatly contribute to the fault diagnosis. However, PLS and its improved methods are unavailable if the data can not be separated into input and output, which prevents the use of PLS in many industrial applications.

2.3.1.4 Canonical Variate Analysis

Canonical Variate Analysis (CVA) is a multivariate statistical technique aiming to determine a relationship between two datasets by maximizing their correlation [100]. It is the statistically optimal rank-reduced model for multivariate regression problems and hence wildly used for system identification, filtering, and adaptive control [101]. By considering the past and further information, CVA is also used for dynamic processes monitoring. It maximally correlates the past and future data of healthy cases and determines a reduced-order model for them [139, 142, 154]. Let us consider the following linear model

$$\mathrm{x}(t+1) = \mathrm{A}\mathrm{x}(t) + \mathrm{B}\mathrm{u}(t) + \mathrm{K}\mathrm{v}(t) \tag{2.48}$$

$$\mathrm{y}(t) = \mathrm{C}\mathrm{x}(t) + \mathrm{D}\mathrm{u}(t) + \mathrm{v}(t) \tag{2.49}$$

where A, B, C, D, K are coefficient matrices ; x, y, u, and v are the state vector, output vector, input vector, and process noises, respectively. The calculation procedure of CVA is summarized as follows.

1. Step 1 : Construct the past data vector $\mathbf{p}_k$ and future data vector $\boldsymbol{f}_k$

$$\mathbf{p}_k = \left[\mathbf{u}_{k-1}^T \mathbf{u}_{k-2}^T \cdots \mathbf{u}_{k-L_p}^T \mathbf{y}_{k-1}^T \mathbf{y}_{k-2}^T \cdots \mathbf{y}_{k-L_p}^T\right]^T \in \mathbb{R}^{mL_p} \tag{2.50}$$

$$\boldsymbol{f}_k = \left[\mathbf{y}_k^T \mathbf{y}_{k+1}^T \cdots \mathbf{y}_{k+L_f-1}^T\right] \in \mathbb{R}^{m_y L_f} \tag{2.51}$$

where $m = m_y + m_u$, m_y and m_u are the dimensions of output and input vectors ; L_p and L_f are the lags of the past and future data, respectively.

2. Step 2 : Arrange the past data vector and future data vector into Hankel matrices

$$\boldsymbol{Y}_p = \left[\mathbf{p}_{k+1}^T \mathbf{p}_{k+2}^T \cdots \mathbf{p}_{k+M}^T\right] \in \mathbb{R}^{mL_p \times M} \tag{2.52}$$

$$\boldsymbol{Y}_f = \left[\mathbf{f}_{k+1}^T \mathbf{f}_{k+2}^T \cdots \mathbf{f}_{k+M}^T\right] \in \mathbb{R}^{m_y L_f \times M} \tag{2.53}$$

where $M = N - L_p - L_f - 1$, N is the number of samples.

3. Step 3 : Calculate the sample covariance and the cross-covariance of the past and future samples

$$\boldsymbol{\Sigma}_{pp} = \frac{1}{M-1} \boldsymbol{Y}_p \boldsymbol{Y}_p^T \in \mathbb{R}^{mL_p \times mL_p} \tag{2.54}$$

$$\boldsymbol{\Sigma}_{ff} = \frac{1}{M-1} \boldsymbol{Y}_f \boldsymbol{Y}_f^T \in \mathbb{R}^{m_y L_f \times m_y L_f} \tag{2.55}$$

$$\boldsymbol{\Sigma}_{fp} = \frac{1}{M-1} \boldsymbol{Y}_f \boldsymbol{Y}_p^T \in \mathbb{R}^{m_y L_f \times mL_p} \tag{2.56}$$

4. Step 4 : Perform the singular value decomposition on the scaled Hankel matrix $\boldsymbol{\mathcal{H}}$

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\Sigma}_{ff}^{-1/2} \boldsymbol{\Sigma}_{fp} \boldsymbol{\Sigma}_{PP}^{-1/2} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T \tag{2.57}$$

5. Step 5 : Select the first n canonical variables (CVs) that span the state subspace. Therefore we obtain the reduced matrix $\mathbf{V}_n$ consisting of the first n columns of $\mathbf{V}$.

6. Step 6 : Calculate the state vector $\mathbf{z}_k$ and the model residual vector $\mathbf{e}_k$ for a past data vectors $\mathbf{p}_k$ at a certain time point k.

$$\mathbf{z}_k = \mathbf{V}_n^T \boldsymbol{\Sigma}_{PP}^{-1/2} \mathbf{p}_k \tag{2.58}$$

$$\mathbf{e}_k = (\mathbf{I} - \mathbf{V}_n \mathbf{V}_n^T) \boldsymbol{\Sigma}_{PP}^{-1/2} \mathbf{p}_k \tag{2.59}$$

By calculating the SPE statistic of the residual vector, CVA usually outperforms DPCA and PLS in fault diagnosis tasks since it takes into account the time-correlation information of signals and state equations of a system [139, 154]. Later, the Canonical Variate Dissimilarity Analysis (CVDA) approach was proposed to improve the sensitivity to incipient faults [142]. However, the limitations of CVA-related approaches are the requirement of the known system's state equations and larger detection delay time compared to sample-based approaches, such as PCA, PLS, and ICA.

2.3.2 Machine learning techniques for one-class classification problem

In the big data era, data-driven methods, such as machine learning and deep learning techniques, are increasingly popular and play a key role in industrial applications. In particular, Neural Network-based approaches, e.g. [30, 179], are widely used to solve fault diagnosis problems for their powerful feature representation capability in non-linear cases. However, one major issue of applying data-driven methods to industrial applications is the limited number of high-quality training samples. The good performance of neural network-based approaches highly relies on a mass of labeled data for both healthy and faulty scenarios. This requirement is barely satisfied for most industrial applications since the faulty data of the objected system is scarce or unavailable. Although fault diagnosis methods can be developed based on the knowledge or the data of known faulty behaviors, applying these fault diagnosis methods to unknown faulty behaviors is unacceptable. Therefore, the one-class classification (OCC) problem is raised and discussed to cater to the practical necessity of industrial cases. The OCC problem aims to identify objects of a specific class (the healthy samples in the fault diagnosis scenario) by determining a classification boundary from the training set only containing one class of samples. This problem is more difficult than the traditional classification problem and faces several challenges, like the outlier issue, inadequate sensitivity for small deviation, and the non-linear nature of data.

Typically, three kinds of OCC approaches can be distinguished : *density estimation methods*, *boundary methods*, and *reconstruction methods* [170]. In spite of some overlap among these three kinds of methods, they have different characteristics for solving the OCC problem.

- *The density estimation methods* create a classifier and set a threshold by estimating the density model of samples, like the Elliptic Envelope method [151]. When the model properly fits the data, and the sample number is sufficiently high to ensure the model's parameters can be well estimated, the density estimation methods are good choices for their simplicity and high efficiency.
- *The boundary methods* aim at determining boundaries around target points to distinguish the healthy and faulty cases. Some kind of distance is mainly considered as the description of samples' position in the space, which is then optimized to obtain boundaries. The typical boundary methods are k-centers [201], One-Class Support Vector Machine (OC-SVM) [93], and isolation forest [65, 114].
- *The reconstruction methods*' main idea lies in choosing and fitting a general model which can only describe healthy data in terms of the state of the model but lead to dramatically different outputs for faulty samples. Widely used reconstruction methods are the auto-encoders method [180] using the Neural network technique.

Recently, OCC techniques tend to become promising solutions for fault diagnosis in industrial applications, particularly for the challenging task of incipient fault detection. In this section, we selectively review several wildly used OCC techniques which can be applied to the fault diagnosis problem.

2.3.2.1 One-class support vector machine

The ordinary support vector machine distinguishes between two classes of samples by using a hyper-plane [138]. Similarly, the idea behind the One-Class Support Vector Machine (OC-SVM) is to distinguish the target class from other classes by using a hyper-sphere in the feature space. The hyper-sphere defined by radius r and center c encompasses all the training samples. OC-SVM tries to determine such a hyper-sphere with the minimum radius. The same as the ordinary SVM, OC-SVM also uses the kernel function to project data into a high-dimensional feature space where the samples are separated. Formally, this idea is described as solving the following constrained optimization problem.

$$\begin{aligned} &\underset{r,c}{\text{minimize}}\ r^2 \\ &\text{subject to } \|\Phi(x)_i - c\|^2 \leq r^2, \quad \forall i = 1, 2, \cdots, N \end{aligned} \tag{2.60}$$

where $\Phi(x)$ is the non-linear transformation usually evaluated by a kernel function, such that $\kappa(x_i, x_j) =< \Phi(x_i), \Phi(x_j) >$.

However, the constraint of the above formulation is highly restrictive, resulting in a poor accuracy performance and weak robustness to outliers. Therefore, a flexible version of OC-SVM is proposed to tolerate the presence of outliers in the training set [157, 171], which is given as

$$\begin{aligned} &\underset{r,c,\zeta}{\text{minimize}}\ r^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\zeta_i \\ &\text{subject to } \|\Phi(x)_i - c\|^2 \leq r^2 + \zeta_i, \quad \forall i = 1, 2, \cdots, N \end{aligned} \tag{2.61}$$

where $\nu > 0$ is a parameter balancing the importance of sphere volume and the number of outliers. The optimization problem involves finding out the radius, center, and a set of slack variables ζ_i. Then by using the Karush-Kuhn-Tucker (KKT) optimality conditions, the center c is obtained as

$$c = \sum_{i=1}^{N}\alpha_i\Phi(x_i) \tag{2.62}$$

where α_i is obtained by solving the following optimization problem

$$\begin{aligned} &\underset{\alpha}{\text{maximize}}\ \sum_{i=1}^{N}\alpha_i\kappa(x_i, x_i) - \sum_{i,j=1}^{N}\alpha_i\alpha_j\kappa(x_i, x_j) \\ &\text{subject to } \sum_{i=1}^{N}\alpha_i = 1 \text{ and } 0 \leq \alpha_i \leq \frac{1}{\nu N}, \quad \forall i = 1, 2, \cdots, n \end{aligned} \tag{2.63}$$

According to the value of α, each sample belongs to one of three cases : sample is inside the hyper-sphere if $\alpha = 0$; sample is outside the hyper-sphere if $\alpha = \frac{1}{\nu N}$; sample lies on the boundary if $0 < \alpha < \frac{1}{\nu N}$. The support vectors are defined as the samples with non-zero α, which composes of the set $\mathcal{I}_{sv}$. The optimal radius is determined as $r = \|\Phi(x_i) - c\|$ with x_i lying on the boundary, where the distance is computed as follows.

$$\|\Phi(x) - c\|^2 = \sum_{i,j\in\mathcal{I}_{sv}}\alpha_i\alpha_j\kappa(x_i, x_j) - 2\sum_{i\in\mathcal{I}_{sv}}\alpha_i\kappa(x_i, x) + \kappa(x, x) \tag{2.64}$$

When $\|\Phi(x_i) - c\| < r$, one can make a decision that sample x_i belongs to the target class.

In the fault diagnosis domain, the OC-SVM technique allows to detect faults without using faulty samples, which is more practical for industrial applications. Two major advantages of OC-SVM for fault diagnosis tasks can be highlighted : (i) By using the kernel trick, this technique can effectively handle complex data distribution ; (ii) It also has strong robustness against outliers by introducing slack variables in the optimization procedure. Therefore, it has achieved high fault detection accuracy in different industrial scenarios, such as mechanical systems [52], chemical processes [7], and power systems [78]. However, the effectiveness of OC-SVM approaches for incipient diagnosis is still unclear, there are still waiting for further investigation.

2.3.2.2 k-centers

The k-center approach was originally proposed to deal with challenging problems in the fault detection task [201], such as applications on high-dimensional data with small sample sizes and dynamical systems. The idea of the k-centers approach is to find a simple description of a domain that only indicates the healthy behavior of systems. Instead of considering the density of samples, this approach extracts critical samples from the training set as supporting points of the healthy domain. The selection of k support objects, denoted as β_i, is called the domain approximation, which actually determines k centers of small hyper-spheres with equal radius r. The objective of the k-center approach is to minimize the radius while all the training samples are included in the healthy domain. Formally, the maximum of the minimum distances between training samples and the support objects is minimized, such as

$$\text{minimize } r = \max_i(\min_k \|\boldsymbol{x}_i - \boldsymbol{\beta}_k\|^2) \tag{2.65}$$

where the Euclidean distance is used as the distance measure. The following procedure is performed to search for the optimal support objects.

1. Randomly choose k samples from the training set as the support objects.
2. Calculate the radius r using Eq.(2.65).
3. If the radius r is smaller than the previous one, update the support objects.
4. Repeat the above steps until radius r cannot decrease.

After determining the support objects, the distance of a new sample is calculated as.

$$d_k(\boldsymbol{x}) = \min_k \|\boldsymbol{x} - \boldsymbol{\beta}_k\|^2 \tag{2.66}$$

When $d_k(\boldsymbol{x}) > r$, sample $\boldsymbol{x}$ is recognized as a faulty sample (not belonging to the healthy class). Obviously, although the k-center approach is simple to perform, its performance is easily affected by the outliers mixed into the training samples [197]. Since outliers are far away from most training samples, a larger radius will be obtained to include all the training samples.

2.3.2.3 Auto-encoder

Autoencoder is an artificial neural network aiming to learn the concise representation of original input data by regenerating the output data from the input one [25, 173]. There

are two processes of the autoencoder in the training stage : encoding and decoding. The encoding part attempts to transform input data to a low dimensional vector so as to refine the data and preserve the critical information. In this process, data are mapped into a feature (coding) space using nonlinear transformation. The coding is an effective representation containing the necessary information to recover input data, and thereby is used as a feature for dimensionality reduction. The decoding process, recovering the original input data based on the coding vectors, is the inverse operation of the previous step. The same as the encoding process, a nonlinear function is used in the decoding step. The objective of the autoencoder is to minimize the difference between the input and output of the model.

Autoencoder is performed as a feedforward artificial neural network constructed by input layer, hidden layers, and output layer. Given the input data $\boldsymbol{X} \in \mathbb{R}^m$, the encoder is a nonlinear function $G : \mathbb{R}^m \to \mathbb{R}^l$ $(l < m)$, such that

$$\boldsymbol{Z} = G(\boldsymbol{X}), \quad \boldsymbol{Z} \in \mathbb{R}^l \tag{2.67}$$

Similarly, the decoder is another nonlinear function $G' : \mathbb{R}^l \to \mathbb{R}^m$, such that

$$\hat{\boldsymbol{X}} = G'(\boldsymbol{Z}) \tag{2.68}$$

As pointed by Cybenko that any nonlinear function can be fitted by the functions with following form to an arbitrary degree of precision [35]

$$\boldsymbol{Z} = \sigma(\boldsymbol{\omega}\boldsymbol{X} + \boldsymbol{b}) \tag{2.69}$$

where σ is a continuous and monotonically increasing function (activation function) usually selected as the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.70}$$

$\boldsymbol{\omega}$ is the weight matrix and $\boldsymbol{b}$ is the bias vector. In the training procedure, these two parameters are updated iteratively through the backpropagation of the reconstructed error ϵ.

$$\epsilon(\boldsymbol{X}, \hat{\boldsymbol{X}}) = \|\boldsymbol{X} - \hat{\boldsymbol{X}}\|^2 = \|\boldsymbol{X} - G'(G(\boldsymbol{X}))\|^2 \tag{2.71}$$

Then the decoder is represented as

$$\hat{\boldsymbol{X}} = \sigma(\boldsymbol{\omega}'\boldsymbol{Z} + \boldsymbol{b}') \tag{2.72}$$

where $\boldsymbol{\omega}'$ and $\boldsymbol{b}'$ are also updated individually in the training procedure. If the reconstructed value of new sample $\boldsymbol{X}$ is much different from the original one, i.e.

$$\epsilon(\boldsymbol{X}, \hat{\boldsymbol{X}}) > th \tag{2.73}$$

where th is a decision threshold, this sample likely does not belong to the target class.

Due to the powerful feature learning ability, the autoencoder is commonly applied to fault diagnosis tasks and realizes promising performance in these tasks. [129, 160]. However, for the small difference between incipient faults and healthy samples, the healthy region developed by AE seems not accurate enough [195]. The sensitivity of AE for incipient faults diagnosis needs to be improved.

2.3.2.4 Isolation Forest

Isolation Forest (IF) is a non-parametric and ensemble-based OCC approach with linear time complexity and high accuracy, which can handle large amount of data [65, 114, 115]. It uses the spatial partitioning strategy to isolate anomalies (faulty samples) rather than profiles the healthy training samples like the above OCC approaches. This approach uses binary trees to separate training samples, where anomalies have a shorter partitioning path in a tree structure since points with low density are easily located with fewer partitioning. It means that the depth of instances (partitioning path) in the developed binary tree indicates the sparsity of data. Based on this isolation characteristic, IF can detect anomalies with few false alarms.

IF uses the ensemble strategy to obtain a convergence result. n binary trees of the IF are structured by recursively separating data into two parts (left child and right child) with the randomly selected attribute x_i of the data and a split value th until the following conditions are satisfied : (i) the tree reaches a height limit ; ii) only one sample in each subspace. The algorithm becomes more stable as the number of trees increases. After the training procedure is completed, a score indicating the degree of anomaly is required. As mentioned before, the path length of data is a crucial value in evaluating the degree of anomaly. Let $h_p(x)$ be the path length of sample x, then the average path length is estimated as

$$\Gamma(n) = 2H_p(n_e - 1) - (\frac{2(n_e - 1)}{n_e}) \tag{2.74}$$

where n_e is the number of external nodes and$H_p(\cdot)$ is the harmonic number that can be calculated as $H_p(n_e) = \ln(n_e) + 0.5772$ (Euler's constant). Then the anomaly score of sample x given n_e is defined as

$$\mathrm{s}(x, n_e) = 2^{-\frac{E(h_p(x))}{\Gamma(n_e)}} \tag{2.75}$$

where $E(h_p(x))$ is the average path length of n isolation trees. The anomaly score takes values in the range of $[0, 1]$, where the value of 1 indicates the high possibility of the sample being anomalous, while the value of 0 means the high possibility of normal (healthy) sample.

Although the IF technique can be used to detect a fault, it is not a good solution to detect incipient faults since it is only effective for samples showing large deviations. Therefore, it is common to use IF to remove outliers from raw data and then apply other techniques for fault detection [87, 209].

2.3.3 Distance measure for fault diagnosis

The distance measure is a key for summarizing the dissimilarity between the healthy reference feature and the testing feature. Then based on the distance measure, a final decision is made for the fault detection task by comparing the analysis index with a detection threshold. In this section, we consider two distributions P and Q, whose Probability Density Functions (PDF) are f and g, and their Cumulative Density Function (CDF) are $\mathbb{F}$ and $\mathbb{G}$, respectively.

2.3.3.1 Divergence

Divergence is a kind of statistical distance measuring the closeness (similarity) between two probability distributions. It is different from metric (a general notion of physical distance), which is a non-negative real-valued function satisfying the identity of indiscernibles, symmetry, and the triangle inequality. Divergences D satisfy the non-negativity and identity of indiscernibles, such that

$$D(f||g) \geq 0, \text{for all } f \text{ and } q \tag{2.76}$$

$$D(f||g) = 0, \text{if and only if } f = g \tag{2.77}$$

However, divergences do not always satisfy symmetry. Therefore divergence should be specified as "from f to g" instead of "between f and g". In addition, divergences do not satisfy triangle inequality since it is a positive-definite quadratic form (squared distance). Here, we briefly review several important and widely mentioned divergences in fault diagnosis applications, such as Kullback–Leibler divergence, Jensen-Shannon divergence, and f-divergence.

- **Kullback-Leibler divergence** (KLD) measures the loss of information when replacing a probability with another one. It was originally introduced by Solomon Kullback and Richard Leibler in [98] and referred to as relative entropy. The definition of the relative entropy $I(f||g)$ is

$$I(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx \tag{2.78}$$

  For most practical applications, the symmetric version of relative entropy is more useful and convenient for calculation. Hereafter, we call its dual form as Kullback-Leibler divergence, which is defined as

$$D_{kl}(f||g) = I(f||g) + I(g||f) \tag{2.79}$$

- **Jensen-Shannon Divergence** (JSD), also known as information radius (IRad) [136] or total divergence to the average [36], was developed as a symmetrized and smoothed version of relative entropy. The Jensen-Shannon divergence is defined as

$$D_{js}(f||g) = \frac{1}{2}(I(f||\mathcal{M}) + I(g||\mathcal{M})) \tag{2.80}$$

  where $\mathcal{M} = \frac{1}{2}(f + g)$ is the mixture distribution of f and g. Jensen-Shannon divergence has a finite value with an upper bound of $log_\rho(2)$ in base ρ, i.e.

$$0 \leq D_{js}(f||g) \leq log_\rho(2) \tag{2.81}$$

- **f Divergence** is a general form of many divergences, such as Kullback-Leibler divergence, Jensen-Shannon divergence, and total variation distance [42]. Given a convex function, such that $F : [0, \infty) \to (-\infty, \infty]$ and $f(1) = 0$, f-divergence is defined as

$$D_F(f||g) = \int F(\frac{f(x)}{g(x)}) g(x) dx \tag{2.82}$$

  f divergence satisfies linearity, i.e.

$$D_{\sum_i a_i F_i} = \sum_i a_i D_{F_i} \tag{2.83}$$

  Some examples of f divergence are given in Table 2.1.

TABLE 2.1 – Examples of f divergence

Function F	Divergence
$x\log(x)$	Relative entropy [4]
$(x-1)log(x)$	KL divergence [4]
$-(x+1)log(\frac{1+x}{2})+xlog(x)$	JS divergence [4]
$\|x-1\|$	Total variation distance [42]
$\frac{1}{2}(\sqrt{x}-1)^2$	Squared Hellinger distance [73]
$(x-1)^2$	Pearson's chi-square divergence [143]

Recently, divergence metric is prevalent in fault diagnosis domain, serving as the diagnosis index [40, 66, 68, 69, 217]. Compared to typical indexes, namely T^2 and SPE statistics, the benefits of using divergence metrics are the high sensitivity to tiny deviations of data and the strong robustness to noise. However, the divergence calculation is based on the features' PDF, which should be estimated accurately. Kernel Density Estimator (KDE) is a common non-parametric approach for estimating PDF from samples, but it is usually vulnerable to the setting of parameters and the number of samples. Particularly, the divergence metric can not describe the distance of two sample sets when their probability distributions are far away [107, 113]

2.3.3.2 Kolmogorov Smirnov distance

Kolmogorov–Smirnov (KS) test is a non-parametric test considering the equality of two probability distributions [16, 119]. In practice, it is used to test whether a given sample is drawn from the reference probability distribution, which is called the one-sample KS test. When we consider two sample sets, the KS test is performed to tell if the two sample sets are drawn from the same distribution, which is called the two-samples test. The one-sample KS test produces a distance evaluating the difference between the sample's empirical probability distribution $\mathbb{F}_n$ and the CDF of the reference distribution $\mathbb{G}(x)$, such as

$$D_{ks} = \max|\mathbb{F}_n(x) - \mathbb{G}(x)| \tag{2.84}$$

The empirical probability distribution $\mathbb{F}_n(x)$ of sample x is calculated by

$$\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{[-\infty,x]}(x_i) \tag{2.85}$$

where $\mathcal{I}_{[-\infty,x]}(x)$ is an indicator function, such that

$$\mathcal{I}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \tag{2.86}$$

In the two-samples case of the Kolmogorov Smirnov test, the distance between two CDFs is calculated as follows,

$$D_{ks} = \max|\mathbb{F}(x) - \mathbb{G}(x)| \tag{2.87}$$

KS distance is also widely used in fault diagnosis tasks [60, 178]. In contrast to divergence metrics, KS distance is computed based on features' CDF, which is easy to estimate empirically. Therefore, using KS distance as the diagnosis index can reflect the dissimilarity between two sample sets and avoid the influence of external parameters. However, KS distance is usually not as sensitive as divergence metrics [215].

2.3.3.3 Wasserstein distance

The Wasserstein distance, also known as Kantorovich–Rubinstein metric, is used to quantify the distance between two probability distributions in a given metric space. Intuitively, the distance measurement can be viewed as a transportation problem that tries to find an optimal solution with a minimum cost to transport one pile of the earth (distribution) to another. Because of this analogy, Wasserstein distance, in some cases, is also called Earth Mover Distance (EMD), and the optimal transport theory provides efficient algorithms to calculate the distance.

Let $\mathcal{J}(\mathbb{F}, \mathbb{G})$ denote the collection of all the joint distributions with the marginals F and G. The p^{th} $(p \geq 1)$ Wasserstein distance between P and Q is :

$$D_w(P, Q) = \left(\inf_{J \in \mathcal{J}(P,Q)} \int \|x - y\|^p d\mathcal{J}(x, y) \right)^{1/p} \tag{2.88}$$

According to the optimal transport theory, the p^{th} Wasserstein distance between P and Q on the measure space $\mathbb{R}$ is :

$$D_w(P, Q) = \left(\int_0^1 |\mathbb{F}^{-1}(x) - \mathbb{G}^{-1}(x)|^p dx \right)^{1/p} \tag{2.89}$$

Let $P = \mathcal{N}(m_1, C_1)$ and $Q = \mathcal{N}(m_2, C_2)$ be two non-degenerate Gaussian distributions with the mean center m_1 and m_2, and symmetric positive semi-definite covariance matrices C_1 and C_2, the 2-Wasserstein distance between P and Q is

$$D_w(P, Q) = \|m_1 - m_2\|^2 + \text{trace}(C_1 + C_2 - 2(C_2^{1/2} C_1 C_2^{1/2})^{1/2}) \tag{2.90}$$

Wasserstein distance is increasingly used in fault diagnosis tasks for its simple calculation and reliable performance [31, 222]. Unlike divergence metrics, Wasserstein distance can measure the distance between two sample sets even if their distributions are far from each other. It is more reliable than KS distance since it considers the global distance of two distributions instead of the local maximum. However, this measurement's sensitivity to incipient faults needs further investigation and comparison with divergence metrics, though they were partly discussed in work [215].

2.4 Conclusion

Fault diagnosis and prognosis are increasingly important for modern complex systems and become hot topics in the signal processing domain with wide applications. This chapter recalls the fault's definition and then classifies faults into three classes : abrupt fault, intermittent fault, and incipient fault, according to their time dependencies. Abrupt and intermittent faults are usually more pronounced and relate to severe performance degeneration of systems for their large fault amplitude or long duration. Incipient fault does not dramatically affect systems' performance at its early stage but may result in severe system failures later. Early diagnosis of incipient fault is crucial to discover the potential risk of system breakdown. However, this kind of fault creates tiny deviation that

can easily be confused with noise or other disturbances, which is more challenging than others.

According to different goals, this chapter summarizes three tasks for the fault diagnosis problem : namely fault detection, faulty sources isolation, and fault severity estimation. Fault detection is the first step and also the foundation of fault diagnosis. Subsequently, the fault isolation and fault severity estimation procedures begin to locate the key factors affecting the systems' performance and assess the severity of faults. Over the last few decades, a large number of fault diagnosis approaches have been studied and presented, where the data-driven approaches attract more attention than model-based approaches for their flexibility and robustness. Accordingly, this chapter reviews the development of data-driven approaches and summarizes them into four steps : pre-processing, feature extraction, feature analysis, and decision-making. Feature extraction and feature analysis steps are two critical parts of the data-driven approaches. Signal processing techniques, statistical techniques, and machine learning approaches are frequently used for feature extraction. This chapter reviews state-of-the-art techniques for feature extraction, including multivariate statistical and one-class classification techniques. Regarding the feature analysis step, this chapter pays more attention to the probability-based metric as the state-of-the-art analysis tools, such as divergence, Kolmogorov Smirnov distance, and Wasserstein distance.

Despite the significant progress, there are some challenges in the fault diagnosis problem, such as the unsatisfying performance of the existing approaches for non-Gaussian distributed data, non-linear processes, and dynamic systems. In addition, the low sensitivity for tiny deviation, the effect of outliers in the training samples, and the insufficient number of faulty samples are also great challenges for fault diagnosis of complex systems. These challenges motivate us to propose advanced feature extraction and analysis techniques. Recently, OCC techniques aiming to identify one class of samples tend to become promising solutions for fault diagnosis in industrial applications, particularly for the challenging task of incipient fault detection. Therefore, the rest of this work will discuss how to apply the OCC idea to develop fault diagnosis approaches.

3

Healthy region approximation for fault diagnosis

3.1 Introduction

This chapter aims to discuss the benefit of using the One Class Classification (OCC) idea to tackle the challenges faced by the existing fault diagnosis methodologies, such as the low sensitivity to tiny deviations, poor performance for non-Gaussian distributed data, non-linear processes, and dynamic systems [198]. In the OCC idea, the fault detection task is regarded as a classification problem, and the training process only relies on healthy data. More precisely, OCC approaches are devoted to developing a healthy region for the fault detection task. Subsequently, the obtained healthy region is used to decide whether samples are healthy or not. If samples are in the healthy region, they are recognized as healthy, otherwise faulty. Therefore, the accuracy of the healthy region is crucial to the fault detection task. Due to the complex data structure, traditional approaches, like PCA-based methods [56, 89], usually fail to develop an effective healthy region. Moreover, the detection of tiny deviations places a higher demand on the accuracy of healthy regions. To address these challenges, this chapter first formulates the healthy region approximation problem and consider ineffective examples of healthy region. Then, we propose our solution for healthy region approximation by using the Local Mahalanobis Distance (LMD). The benefits of using LMD for healthy region approximation, such as high accuracy, effectiveness for complex data, and robustness against outliers, are highlighted by comparing the performance of the proposed method with the other OCC approaches. After obtaining an accurate healthy region, fault detection, faulty variables isolation, and fault severity estimation approaches can be developed.

3.2 Problem formulation

Let's consider a sample vector $\boldsymbol{x} \in \mathbb{R}^m$ with m variables representing features or original signals, for example, electrical signal, temperature, pressure, and concentration. In the fault diagnosis context, it is natural to first determine if $\boldsymbol{x}$ is healthy or not. As reviewed in the last chapter, fault detection approaches can be simplified as a linear or non-linear function $\mathcal{B} : \mathbb{R}^m \to \mathbb{R}^1$, such that

$$\mathcal{B}(\boldsymbol{x}) \leq th, \text{ for healthy samples}$$
$$\mathcal{B}(\boldsymbol{x}) > th, \text{ for faulty samples}$$

where th is the detection threshold. From the view of spatial distribution, $\boldsymbol{x}$ belongs to one of the two regions in m dimension space. They are the healthy region denoted as Ω_H, and the faulty one denoted as Ω_F, where $\Omega_H \cap \Omega_F = \emptyset$ and $\Omega_H \cup \Omega_F = \mathbb{R}^m$. From this point of view, fault detection turns to determine such a separating hyperplane $\mathcal{B}(\boldsymbol{x}) - th = 0$ satisfying

$$\mathcal{B}(\boldsymbol{x}) - th \leq 0, \boldsymbol{x} \in \Omega_H$$
$$\mathcal{B}(\boldsymbol{x}) - th > 0, \boldsymbol{x} \in \Omega_F$$

The output of function $\mathcal{B}$ serves as a detection index and is also expected to contain significant information about the fault, such as its root causes and amplitude. Subsequently, faulty variables isolation and fault severity estimation approaches are developed based on the detection index.

Although supervised approaches, such as Support Vector Machines (SVM), can effectively achieve an optimal separation based on negative and positive samples [116], it is impractical for industrial applications. The lack of sufficient faulty samples in the training step induces the more challenging OCC problem. A primary solution for the OCC problem is to obtain the description of the healthy region in feature space based on historical fault-free samples, which is called healthy region approximation. For example, a hypersphere in a m dimension space is commonly used to approximate the healthy region Ω_H. Common statistics criterion, namely Hotelling's T^2 and SPE(Q), and their mixture index can be used as examples of healthy region approximation in the PCA framework. Particularly, Hotelling's T^2 is the particular case of the SPE(Q) statistics normalizing each dimension [56, 89]. These three statistics criteria can be expressed in the form of separating hyperplane as follows

- Hotelling's T^2 : $\mathcal{B}(\boldsymbol{x}_{pc}) = \boldsymbol{x}_{pc}\Lambda_{pc}^{-1}\boldsymbol{x}_{pc}^T$
- SPE(Q) : $\mathcal{B}(\boldsymbol{x}_{res}) = \boldsymbol{x}_{res}\boldsymbol{x}_{res}^T$
- Mixture index : $\mathcal{B}(\boldsymbol{x}) = \frac{\boldsymbol{x}_{pc}\Lambda_{pc}^{-1}\boldsymbol{x}_{pc}^T}{th_{T^2}} + \frac{\boldsymbol{x}_{res}\boldsymbol{x}_{res}^T}{th_{SPE}}$

where $\boldsymbol{x}_{pc}$ and $\boldsymbol{x}_{res}$ are transformed data with zero mean in principal subspace and residual subspace, respectively; $\Lambda_{pc} = \text{diag}(\lambda_1, \cdots, \lambda_j, \cdots, \lambda_l)$ is the corresponding eigenvalues matrix whose element λ_j is also the variance of the j^{th} projected data in the principal space. th_{T^2} and th_{SPE} are the thresholds of Hotelling's T^2 and SPE statistics, respectively. However, Hotelling's T^2, SPE, and their mixture are only effective for elliptically distributed data, e.g., multivariate Gaussian distribution and multivariate t-distribution. Two toy examples of input data are shown in Fig.3.1, where the data in Fig.3.1 (a) (b) (c) are assumed as Gaussian distributed and data in Fig.3.1 (d) (e) (f) are non-Gaussian. Hotelling's T^2, SPE, and their mixture index are used to develop a healthy region for the input data, respectively.

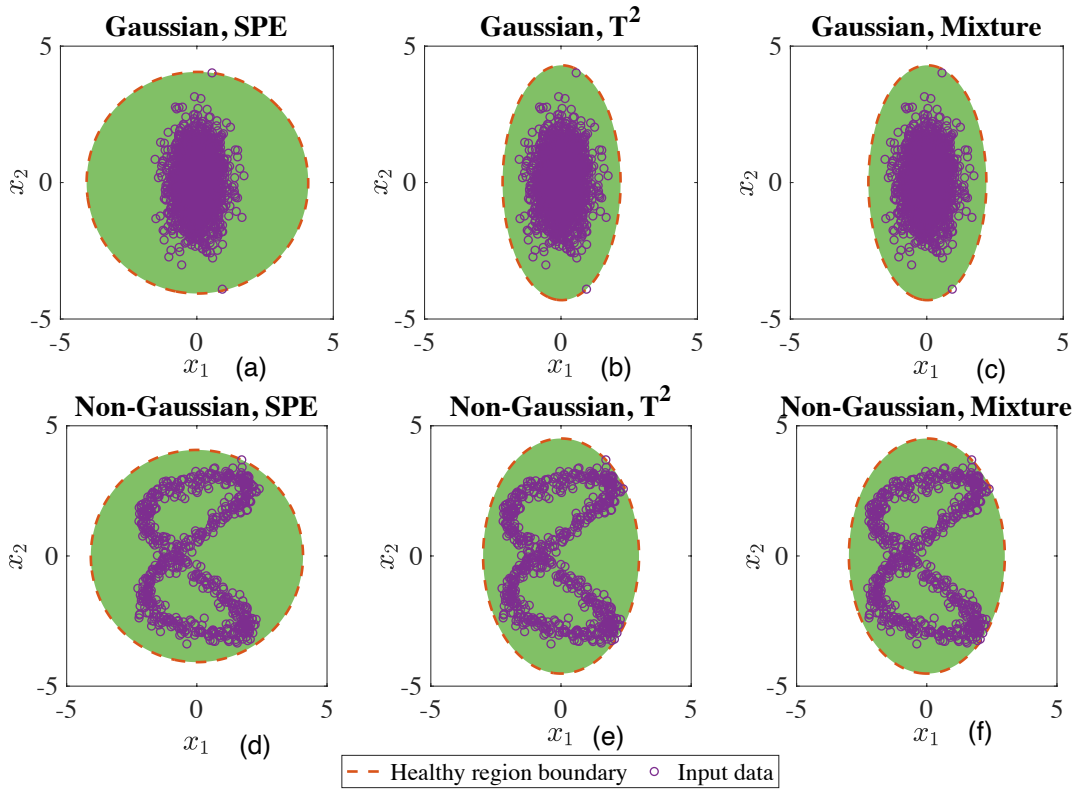

FIGURE 3.1 – Healthy region approximation for the projected data : (a) SPE statistic for Gaussian distribution ; (b) Hotelling's T^2 for Gaussian distribution ; (c) Mixture index for Gaussian distribution ; (d) SPE statistic for non-Gaussian distribution ; (e) Hotelling's T^2 for non-Gaussian distribution ; (f) Mixture index for non-Gaussian distribution ;

- In Fig.3.1 (a), the input data is Gaussian distributed with different variances. SPE statistic approximates the healthy region as a circle without normalizing each dimension. Since the variance of each dimension is different, the developed healthy region is too large in the horizontal direction.
- Fig.3.1 (b) using the same data as Fig.3.1 (a). Hotelling's T^2 creates an ellipse as the healthy region, which can better fit the data. Therefore this statistical criteria is effective for healthy region approximation of Gaussian distributed data.
- In Fig.3.1 (c), the result is close to Fig.3.1 (b). Therefore the mixture index is effective for Gaussian distribution.
- Fig.3.1 (d) shows that when the input data is non-Gaussian, SPE statistic is still ineffective for the too large healthy region.
- In Fig.3.1 (e), although the healthy region developed by Hotelling's T^2 is more compact than by SPE statistic, it is too large to describe the healthy region efficiently. As a result, Hotelling's T^2 and SPE statistics are ineffective for non-Gaussian distribution.
- In Fig.3.1 (f), the result is similar to Fig.3.1 (e), meaning that the mixture index is also ineffective for non-Gaussian distribution.

Besides evaluating the performance of obtained healthy regions in a graphical way, we now introduce the following two numerical criteria to show their effectiveness. On the one hand, the primary goal of healthy region approximation approaches is to develop a region to include as many training fault-free samples as possible. Therefore, the accuracy

criterion ACC_H is defined as

$$\text{ACC}_H = \frac{\text{Number of samples in the region}}{\text{Number of fault-free samples}} \times 100\% \tag{3.1}$$

A larger ACC_H (closer to 100%) indicates better performance for healthy region approximation in terms of accuracy. On the other hand, a good healthy region is expected to be tighter, meaning that it should precisely fit the shape of data distribution instead of being too large. To evaluate the fitness of a healthy region to fault-free samples, we consider the Intersection over Union (IoU) criterion, which is widely used in the image object detection task [57, 58]. We regard fault-free samples as objects and the healthy region as their detected area. The IoU criterion is defined as the ratio of the intersection and the union of objects' area to the detected area, such that

$$\text{IoU} = \frac{\text{Area}_{\text{inter}}}{\text{Area}_{\text{union}}} \times 100\% \tag{3.2}$$

To calculate the area of objects and healthy region, we first transform a point into a small cell of a grid. Then we can compute the objects' area by locating all the corresponding cells of fault-free sample points and calculate the area of healthy region in the same way. The intersection and union areas are easily determined and the IoU criterion can be calculated. Obviously, a large IoU value indicates a good approximation performance. When the healthy region perfectly fits fault-free data, IoU is equal to 100%.

As shown in Table 3.1, although all approaches achieve 100% accuracy, their IoU performance is depressing. Generally, a high accuracy performance but a low IoU value indicates the developed healthy region is significantly larger than the data distribution. For both Gaussian and non-Gaussian data, T^2 and the mixture index have similar performance, and they are better than the SPE index. This example shows that a healthy region approximation using a hypersphere, such as a circle and an ellipse in 2-dimensional space, is sometimes not accurate enough for irregular distribution. Although hypersphere is still an effective model to describe a healthy region by transforming original data into features grouped at a center, such transformation is always difficult to obtain efficiently. This issue motivates us to discuss a more flexible way for healthy region approximation in the next section.

TABLE 3.1 – Healthy region approximation performance of Hotelling's T^2 and SPE indexes in the PCA framework for the example given in Fig.3.1

Methods	Gaussian		Non-Gaussian	
	Acc_H (%)	IoU (%)	Acc_H (%)	IoU (%)
SPE	100	28.13	100	31.99
T^2	100	37.04	100	35.32
Mixture	100	37.06	100	35.61

3.3 Healthy region approximation with multiple centers

A common way to develop a healthy region for the given fault-free samples is to determine a hypersphere with a center and radius, as shown in the last section. When

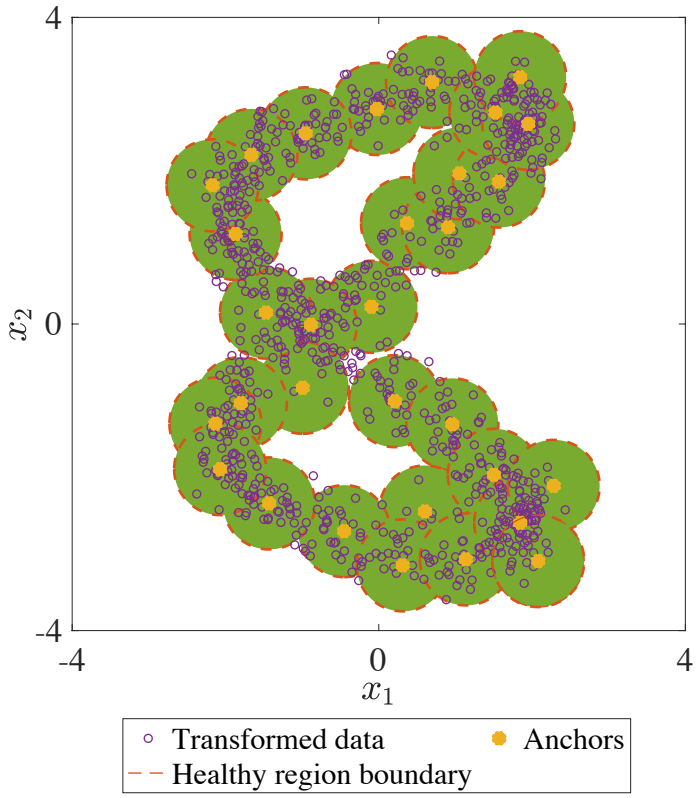

FIGURE 3.2 – Proposed multiple centers idea for healthy region approximation

samples are irregular in their original space, they should be projected to a new space where a hypersphere effectively encompasses training samples. A typical example of this idea is the KPCA with Hotelling's T^2 statistic. Alternatively, the more straightforward idea is to focus on the local information of samples' spatial position. More precisely, in this work, we propose that each local region of training samples is modeled as a hypersphere, and the healthy region is composed of multiple hyperspheres with different centers and radius. However, this naive idea leads to a high-computational issue for a large number of parameters. To simplify the problem, we assign the same radius but determine different centers for these hyperspheres. We also allow the hyperspheres to overlap each other. As illustrated in Fig.3.2, with the position of centers and the radius's value being properly specified, the overlapping hyperspheres can perfectly cover all the fault-free samples. These centers, denoted as $\boldsymbol{C}$ and referred to as anchors hereafter, contain essential spatial information about the healthy region. Likewise, the radius denoted as r_L specifies the margin of a local region. In other words, an approximate healthy region is completely determined by an anchor set $S = \{\boldsymbol{C}_1, \cdots, \boldsymbol{C}_k, \cdots\}$, $S \subseteq \Omega_H$ and the healthy region margin r_L.

Since this idea divides the healthy region into multiple small local regions, selecting an appropriate size of local region is crucial to obtain an accurate description of Ω_H. Indeed, a smaller size can help to improve the accuracy of the description by preserving most details. An extreme case is $r_L \to 0$, where all the training samples are taken as centers of regions. Although all samples' information is preserved in this case, any tiny difference between a new sample and the training samples is not allowed, resulting in the low generalizability of the model. This situation is also referred to as the over-fitting issue in the machine

learning domain. In contrast, if the size is too large, the developed model is inflexible to fit the training samples. The extreme case is the single hypersphere case, which is equivalent to Hotelling's T^2 statistic. In order to obtain reasonable sizes of local regions, a distance measurement should be first defined to decide which samples are close and should be regarded as neighbors. This concern motivates us to propose a particular distance named the local Mahalanobis distance (LMD), based on which the anchors-generation and region margin selection approaches are developed to determine a healthy region.

3.3.1 Local Mahalanobis distance calculation

In the concept of healthy region approximation, distance measure [12] plays a central role, motivating the following discussion on the distance measure. In multivariate applications, Mahalanobis distance attracts more interest than Euclidean distance because the former is unitless and scale-invariant and considers the correlations of variables. Let us suppose that $\boldsymbol{x}, \boldsymbol{y}$ are two sample vectors generated from the same distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. Then the multivariate Mahalanobis distance of $\boldsymbol{x}$ with respect to $\boldsymbol{\mu}$ is defined as :

$$d_M(\boldsymbol{x}, \boldsymbol{\mu}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})} \tag{3.3}$$

The Mahalanobis distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$d_M(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})} \tag{3.4}$$

By using this distance, data is first transformed to a new space where axes are independent and then re-scaled to unit variance. So, the Mahalanobis distance is equivalent to a standard Euclidean distance in the transformed space. There are two main properties of the Mahalanobis distance for Gaussian distributed data : (i) the probability density of samples uniquely determines the Mahalanobis distance $d_M(\mathbf{x}, \boldsymbol{\mu})$; (ii) its squared value d_M^2 follows a chi-squared distribution. However, for non-Gaussian distributed data, Mahalanobis distance fails to indicate the samples' order in probability density.

To provide an insight into how Gaussian and non-Gaussian distributed data impact Mahalanobis distance, we present an example of samples and their probability density contour lines in Fig.3.3.

- First, let's consider that sample vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^2$ are generated from the Gaussian distributed function $\mathcal{N}$ with the following probability density P :

  $$\mathrm{P}(\mathbf{x}_1|\mathcal{N}) < \mathrm{P}(\mathbf{x}_2|\mathcal{N}) = \mathrm{P}(\mathbf{x}_3|\mathcal{N}) \tag{3.5}$$

  As shown in Fig.3.3 (a), $\mathbf{x}_1$ is far from the density center, while $\mathbf{x}_2$ and $\mathbf{x}_3$ are closer and on the same density contour line. Fig.3.3 (b) illustrates the transformed samples as the intermediate results of Mahalanobis distance, viz. $\mathbf{x}^T\boldsymbol{\Sigma}^{-1/2}$, from which one can easily know that :

  $$d_M(\mathbf{x}_1, \boldsymbol{\mu}) > d_M(\mathbf{x}_2, \boldsymbol{\mu}) = d_M(\mathbf{x}_3, \boldsymbol{\mu}) \tag{3.6}$$

  This implies their correct probability order.

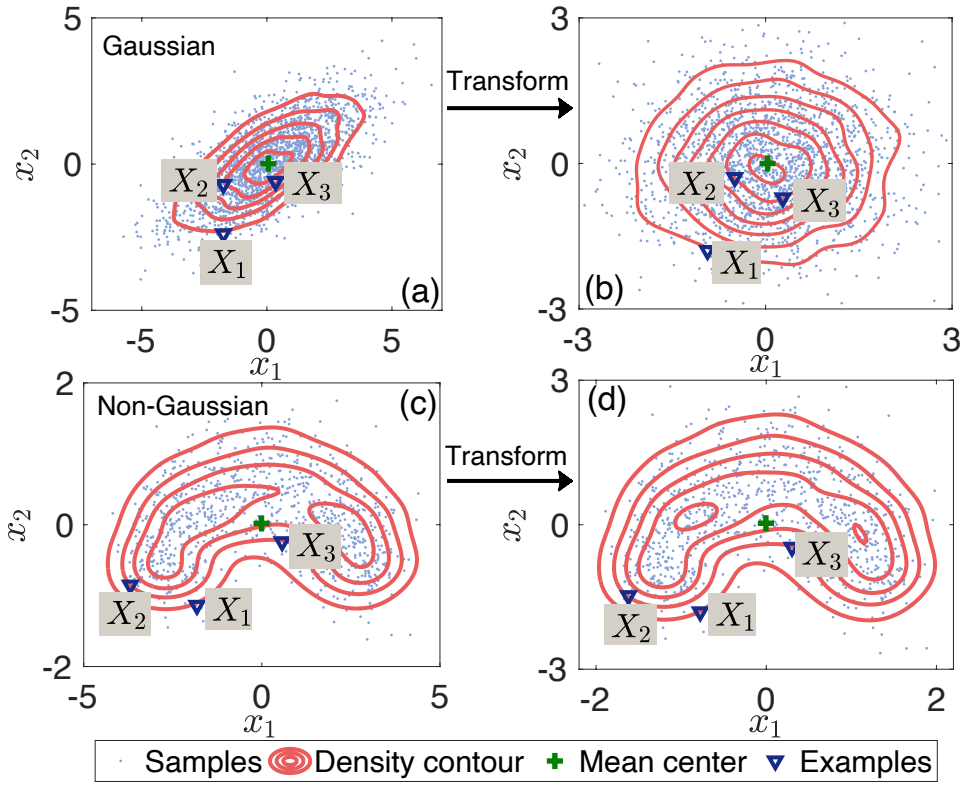

FIGURE 3.3 – a) Gaussian distributed data and density curves ; b) Data after Mahalanobis transformation ; c) Non-Gaussian distributed data and density curves ; d) Data after Mahalanobis transformation

- In the second case, when the same sample vectors are non-Gaussian distributed with the same probability conditions previously mentioned in Eq.(3.5) (see Fig.3.3 (c) and (d)), the Mahalanobis distance, is then :

$$d_M(\mathbf{x}_2, \boldsymbol{\mu}) < d_M(\mathbf{x}_1, \boldsymbol{\mu}) < d_M(\mathbf{x}_3, \boldsymbol{\mu}) \tag{3.7}$$

Indeed, this does not correspond to the correct vectors' probability order. Mahalanobis distance with the center $\boldsymbol{\mu}$ is disabled to correctly order samples from an irregular distribution. One may notice from the example that the real key to this issue is the distance between tested samples and the healthy region in the studied space. Therefore, we consider the minimum Mahalanobis distance between a sample and the healthy region as the measurement of their closeness. Then, this minimum distance is called hereafter the Local Mahalanobis Distance (LMD) for its local consideration of the healthy region and is denoted as D. Formally, LMD is defined as

$$D_L(\boldsymbol{x}; \Omega_H) = \min_k \{d_M(\boldsymbol{x}, \boldsymbol{C}_k) \,|\boldsymbol{C}_k \in S\} \tag{3.8}$$

where S is the anchor set. The selection of anchors, called the healthy region approximation, aims to generate the anchor set from the training samples as the approximation of Ω_H. Let κ be the element number of S. According to Eq.(3.8), three properties of D_L are summarized as follows :

1. $D_L \in \boldsymbol{R}^+$.
2. $\lim_{k \to +\infty} D_L(\boldsymbol{x}; \Omega_H) = 0$ *for* $\boldsymbol{x} \in \Omega_H$

It means that the ideal LMD value of the healthy samples is 0. Since $\boldsymbol{x} \in \Omega_H$ and $\Omega_H = \lim_{k \to +\infty} S$ we have

$$D_L = \min_k \{d_M(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x}, \boldsymbol{y} \in \Omega_H\}$$

Then, there is always a point $\boldsymbol{y}$ equal to $\boldsymbol{x}$, and the minimum value is always 0.

3. According to the theorem of extreme value statistics [9, 95], the distribution of D_L can be modeled by the following general form named Generalized Extreme Value distribution (GEV) for minima

$$\Phi_G(D_L; \rho, \beta, \tau_a) = 1 - \exp\left\{ - \left[1 + \tau_a \left(\frac{\rho - D_L}{\beta} \right) \right]^{-1/\tau_a} \right\} \qquad (3.9)$$
$$-\beta - \tau_a(\rho - D_L) \leq 0, \qquad \beta > 0$$

where ρ, β, τ_a are the location, scale and shape parameters, respectively.

The three properties of LMD will be used to obtain optimal anchors and determine the region margin.

3.3.2 Anchors generation

As mentioned previously, the healthy domain approximation for our proposal has to be done using a set of anchors. This part aims at generating these anchors based on fault-free samples. According to the definition of LMD, its computational cost is mainly related to the number of elements of S. Although we will have more accurate results with a larger number of anchors, we notice that reducing the number of elements to some extent will not cause severe performance degradation. As an example, we randomly select anchors from fault-free samples and present in Figure 3.4 the evolution of the LMD according to the anchors' number K_{an}. It is clear that slightly decreasing the number of anchors can keep reasonable accuracy but reduce computational cost. Therefore, the first goal of the anchors-generation procedure is to remove redundancy and extract critical spatial information from the original sample set. More precisely, geometrically close samples are identified and merged as an anchor. Although clustering algorithms, like k-means [70], can achieve a similar goal, they are either easily affected by outliers or cannot be optimized in the proposed framework. The second goal of the anchors-generation procedure is to eliminate the impact of outliers and increase the generalization ability of the anchor set. To that end, we use the local density information to identify outliers by restricting the minimum number of samples in a local region. If the samples number in a local region is too small, these samples are recognized as outliers and will be excluded from generating anchors. Finally, the robust anchors generating algorithm is given as follows.

In this algorithm, two parameters are required, namely the region radius γ_{an} and the limitation number η_{an}. The region radius specifies the size of local regions in which original samples will be merged since they are close spatially. In other words, the region radius controls the approximation accuracy and the number of anchors. Furthermore, the proposed algorithm counts the samples number in each local region and compares them with the limitation number η_{an}. When the number is smaller than η_{an}, indicating a low local density of samples, the corresponding region will be ignored since samples in this

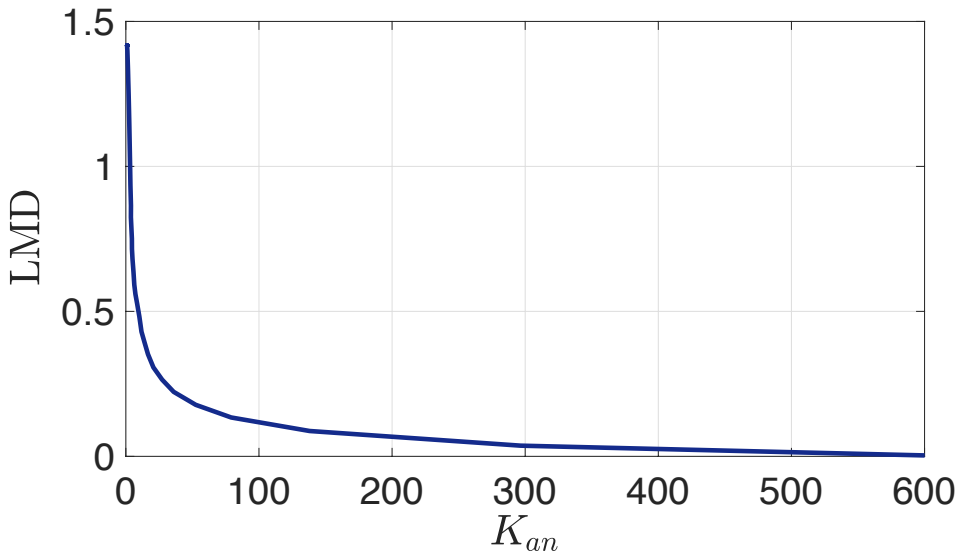

FIGURE 3.4 – LMD evolution with the number of anchors

Algorithm 1 Anchors-generating algorithm.

Input: Fault-free samples $\boldsymbol{X}^* = \{\boldsymbol{x}_i\}_{i=1}^N$, local radius γ_{an}, minimum number of local samples η_{an}
Output: Anchor set S

$\boldsymbol{\mu} \leftarrow \frac{\sum_{i=1}^N \boldsymbol{x}_i}{N}$
$\boldsymbol{C}_1 \leftarrow \boldsymbol{\mu}$
$\boldsymbol{\Sigma} \leftarrow \text{cov}(\boldsymbol{X}^*, \boldsymbol{X}^*)$
$d_M(\boldsymbol{x}_i) \leftarrow \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})^T}$
Rearrange fault-free samples as $\boldsymbol{X}'^*$ in ascending order according to $d_M(\boldsymbol{x}_i)$
$k \leftarrow 1$
repeat
    $k \leftarrow k + 1$
    Take one sample $\boldsymbol{x}_i$ from $\boldsymbol{X}'^*$
    Find out $Z_k = \{\boldsymbol{x}_l | d_M(\boldsymbol{x}_i, \boldsymbol{x}_l) < \gamma_{an}, l > i, \boldsymbol{x}_l \in \boldsymbol{X}'^*\}$
    $n_k \leftarrow$ samples number in Z_k
    if $n_k < \eta_{an}$ **then** Continue
    end if
    $\boldsymbol{C}_k \leftarrow \frac{1}{n_k} \sum_{q=1}^{n_k} \boldsymbol{x}_q (\boldsymbol{x}_q \in Z_k)$
    Remove Z_k from $\boldsymbol{X}'^*$
until $\boldsymbol{X}'^*$ is empty
$S \leftarrow \{\boldsymbol{C}_1, \cdots, \boldsymbol{C}_k\}$
return S;

region are potential outliers. Therefore, η_{an} should be firstly selected according to the quality of training data. Generally, the more outliers contained in the training data, the larger η_{an} should be set. Then the local region radius γ_{an} can be selected to minimize the approximated error, which is derived as :

$$\text{Err} = \sum_{i=1}^{N} D_L(\boldsymbol{x}_i) \tag{3.10}$$

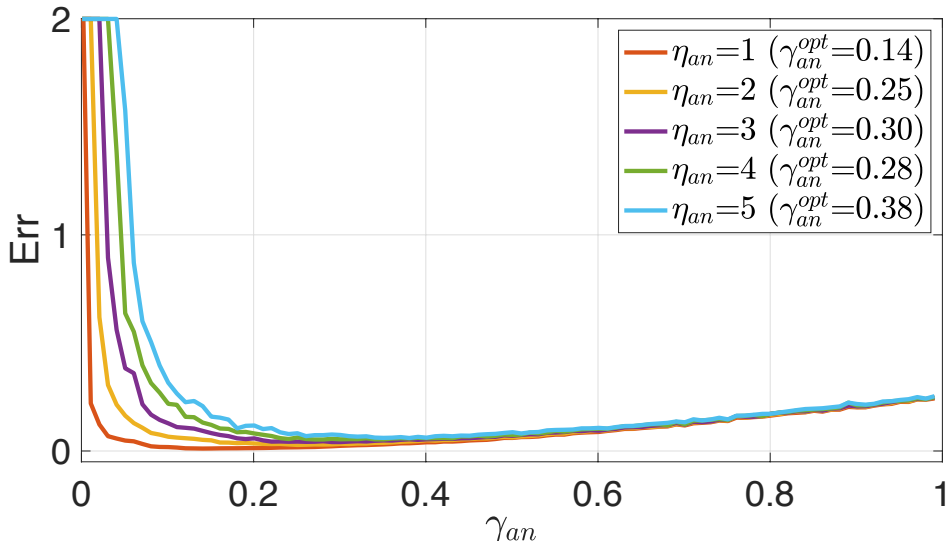

FIGURE 3.5 – Evolution of the approximation error along with γ_{an}

Fig.3.5 shows the evolution of the approximation error along with γ_{an} for different settings of η_{an}. For each η_{an} setting, the evolution curve of the error is convex, guaranteeing that a minimum value can be found in the searching procedure of the optimal value γ_{an}^{opt}.

3.3.3 Margin selection based on PDF estimation

The previous subsection specifies the centers of a healthy region but remains the selection of a margin to determine the region ultimately. Margin is as important as centers to obtain a robust and accurate healthy region. It is also crucial for the fault detection procedure to trade off the false alarm and fault detection rates. Instead of determining different radii for each local region, we simplify the margin selection problem by assigning the same radius for all local regions to avoid high computational cost and over-fitting issue.

Like the anchors-generation procedure, outliers dramatically affect the margin selection. For example, the k-centers approach determines a set of regions to cover over all the training samples and meanwhile minimizes the maximum radius of those regions [201]. This approach usually leads to a larger healthy region for the ineffective margin selection when outliers are present. This subsection presents a robust approach to obtain a decision boundary for healthy region approximation by mitigating the influence of outliers. To that end, we should first identify the potential outliers and then exclude them from determining a margin. According to the properties of LMD, the major characteristic of outliers is their distinct large LMD values. If we consider the probability distribution of D_L, one can easily identify outliers since their LMD values are located in the largest part of the CDF curve. Therefore, we propose to select the corresponding region margin based on the probability model with a given significance level α.

To estimate the probability distribution of LMD for training samples, we introduce the Generalized Extreme Value distribution as follows. For more details on extreme value statistics, readers can refer to [9, 95]. According to the theorem of extreme value statistics, the probability distribution of LMD is established based on GEV model for minima, i.e.,

Eq.(3.9). To estimate the model's parameters, we minimize the mean square error of the estimated general extreme value distribution model and the empirical cumulative density. Then the model's parameters are determined when the error reaches the minimum. This procedure is described by the following convex optimization problem.

$$\begin{aligned} \underset{\rho,\beta,\tau_a}{\text{minimize}} \ & \frac{1}{N}\sum_{i=1}^{N}\left[\Phi_G\left(D_L(\boldsymbol{x}_i);\rho,\beta,\tau_a\right)-\mathbb{F}_e(D_L(\boldsymbol{x}_i))\right]^2 \\ \text{subject to} \ & -\beta-\tau_a(\rho-D_L(\boldsymbol{x}_i))\leq 0, \qquad \beta>0 \end{aligned} \tag{3.11}$$

where $\Phi_G(D_L(\boldsymbol{x}_i);\rho,\beta,\tau_a)$ is the GEV distribution function and $\mathbb{F}_e(\cdot)$ is the empirical cumulative density. To solve the convex optimization problem with multiple constraints, we employ the Nelder–Mead method to search for the solution iteratively until the objective function (Eq.3.11) converges [132]. Nelder–Mead method is a numerical technique allowing to search minimum or maximum in a multidimensional space. It often serves as an essential tool for convex optimization problems when derivatives of the objective function are unknown or difficult to calculate [99, 132, 163]. After determining the model's parameters, the healthy region margin r_L is determined with the given significance level α.

$$r_L=\Phi_G^{-1}(\alpha;\rho,\beta,\tau_a)=\rho+\frac{\beta}{\tau_a}-\frac{\beta\left[-\ln\left(1-\alpha\right)\right]^{-\tau_a}}{\tau_a} \tag{3.12}$$

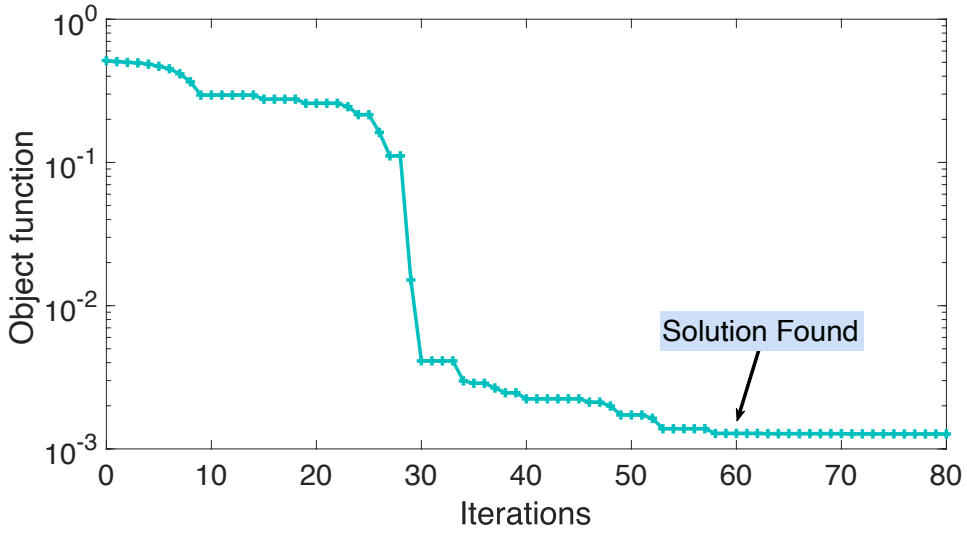

FIGURE 3.6 – Objective function in the solution searching procedure

We give an example to show the evolution of the objective value in the solving procedure. As shown in Fig.3.6, the object value decreases quickly with iterations. After 60 iterations, the solution is found for a low enough objective value (close to 10^{-3} in this example). After the parameters of the GEV model is determined, Fig.3.7 displays the estimated and the empirical cumulative density functions of the training samples' LMD values. The comparison of the two curves indicates that the obtained model has a good approximation for empirical data.

Finally, an effective healthy region can be obtained based on the generated anchors and the determined region margin. We apply the proposed approach to the toy example given in section 3.2. Fig.3.8 illustrates the developed healthy regions for the Gaussian

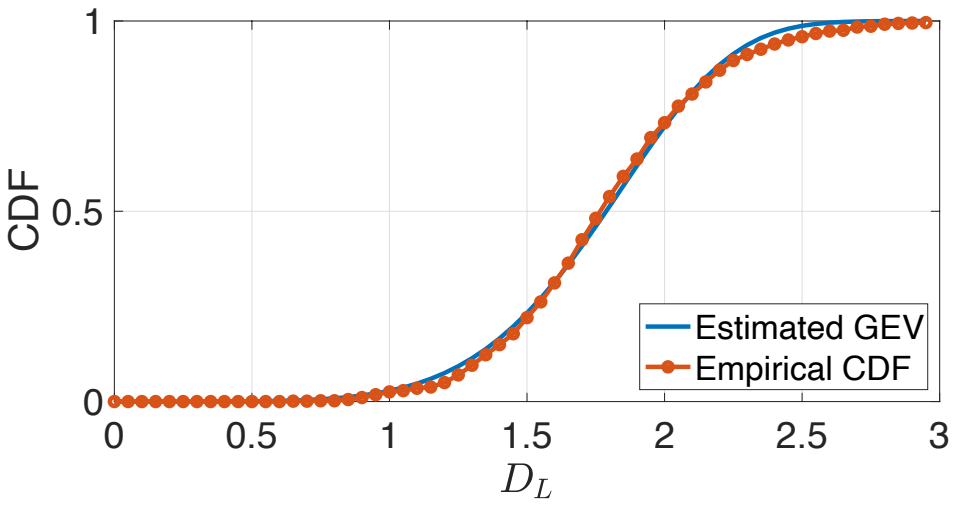

FIGURE 3.7 – Estimated and empirical CDF curve

and non-Gaussian data. Intuitively, the LMD-based healthy regions are notably more effective than T^2, SPE, and mixture indexes since they fit the data distribution precisely. The numerical criteria given in Table 3.2 also indicate the significant improvement of the proposed approach compared to T^2, SPE, and mixture indexes. Although the accuracy of the proposed method is 1% less than the three traditional indexes, their IoU values are 27.96% larger for Gaussian data and 43.68% larger for non-Gaussian data. The benefit of using the LMD proposal for healthy region approximation is then highlighted.

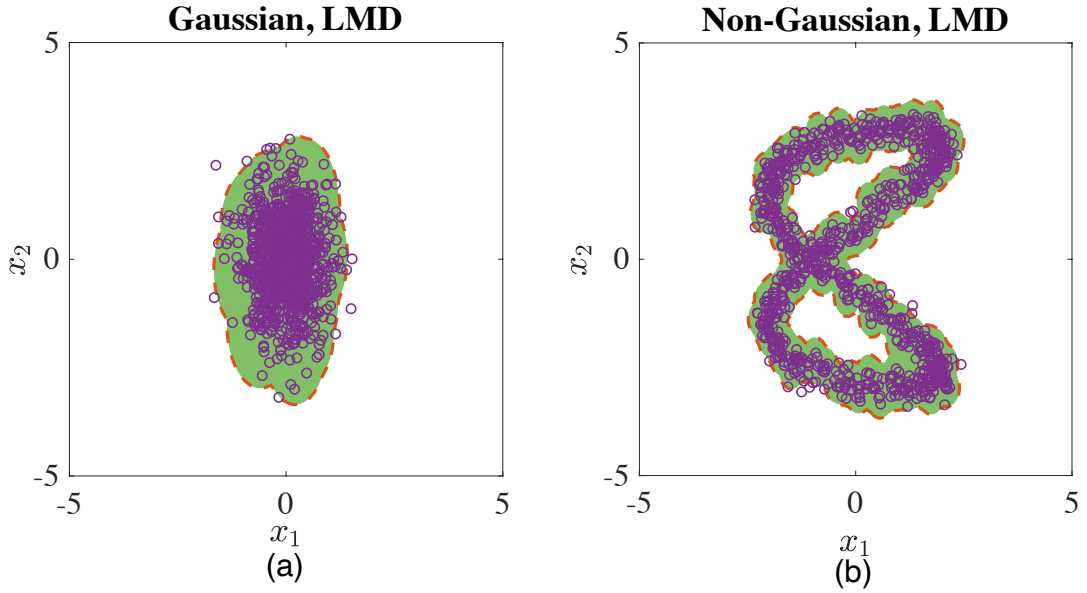

FIGURE 3.8 – The developed healthy region by using the proposed healthy region approximation approach : (a) Gaussian distributed data ; (b) Non-Gaussian distributed data.

TABLE 3.2 – Healthy region approximation performance of the proposed method

Methods	Gaussian		Non-Gaussian	
	Acc_H (%)	IoU (%)	Acc_H (%)	IoU (%)
LMD	99.0	65.02	99.0	79.29

3.4 Performance analysis

This section shows the performance of the proposed healthy region approximation approach for different kinds of data and highlights its superiority by comparing it with other OCC approaches. Let us consider four cases of data as follows :

- Case 1 : Two independent variables following standard Gaussian (normal) distribution, i.e., $\boldsymbol{x}_1, \boldsymbol{x}_2 \sim \mathcal{N}(0, 1)$
- Case 2 : 2-dimensional Gaussian distribution.

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2] \sim \mathcal{N}(\mathbf{0}, \begin{bmatrix} 0.8 & 0.3 \\ 0.3 & 1 \end{bmatrix})$$

- Case 3 : Two periodic signals $\boldsymbol{x}_1 = 2\cos(t)$, $\boldsymbol{x}_2 = 2\sin(2t - \frac{\pi}{2})$
- Case 4 : Mixed Gaussian distribution $\boldsymbol{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]^T$, where

$$\mathbf{x}_1 \sim \mathcal{N}([0, -2], \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix})$$
$$\mathbf{x}_2 \sim \mathcal{N}([2, 2], \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix})$$
$$\mathbf{x}_3 \sim \mathcal{N}([-2, 2], \begin{bmatrix} 0.2 & 0 \\ 0 & 0.6 \end{bmatrix})$$

We also introduce outliers samples following the uniform distribution, i.e., $\boldsymbol{X}_o \sim \text{U}(-5, 5)$.

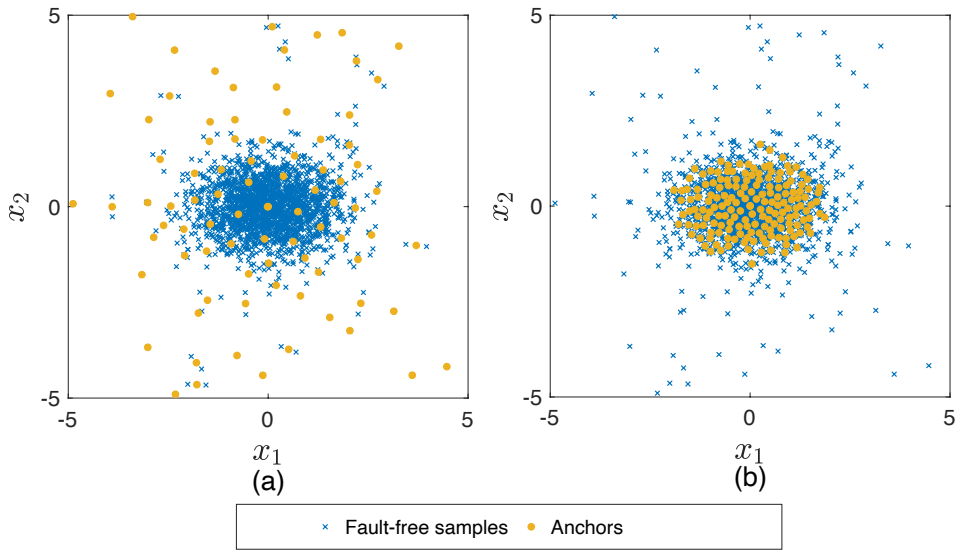

FIGURE 3.9 – Output of anchor-generating algorithms : (a) without considering outliers ($\eta_{an} = 0$) ; (b) considering outliers ($\eta_{an} = 1$).

Outlier is a major external factor affecting the performance of healthy domain approximation. The proposed anchors-generating algorithm uses local density information of samples to recognize outliers so as to obtain a more effective anchor set. Accordingly, we first investigate the benefit of using local density information in generating robust anchors. We produce the data of case 1 mixed with 50 outliers. Fig.3.9 (a) shows the generated anchors using the proposed anchors-generating algorithm with $\eta_{an} = 0$, which means the algorithm does not consider the local density information in the training process. In this case, anchors spread over the entire observed area instead of concentrating on the center of the training samples. The corresponding healthy region, in this case, is therefore inaccurate. While, when we set $\eta_{an} = 1$, the algorithm excludes outliers in the training fault-free samples, gathering anchors in the density center of training samples (see Fig.3.9 (b)). The comparative result indicates that using local density information in anchors generation significantly improves the accuracy and robustness of the healthy region approximation.

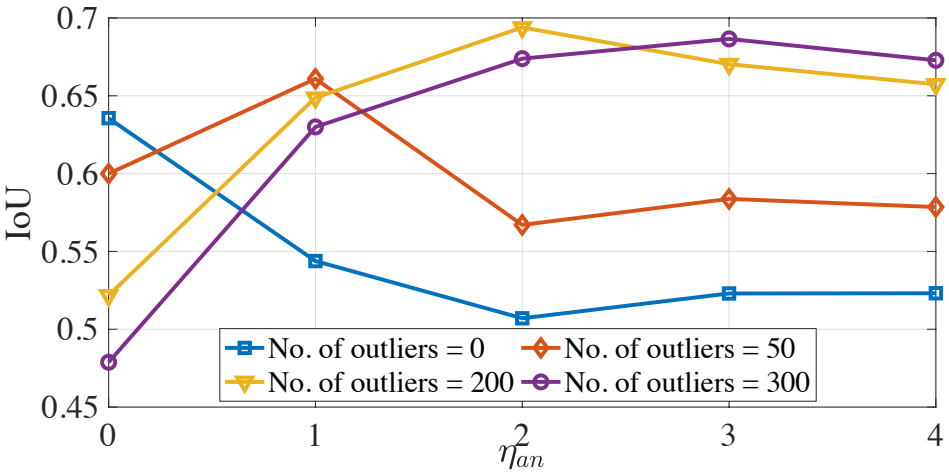

FIGURE 3.10 – Performance evolution of the proposed method with different parameters for varying number of outliers

Secondly, we change the number of outliers and investigate the IoU performance of the proposed method by setting different values for parameter η_{an}. The result is shown in Fig.3.10. When the training data does not contain outliers (the blue curve marked by rectangles), the setting of $\eta_{an} = 0$ results in the best IoU performance. If the outliers number is 50 (the red curve marked by diamonds), the parameter η_{an} should be set as 1 to reach the best IoU performance. The optimal parameter value of η_{an} is 2 and 3 for 200 and 300 outliers, respectively. In summary, η_{an} should be tuned according to the number of outliers : the more outliers, the larger η_{an} value. Even though the exact outliers number is unknown in practice, one can still properly set the parameter by roughly estimating the outliers number.

We compare the healthy regions determined by different OCC approaches, namely OC-SVM [93, 138, 157], IF [65, 114, 115], k-centers [201], AE [126, 180], the PCA technique with the mixture index of T^2 and SPE statistics (PCA-MIX) [56, 89], and our proposal. This experiment does not consider outliers but focuses on the methods' ability to handle non-linear data. Accordingly, we produce the above four cases of data and apply these methods to the data. As illustrated in Fig.3.11, the healthy regions of our proposal and the OC-SVM can fit well with all the data, indicating that they are effective for linear and non-linear cases. The IF and K-centers approaches are also flexible in describing a healthy

region for non-linear data. However, the former is too strict to produce a boundary, while the latter tends to a larger healthy region. In the subsequent fault detection procedure, a smaller healthy region usually leads to a high false alarm rate, while a larger one leads to a high miss detection rate. The AE and PCA approach with mixture index have similar performance for Gaussian cases (case 1 and case 2), which are effective but still not accurate enough compared to our proposal and OC-SVM. For non-linear cases (case 3 and case 4), these two approaches are not flexible to fix data.

TABLE 3.3 – Healthy region approximation performance without outliers

Methods	Case 1		Case 2		Case 3		Case 4	
	Acc_H	IoU	Acc_H	IoU	Acc_H	IoU	Acc_H	IoU
LMD	99.5	**64.70**	99.5	**68.82**	99.5	**70.87**	99.5	**65.23**
OC-SVM	99.4	57.01	99.4	57.78	99.6	64.12	99.6	52.54
IF	78.7	52.98	78.9	48.45	59.34	38.73	69.6	27.46
k-centers	100.0	28.59	100.0	38.18	100.0	50.9	100.0	31.94
AE	100.0	27.89	100.0	38.74	100.0	38.19	100.0	28.02
PCA-MIX	100.0	27.71	100.0	33.56	100.0	23.92	100.0	16.19

We also evaluate the IoU and accuracy performance of these approaches for each case of data, as given in Table 3.3. k-centers, AE, and PCA-MIX achieve the perfect accuracy performance (100%). The healthy regions they developed can cover all training samples. However, their IoU results are low (less than 50%), meaning that healthy regions are too large. Concerning the IoU results, the LMD approach always has the best performance. Its IoU results are significantly larger than the OC-SVM's, even though their healthy region illustrated in Fig.3.11 seems similar. The IF approach is unreliable since its accuracy performance is low. Therefore, the LMD approach has better trade-off performance between accuracy and IoU criteria than all the other approaches.

TABLE 3.4 – Healthy region approximation performance with outliers

Methods	Case 1		Case 2		Case 3		Case 4	
	Acc_H	IoU	Acc_H	IoU	Acc_H	IoU	Acc_H	IoU
LMD	99.5	**59.10**	99.5	**67.19**	99.0	**74.23**	99.0	**70.36**
OC-SVM	99.4	46.26	99.5	48.85	99.5	57.62	99.6	40.78
IF	80.0	53.82	78.8	49.89	71.29	48.04	69.6	27.67
k-centers	100.0	23.72	100.0	17.4	100.0	13.45	100.0	11.38
AE	100.0	22.37	100.0	24.43	100.0	32.04	100.0	23.22
PCA-MIX	100.0	20.53	100.0	32.35	100.0	23.91	100.0	16.30

Next, we investigate the robustness of each approach against outliers by introducing 10 outliers to the training data. The healthy regions of each approach are demonstrated in Fig.3.12. Benefiting from the robustness algorithm, our proposal is not affected by outliers and can determine an accurate healthy region (see the first line of figures in Fig.3.12). While outliers have a slight impact on the OC-SVM approach since it mistakenly includes outliers into the healthy region (see second line of figures in case 1 and case 2). The IF method also shows strong robustness against outliers but establishes a too small healthy region in this experiment. IoU and accuracy performance are summarized in Table3.4. The LMD approach outperforms all the other approaches in terms of IoU and it has

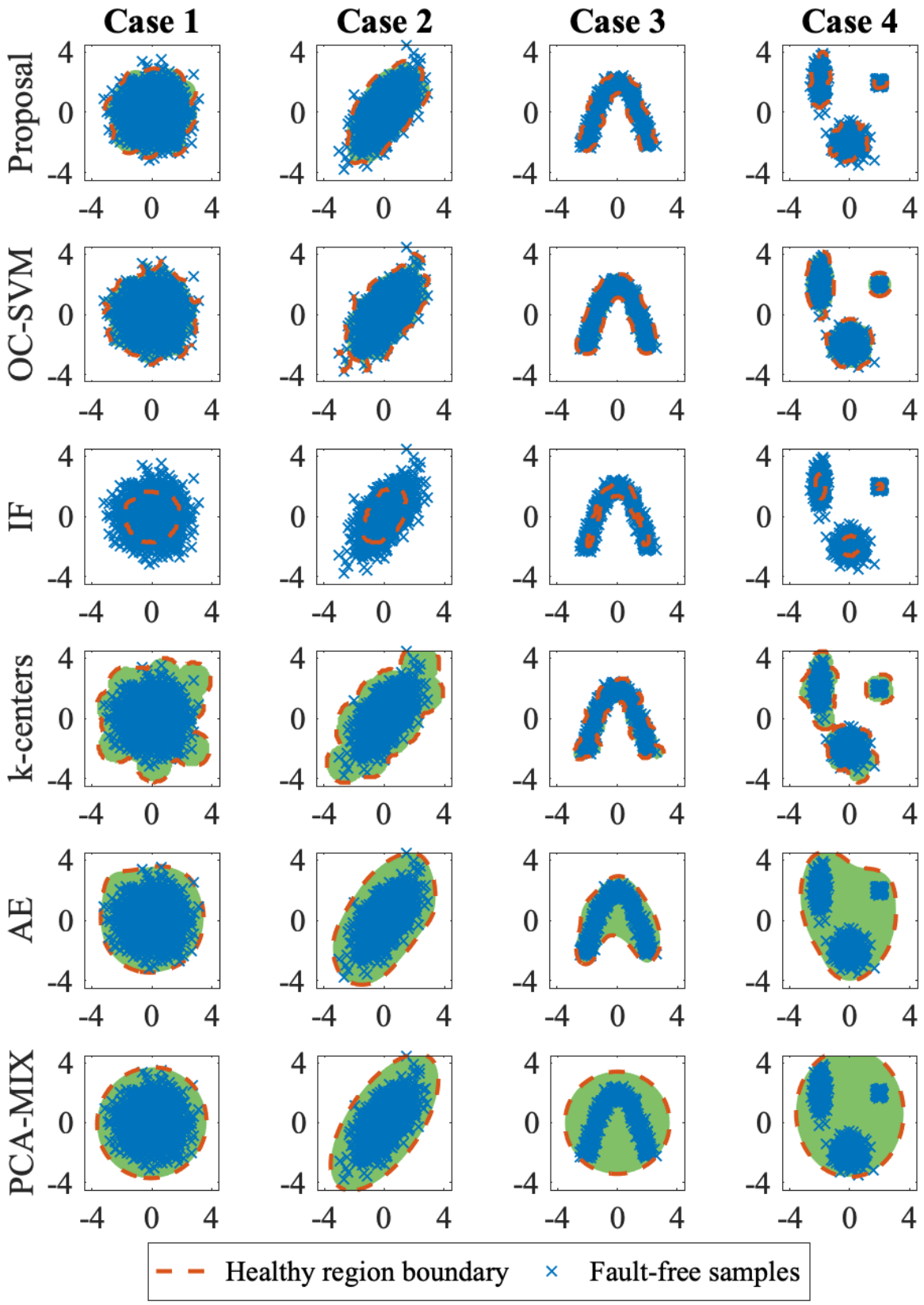

FIGURE 3.11 – The healthy regions of different OCC approaches for the data without containing outlier.

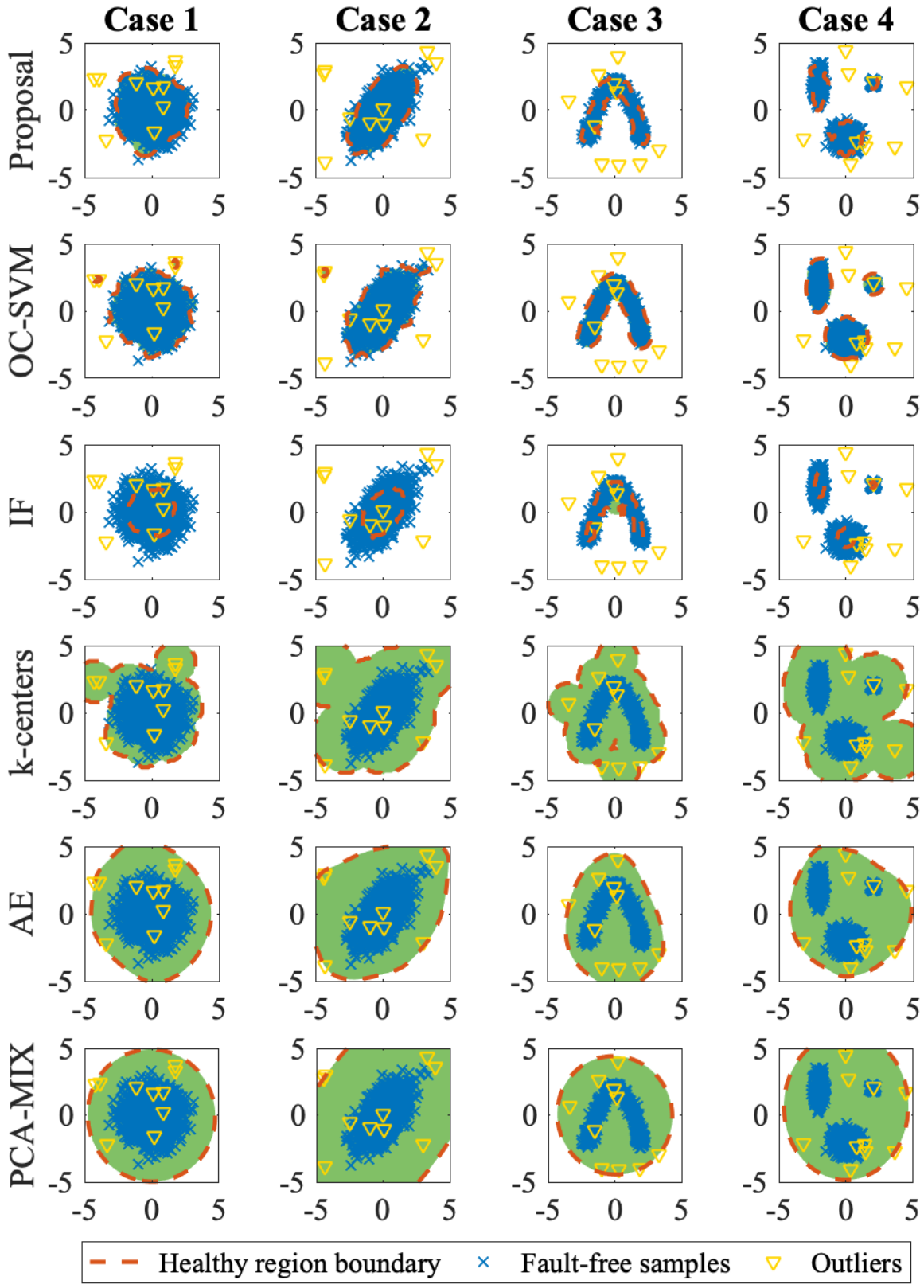

FIGURE 3.12 – The healthy regions of different OCC approaches for the data containing outlier.

high accuracy performance (at least 99%). Hence, this approach has the best trade-off performance in this case with outliers. The obtained performance shows that the LMD proposal is then robust to outliers. k-centers's performance dramatically degenerates for the introduction of outliers, indicating its poor robustness performance against outliers. The outcomes of this experiment clearly demonstrate that the k-centers, AE, and PCA approaches are dramatically affected by outliers, producing too larger regions. Therefore, these three approaches are unreliable when the training data contain outliers.

3.5 Conclusion

The Fault diagnosis topic is facing several challenges like incipient fault diagnosis, non-linear and non-Gaussian data, the lack of fault samples, etc. One-class classification technique is one promising branch to solve the above challenges. According to the OCC idea, fault detection turns to determine a minimum region based on the given healthy samples to separate faulty and healthy samples. Two traditional statistics for fault diagnosis, Hotelling T^2 and SPE, are two examples of the OCC idea, but they are ineffective for non-Gaussian data. To develop an effective healthy region from the training samples, we propose a more flexible way focusing on the local information of spatial distribution. The proposed method describes a healthy region as multiple hyperspheres with different centers and the same radius. Then the healthy region approximation problem is simplified as determining those centers and a radius that minimizes the healthy region. To find the proper size of local regions, a particular distance measurement, called local Mahalanobis distance, is then defined as the minimum Mahalanobis distance between a sample and the healthy region. Based on the LMD, an efficient and robust anchor-generation algorithm is designed to determine the centers of the healthy region. This algorithm removes the redundancy of training samples and extracts critical spatial information. Furthermore, the local density information is used to identify outliers in the training samples so as to improve its accuracy and robustness. This chapter also discusses the properties of LMD, where we show that the probability distribution of LMD in the null case can be modeled as the generalized extreme value distribution. Then the parameters of the GEV model are estimated using the Nelder-Mead method. Based on the estimated probability distribution, we determine the radius of the healthy region.

To highlight the advantages of the proposed healthy region approximation idea, we show its performance by applying it to four cases of simulated data. Comparing two cases of anchors generated with and without the use of local density information, we show that this information can avoid the distractions of outliers in the anchor-generation process and improve the accuracy and robustness of the algorithm. This chapter also compares the healthy region approximation performance of different OCC approaches. The result indicates that the proposed method is effective for both Gaussian and non-Gaussian data. The established healthy region is accurate and flexible to fit different kinds of data. In the experiment without outliers, the proposed method outperforms the IF, k-centers, AE, and the PCA technique with mixture index. While, in the experiment with outliers, the proposed method is the only one not impacted by outliers. In summary, the healthy region developed by the proposed approach is accurate, effective for data without any distribution-type assumption, and robust against outliers, which provides an important foundation for subsequent fault detection, isolation, and severity estimation procedures.

4

Fault diagnosis using local Mahalanobis distance

4.1 Introduction

The chapter 3 developed an advanced healthy region approximation approach with the diagnosis index named the local Mahalanobis distance. It is sensitive to tiny deviations, effective to complex distributed data, and robust against outliers. Based on this technique, this chapter proposes different solutions for fault detection, faulty variable isolation, and fault severity estimation, respectively.

We primarily formulate the fault model and make necessary assumptions for the theoretical discussion on the fault diagnosis problem. Concerning the fault detection problem, two LMD-based solutions are proposed. For time-series signals, we propose the detection scheme combining the LMD with the empirical probability density cumulative sum technique to accurately discover signals' tiny deviations. For time-independent samples, LMD is considered as a feature and the probability-based distance is used for analysis. After fault detection, we develop the LMD-based contribution plot method for faulty variable isolation. As this approach is dedicated to single fault cases, we further propose the improved approach using the reconstruction-based contribution framework with the LMD index (LMD-RBC), which can accurately isolate multiple faulty variables. As for the fault severity estimation task, we first estimate the fault's increasing rate based on the LMD index. Despite the high accuracy, this approach has some limitations, such as the strict assumption of fault evolution as a first-order model, lack of information on fault deviation, and underestimated results. Alternatively, we propose to estimate fault amplitude by using the LMD-RBC method.

For each proposal, we provide a step-by-step guide by taking simulation data as an example. The outcome of each step is illustrated and discussed, and then the effectiveness of each solution is verified.

4.2 Fault modeling

Before discussing on fault diagnosis approaches, it is essential to formulate the considered faults and make necessary assumptions. Let us consider a data matrix with m variables and N observations. A variable (signal) is denoted as $\boldsymbol{x}_j$ $(j = 1 \cdots m)$, and the ith observed sample is written as $\mathrm{x}_i = [x_{i1}, \cdots, x_{ij}, \cdots, x_{im}]$. Without loss of generality, faulty samples can be decomposed into two independent parts : healthy and faulty components.

$$\mathrm{x}_i = \mathrm{x}_i^* + \mathrm{x}_i^f, i = 1, \cdots, N \tag{4.1}$$

In this equation, x_i^* represents the healthy part of the signal (fault-free data), and x_i^f denotes the faulty component. Instead of a strict assumption limiting the application scenarios of the developed fault diagnosis approaches, Eq.(4.1) is a general model offering a convenient way to analyze a faulty behavior. A fault with abnormal gain (multiplicative form) can be modeled as an additive form using Eq.(4.1) [67]. This model establishes a reasonable bridge between the theoretical analysis and engineering problems. Thus it is widely used in industrial applications, such as diagnosing sensors fault in control systems [85, 142], detecting the non-destructive crack of material [69, 217], and assessing fault severity [67, 68].

We consider that a fault occurs from b_oth to Nth observations, meaning that at least one variable in the last $N - b_o + 1$ observations is faulty. Particularly, for time series signals, b_o represents the fault occurrence time. Generally, online monitoring tasks usually require a short delay between the fault detection and the fault occurrence time [142]. For convenience, we illustrated the above notations in Fig.4.1. The observed value of each variable is arranged in the horizontal direction according to their sampling order, where faulty samples started from b_o are marked in red. Then each column of data points is regarded as a sample vector x. Faults may occur in one or more variables. For the case of multiple faulty variables, we specify the index of the first faulty sample as b_o.

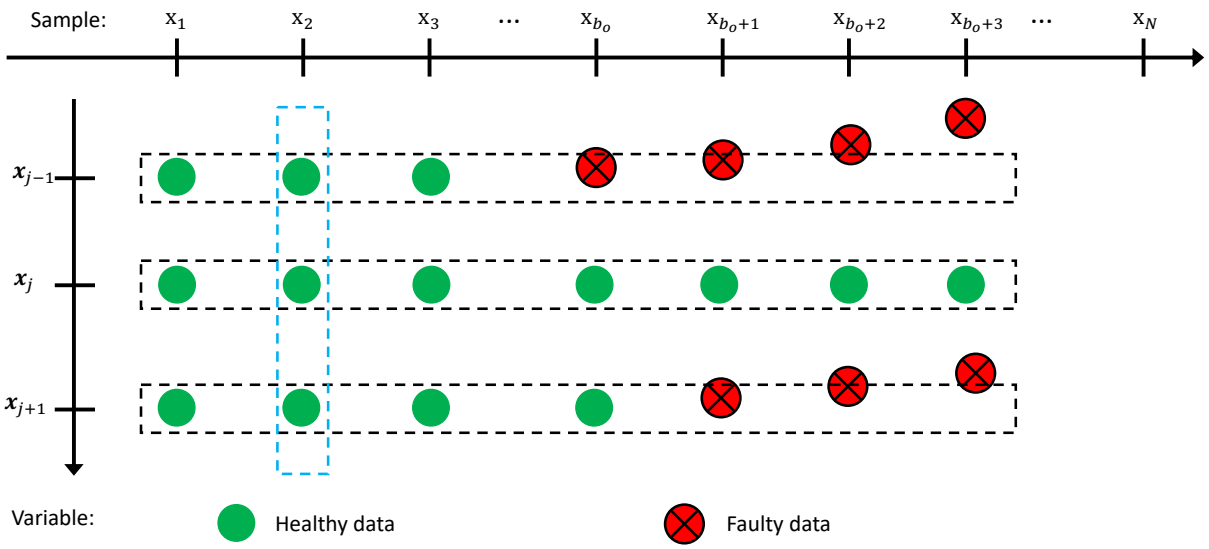

FIGURE 4.1 – Diagram of data points arrangement

After the fault detection, the major attention of the subsequent diagnosis tasks is paid to faulty samples, i.e., the last $N - b_o + 1$ observations. The faulty component is

composed of faulty variables and fault amplitudes. Therefore, we decompose the faulty component into the fault direction matrix $\mathbf{\Theta}$ and the fault amplitude vector Δ, such that $\mathrm{x}_i^f = \mathbf{\Delta}_i\mathbf{\Theta}$. The fault direction matrix $\mathbf{\Theta} \in \{0,1\}^{n_f \times m}$ consists of n_f columns indicating n_f faulty variables. In each column, the position of element 1 indicates a faulty variable. For example, when the 1st and 2nd variables are faulty, we have

$$\mathbf{\Theta} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \end{bmatrix} \tag{4.2}$$

During the fault occurrence, the fault direction is treated as a constant. The fault amplitude vector $\mathbf{\Delta} \in \mathbb{R}^{1\times n_f}$ specifies the amplitudes of n_f faulty variables. Since $\mathbf{\Delta}$ changes with samples, its estimated value can be easily affected by noise and other interference. An effective way to improve the estimation performance is to use prior knowledge of a fault or make the necessary assumption of a faulty behavior. For example, incipient faults are usually described as slowly developing changes with time. Based on this assumption, we can further model the amplitude of a single faulty variable ($n_f = 1$) as a first-order function, such as

$$\Delta_i = \delta \cdot (i - b_o + 1) \cdot T_e, \quad (i \geq b_o) \tag{4.3}$$

where δ is a constant factor in small duration, and T_e is the sampling time. The first-order function is proper to describe the behavior of incipient faults since their changes are slow. By estimating δ as the fault's increasing rate, the severity degree of the fault can be assessed [67,216]. A rapidly rising fault usually indicates a severe level of failure. However, if a fault changes quickly (not incipient faults), we can consider a second-order or exponential function to represent the fault behavior. This will be studied in our future work. Although estimating the fault increasing rate is simple and effective, it can not describe the deviation degree of a fault. Therefore, a more direct way of fault severity estimation is to estimate the fault amplitude Δ, which will be discussed in the following [83–85]

In the fault diagnosis topic, it is essential to evaluate the performance of fault diagnosis approaches at different fault severity levels as well as by considering different noise levels. These two elements simultaneously influence the performance of a method. To take into account the effect of both elements, the fault-to-noise ratio (FNR) is widely used in the evaluation procedure. The definition of FNR is given as :

$$\mathrm{FNR} = 10 \log \frac{p_f}{p_n} \tag{4.4}$$

where p_n is the noise power, and p_f is the faulty component power. Here, we further develop the fault power p_f based on the derived first-order fault model (Eq.(4.3)) as

$$\begin{aligned} p_f &= \frac{\sum_{i=b_o}^{N} \Delta_i^2}{N} = \frac{\sum_{i=1}^{N-b_o+1} \delta^2 i^2 T_e^2}{N} \\ &= \frac{\delta^2 T_e^2 (N - b_o + 1)(N - b_o + 2)(2N - 2b_o + 3)}{6N} \end{aligned} \tag{4.5}$$

Using Eq.(4.5), the FNR can be derived, and the influence of the fault severity for a given noise environment can be studied. Note that generally, the samples' number N is usually large, and $N - b_o$ is the number of faulty samples. For the isolation and estimation procedures, only these faulty samples are considered. A low number of faulty samples will reduce the fault power and increase the difficulty of diagnosis, particularly in a noisy

environment. The faulty samples number, as a factor affecting the diagnosis performance, will be further studied in the next chapter by considering different FNR values.

In order to go through the explanations of the proposal and show a step-by-step validation, we consider, in the following simulations, a multivariate data system example referred to in the paper [200]. The considered eight-variable system with N samples is given follows, and their example data are shown in Fig.4.2.

$$\begin{aligned}
x_1(t) &= 1 + \sin(0.1t) \\
x_2(t) &= 2\cos^3(0.25t) \cdot e^{-t/N} \\
x_3(t) &= \log(x_2(t)^2) \\
x_4(t) &= x_1(t) + x_2(t) \\
x_5(t) &= x_1(t) - x_2(t) \\
x_6(t) &= 2x_1(t) + x_2(t) \\
x_7(t) &= x_1(t) + x_3(t) \\
x_8(t) &\sim \mathcal{N}(0,1)
\end{aligned} \tag{4.6}$$

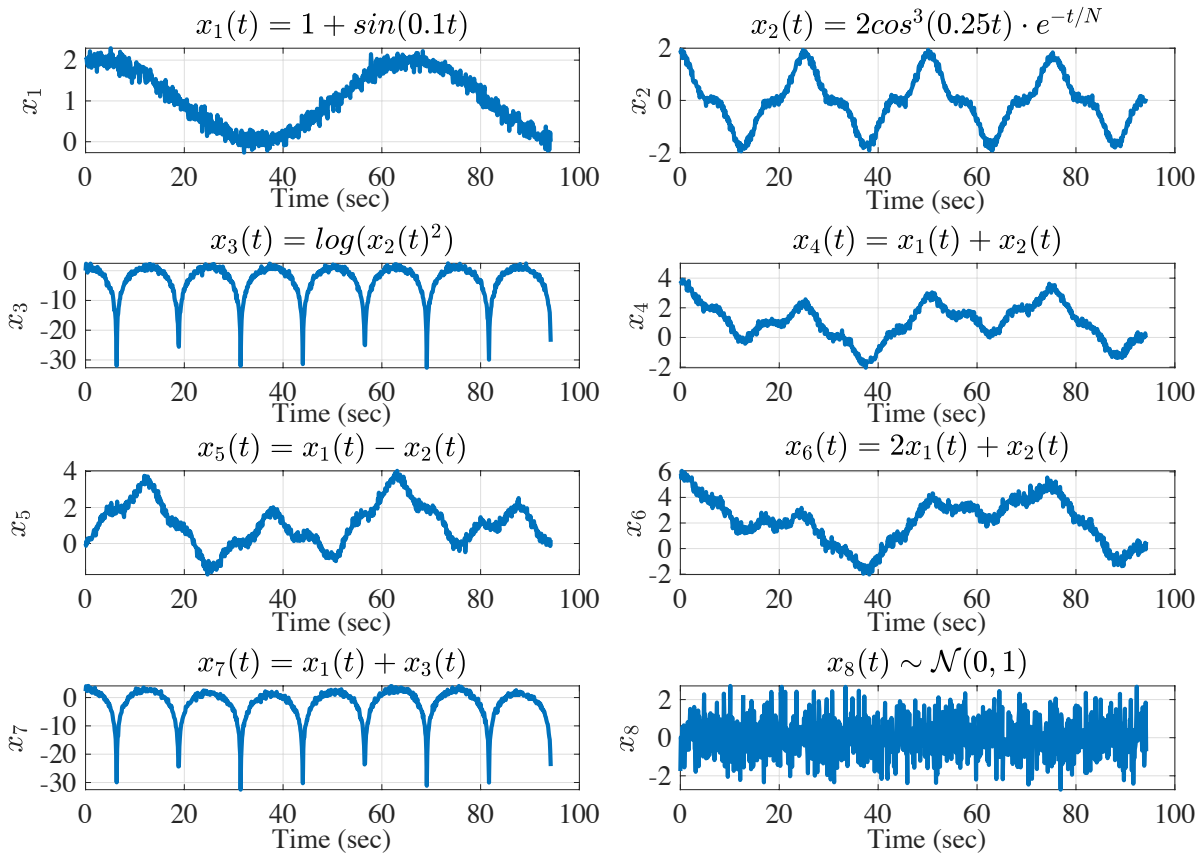

FIGURE 4.2 – Examples of simulation signals

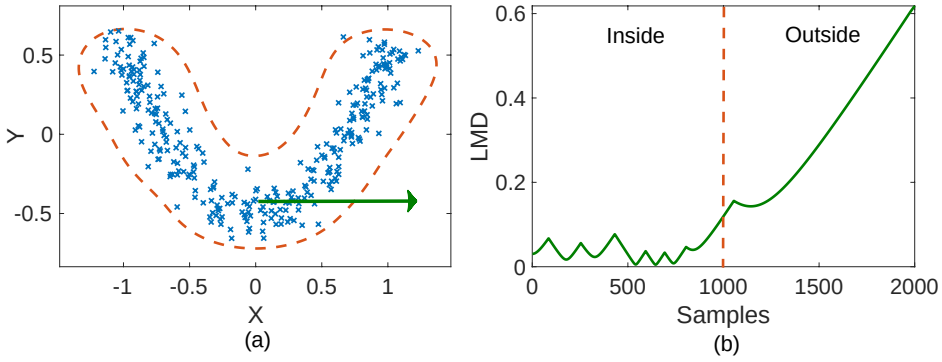

FIGURE 4.3 – (a) 2-Dimensional non-Gaussian distributed samples and the corresponding healthy region ; (b) LMD result of samples moving from the center to the outside of the healthy region.

4.3 Fault detection approaches

As reviewed in chapter 2, fault detection is the foundational task of fault diagnosis, attracting major attention in the literature. Basically, the fault detection task focuses on deciding if observed samples are faulty or not. Moreover, for real-time applications, this task's additional requirement is to detect a fault within a short time after the fault occurrence. This requirement is important, especially for incipient faults, since effective detection in the early stages of failure can dramatically reduce the accident risk. However, incipient fault detection is a real challenge, particularly in a strong noisy environment. Therefore, in this section, we propose effective incipient fault detection approaches using the local Mahalanobis distance.

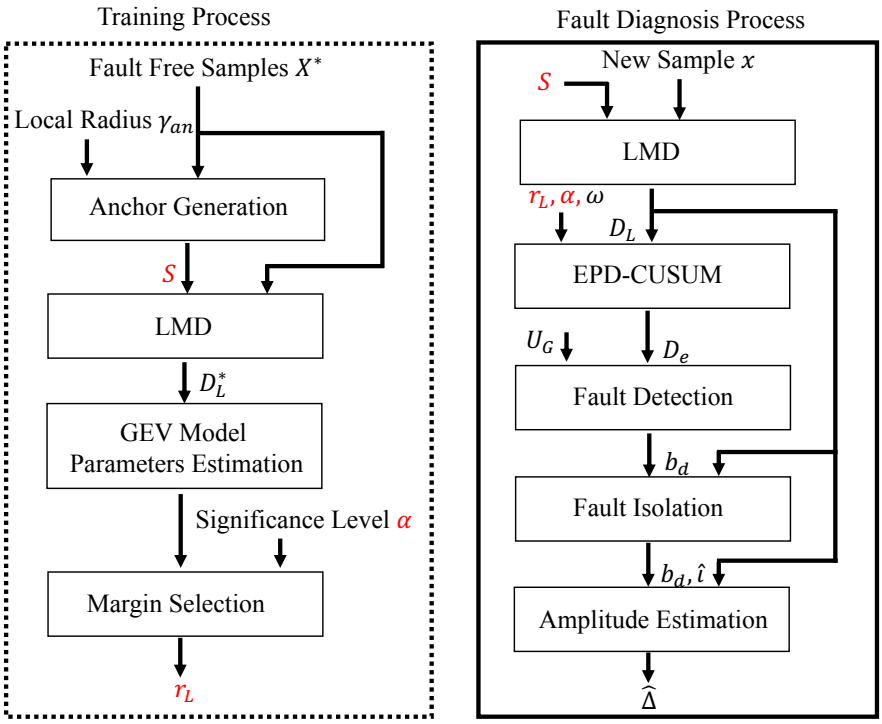

FIGURE 4.4 – Proposed LMD-based diagnosis procedure

To explain the benefit of using LMD as the major index for fault detection, we present 2-dimensional non-Gaussian data as a toy example in Fig.4.3. First, in Fig.4.3 (a), the red dotted curve represents the healthy region and the scatter of points in blue are healthy samples. To visualize the LMD behavior, we consider samples moving from the center to the outside of the healthy region. The LMD results of samples demonstrated in Fig.4.3 (b) fluctuates slightly above 0 when samples are within the region and sharply increases once they go out of the region. Based on this feature, one can effectively recognize an outlier by monitoring its LMD value. Then, our proposal for incipient fault diagnosis is presented in two parts : the training process and the fault diagnosis process. The main steps of these two parts are summarized in Fig.4.4 and described in the following subsections.

4.3.1 Training process

This training process aims to obtain the optimal anchor set S and determine the domain margin r_L to provide a setting for the detection procedure. Besides fault detection, the training result is also used for fault isolation and fault amplitude estimation. In this process, we consider a fault-free data matrix $\boldsymbol{X}^*$ with N_t observations for training. Like most learning-based algorithms, the number of training samples N_t is important for the good performance of the proposed approaches. It should be sufficiently large to estimate the healthy domain accurately. This dimension will be discussed in the next chapter.

The first step of the training process is to perform the anchor-generation algorithm on $\boldsymbol{X}^*$. Here, we set the limitation number $\eta_{an} = 1$. It is a usual value for high-quality samples set without outliers. The anchor-generation algorithm searches for the optimal local radius γ_{an} by minimizing the approximation error. We calculate the approximation error for different local radii values and show the result in Fig.4.5. When the loss reaches its minimum value, the optimization procedure stops, and the corresponding radius is obtained as the optimal one. In this example, the minimum loss and its corresponding radius are marked as a red point in Fig.4.5. Finally, the algorithm yields the optimal local radius $\gamma_{an} = 1.461$ and allows to obtain 178 anchors from 1200 initial training samples.

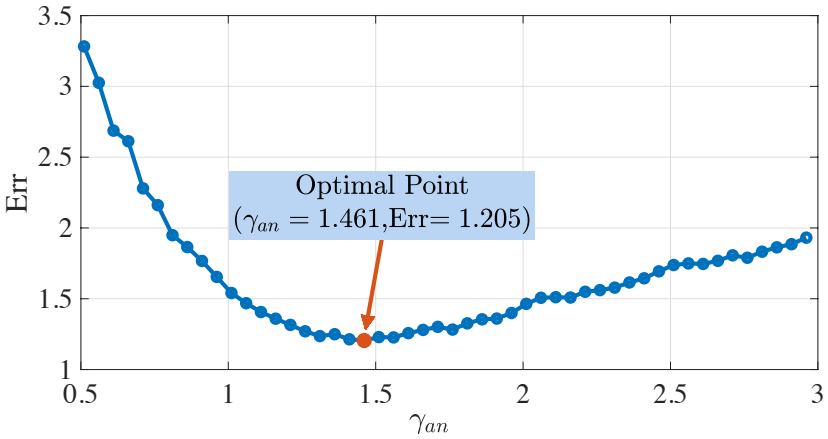

FIGURE 4.5 – Evolution of loss function versus region radius

In the second step, we compute the LMD value of the training samples based on the obtained anchors set. Fig.4.6 illustrates the obtained results, showing that their LMD

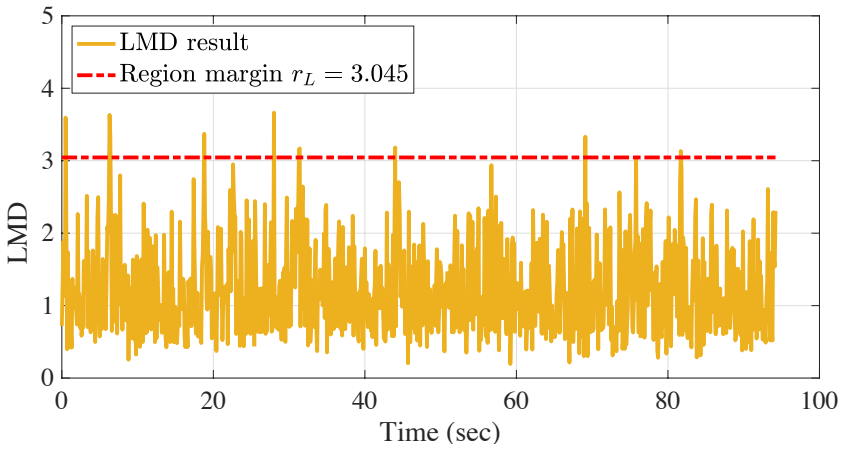

FIGURE 4.6 – LMD value of the fault-free training samples

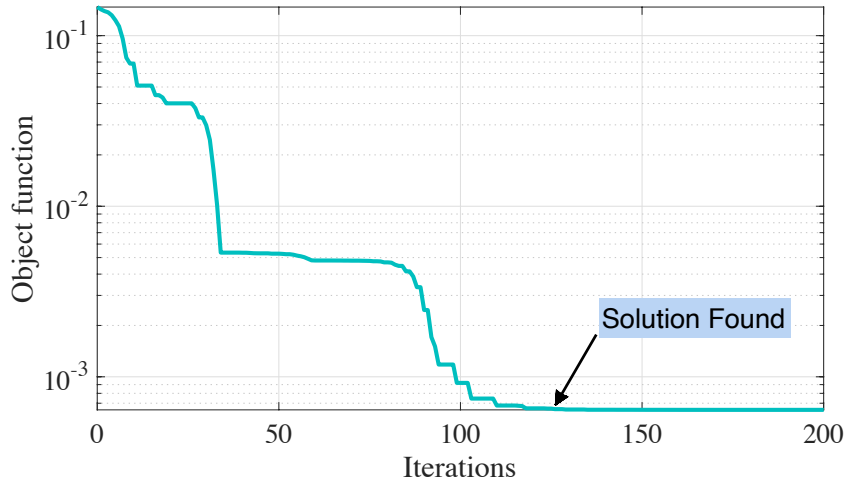

FIGURE 4.7 – Objective function in the solution searching procedure

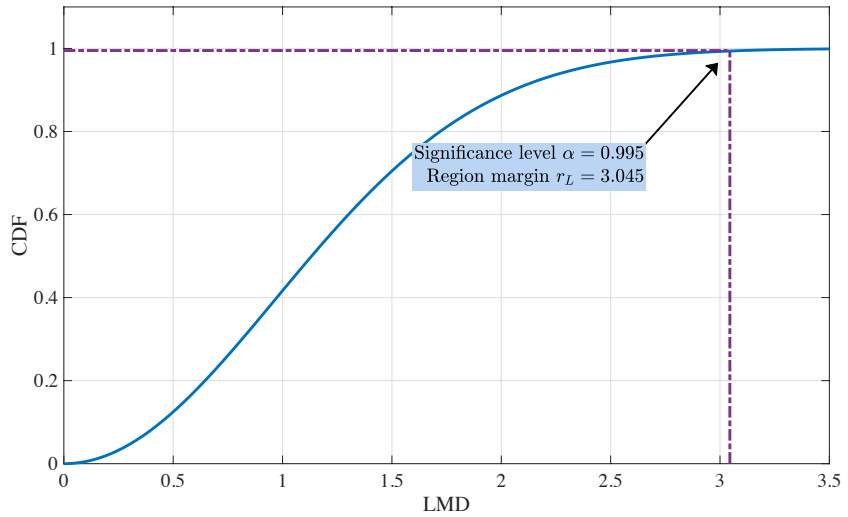

FIGURE 4.8 – Estimated CDF of the LMD value and the selected region margin value corresponding to 99.5% significance level

values are smaller than 4. Then, we need to determine a threshold (region margin r_L) to

distinguish between faulty and healthy samples. To that end, in the third step, we estimate the GEV model based on the LMD values of healthy samples. This step employs the Nelder–Mead method to solve the optimization problem of Eq.(3.11). As shown in Fig.4.7, the value of the objective function decreases quickly with iterations and finally converges to a small value. An effective solution is obtained for the simulation data after 120 iterations, and the objective value is smaller than 10^{-3}. Fig.4.8 displays the estimated CDF of the LMD values in the healthy case, based on which we select the region margin as $r_L = 3.045$ by specifying the significant level as $\alpha = 99.5\%$. It means that if a sample's LMD value is larger than the region margin, we have 99.5% confidence to take the sample as faulty. Based on the elements obtained in this training process (anchor set S, healthy domain margin r_L, and the GEV model), the next part of our proposed procedure (detection process) can be computed using new data that can contain faulty behavior.

4.3.2 Detection process

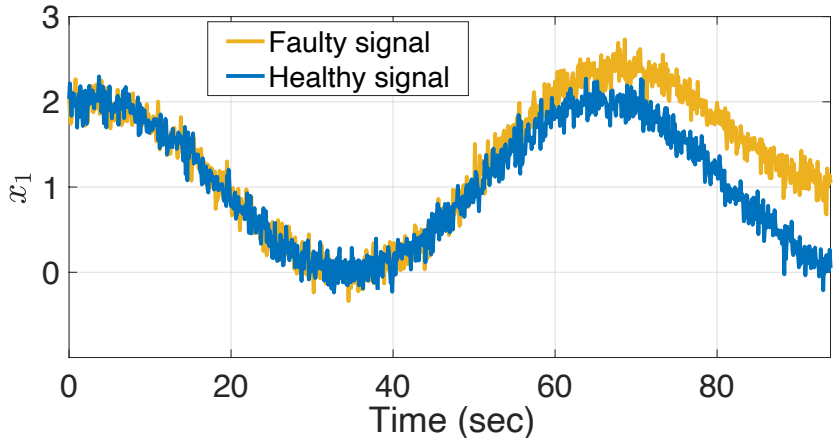

FIGURE 4.9 – Faulty and healthy x_1 simulated signal

In the detection process, we consider a new set of samples with a fault occurring in $\boldsymbol{x}_1$ at $t = 47s$, i.e., the 47th sample, with the FNR= $10dB$. As shown in Fig.4.9, the faulty signal of $\boldsymbol{x}_1$ gradually deviates from the healthy one. At the beginning of the fault (around 47 seconds), the change is tiny and is difficult to detect. To detect a fault, we first calculate the LMD values of the new samples based on the anchors set S. The LMD results of the faulty data are illustrated in Fig.4.10, where the real fault occurrence time ($t = 47s$) is marked as the green vertical dashed line, and the healthy region margin is marked as a red dashed line. In the healthy part (the left side of the green dashed line), there are several LMD values larger than the region margin, which are called false alarms. In the faulty part (beyond the green vertical dashed line), LMD index increases along with time and then exceeds the region margin, which means a fault is detected. According to this characteristic, LMD is used as the preliminary monitoring index for fault detection. Although the LMD index is effective for fault detection in the long term, missed detections still exist at the beginning of the fault occurrence (see Fig.4.10). Since the fault amplitude is subtle (incipient fault), those insensitive approaches usually can not allow the detection of incipient fault until the fault become severe. Accordingly, it is still a challenge to detect incipient faults at early stage.

To increase the fault detection accuracy and sensitivity, we focus on the distribution

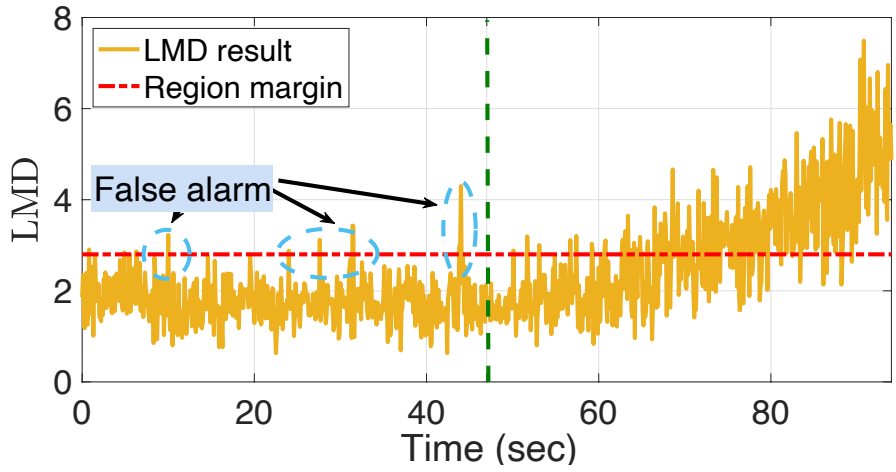

FIGURE 4.10 – LMD results for simulated signals

information of LMD values instead of a sample-based result. After fault occurrences, the distribution of LMD value will change because of faulty behavior. Since the distribution model of LMD values in healthy conditions has been developed, a more effective decision can be made by comparing the distribution of new LMD results with the healthy one. Common metrics for evaluating the difference between two distributions, like KLD and JSD [69, 217], can be used to achieve this goal. However, they usually need to estimate samples' PDF functions before calculating metrics, which leads to some potential problems, such as long computational time and low estimation accuracy. To avoid these problems, we propose a new metric called Empirical Probability Density Cumulative Sum (EPD-CUSUM). This technique is used after LMD calculation and is performed sample by sample. By accumulating the tiny difference of distributions in a recursive manner, it can detect incipient faults and achieves excellent performance, including high sensitivity, robustness and high accuracy.

For unknown data, there are two hypotheses : the healthy one $\mathcal{H}_0$ and the faulty one $\mathcal{H}_1$.

- Hypothesis $\mathcal{H}_0$: there is no fault in the observed samples.
- Hypothesis $\mathcal{H}_1$: a fault has occurred at the b_oth sample.

A fault detection procedure is then to accept or reject the hypothesis $\mathcal{H}_0$ in a test. In this process, the LMD value of an unknown sample is regarded as a random variable. To perform the hypothesis test, we first consider a single comparison event based on the LMD result and represent it by an indicator function $\mathcal{I}(x)$, such that

$$\mathcal{I}(\mathrm{x}_i) = \begin{cases} 1 & D_L(\mathrm{x}_i) > r_L \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

Under the healthy hypothesis, the probability of the event $D_L(\mathrm{x}_i) > r_L$ is equal to $1 - \alpha$. Therefore, the probability of the healthy hypothesis, denoted as $P_{\mathcal{F}_0}$, is calculated as

$$P_{\mathcal{H}_0} = \prod_{i=1}^{N}(1-\alpha) = (1-\alpha)^N \tag{4.8}$$

However, this event's probability is unknown under the healthy hypothesis. To indicate the probability of the fault occurrence in the current sample, we recursively calculate the

empirical probability density P_f via the following equation

$$P_f(\mathrm{x}_i) = (1-\omega)P_f(\mathrm{x}_{i-1}) + \omega\mathcal{I}(\mathrm{x}_i) \tag{4.9}$$

where $P_f(\mathrm{x}_0) = 0$, $\omega \in (0, 1]$ is a weight factor. Similarly, the probability of the faulty hypothesis, denoted as $P_{\mathcal{H}_1}$, is calculated as

$$P_{\mathcal{H}_1} = \prod_{i=1}^{b_o-1}(1-\alpha)\prod_{i=b_o}^{N} P_f(\mathrm{x}_i) = (1-\alpha)^{b_o-1}\prod_{i=b_o}^{N} P_f(\mathrm{x}_i) \tag{4.10}$$

The two probabilities are compared by calculating the log ratio, such that

$$\mathbb{L}(b_o, N) = \ln\left(\frac{P_{\mathcal{H}_1}}{P_{\mathcal{H}_0}}\right) = \sum_{i=b_o}^{N}\ln\left(\frac{P_f(\mathrm{x}_i)}{1-\alpha}\right) \tag{4.11}$$

We further define the instantaneous log ratio for sample x_i as :

$$\mathrm{r}_{in}(i) = \ln\left(\frac{P_f(\mathrm{x}_i)}{1-\alpha}\right) \tag{4.12}$$

Its cumulative sum from 1 to the ith sample is :

$$\mathrm{R}_{in}(i) = \sum_{1}^{i}\mathrm{r}_{in}(i) \tag{4.13}$$

By substituting (4.12) and (4.13) into (4.11), $\mathbb{L}$ can be written as :

$$\mathbb{L}(b_o, i) = \mathrm{R}_{in}(i) - \mathrm{R}_{in}(b_o - 1) \tag{4.14}$$

Since the fault occurrence time b_o is unknown, we estimate the maximum likelihood of the log ratio $\mathbb{L}$, denoted as D_e, as follows

$$D_e(\mathrm{x}_i) = \max_{1\le b_o\le i}\mathbb{L}(b_o, i) = \mathrm{R}_{in}(i) - \min_{[1\le b_o\le i]}\mathrm{R}_{in}(b_o - 1) \tag{4.15}$$

The recursive form of D_e is derived as

$$D_e(\mathrm{x}_i) = \left\{\ln\frac{P_f(\mathrm{x}_i)}{1-\alpha} + D_e(\mathrm{x}_{i-1})\right\}^{+} \tag{4.16}$$

where $\{\wp\}^+ = \wp$ for positive values, otherwise it is equal to 0. When D_e exceeds the control limit U_G, a fault is detected, and the current time is recorded as the fault detection time.

Generally, a large U_G value can reduce the risk of false alarms but increase detection delay. On one hand, to minimize the false alarm rate, we can set this parameter as the maximum D_e value of the healthy samples. On the other hand, the parameters ω should selected to adapt to the noise strength and fault severity. Generally, we should consider a small ω value for low fault severity. In practical usage, there are two ways to tune this parameter.

- When faulty samples are available, we can optimally tune the ω value to maximize the AUC criterion of D_e results.

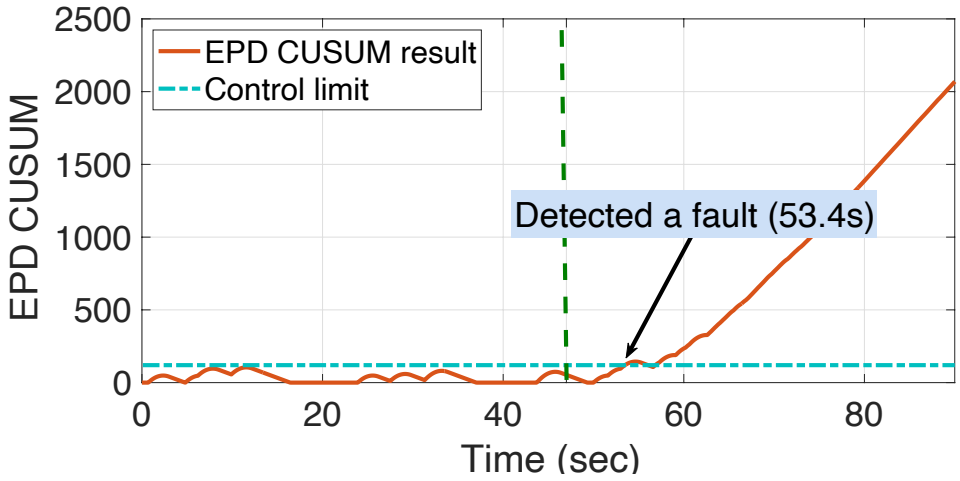

FIGURE 4.11 – EPD CUSUM results for simulated signals, where $\omega = 0.3$, $\alpha = 99.5\%$ and $U_G = 110$.

- Otherwise, ω can be set to an arbitrary value close to 0.5 to achieve excellent global detection performance for different fault severity and noise conditions.

The benefit of the EPD-CUSUM technique is twofold. It significantly avoids false alarms and missed detection since D_e will only increase with time when the LMD result exceeds r_L continuously. Besides, this technique improves detection sensitivity because small changes will be cumulative with time, yielding a striking value of D_e. The final detection result D_e of the simulation data is shown in Fig.4.11. In the healthy part of the samples, all the results are below the control limit U_G without causing false alarms. While in the faulty part (beyond the green vertical dashed line) D_e increases and exceedes the control limit at $53.4s$. A delay between the real fault occurrence time and the detected one can then be noticed ($53.4 - 47 = 6.4s$). Note that for this result, the confidence level α is 99.5%.

4.3.3 LMD feature for fault detection

The above fault detection approach based on LMD effectively recognizes incipient faults in time-series signals. However, for time-independent samples, a widely used method is to extract sample features and then analyze whether the test sample deviates from the reference using probability-based distance [40, 66]. Statistic-based techniques are frequently used to extract meaningful information from raw samples without faults' prior knowledge . For example, principal component analysis can refine the faulty information by projecting data onto an orthogonal and lower-dimensional healthy space preserving the maximum variance [82,177]. Subsequently, probability-based distance, like Kullback-Leibler divergence, is used to measure the tiny dissimilarities between observed and reference probability distributions [21, 55, 86]. Nevertheless, the major limitation of the existing detection approaches should be pointed out : PCA is unsuitable for non-linear data and will cause the loss of information, sometimes degenerating this method's accuracy performance. As introduced in chapter 3, the proposed healthy region approximation method can effectively highlight the tiny change of incipient fault by LMD index and it is effective for non-linear data. Thus, the LMD index can be used as a representative feature for the fault detection task.

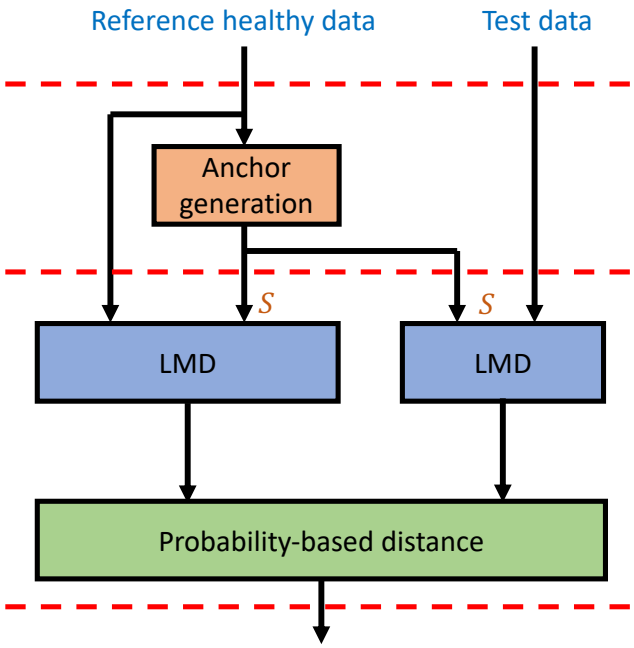

FIGURE 4.12 – Fault detection procedure based on LMD feature and probability-based distance

This subsection discusses the benefits of using the LMD index as a feature and the probability-based distance as feature analysis tools. This proposed detection procedure is displayed in Fig.4.12. The healthy data is first used to generate anchors for the LMD calculation. Next, the LMD features of the healthy reference and test data are extracted, respectively. Finally, the probability-based distance of the two LMD features, such as Kullback-Leibler divergence (D_{kl}) [66], Jensen-Shannon divergence (D_{js}) [217], Wasserstein distance (D_w) [31], and Kolmogorov-Smirnov distance (D_{ks}) [60], are evaluated for detection purposes.

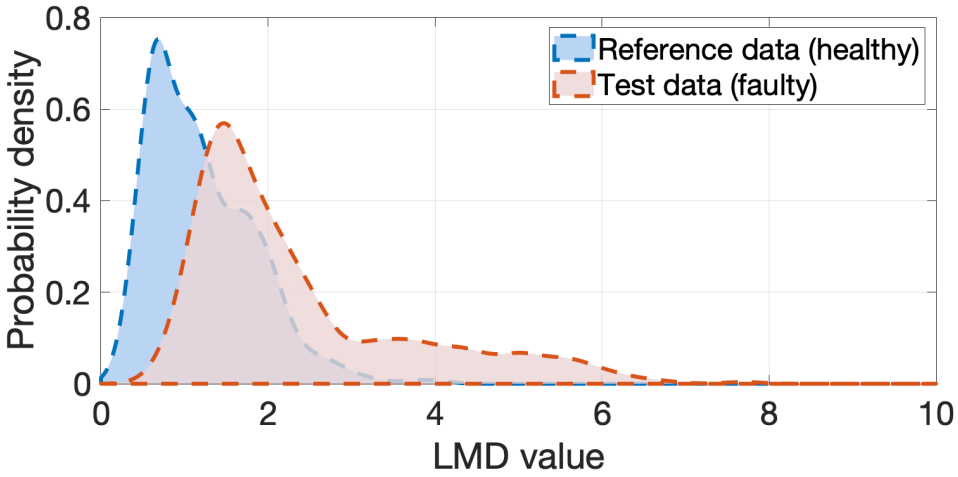

FIGURE 4.13 – Distribution of LMD features for healthy and faulty cases.

We consider the simulation data given in section 4.2 as an example. Fig.4.6 shows the extracted LMD feature of healthy simulation data, and Fig.4.10 illustrates the LMD feature of the faulty case. The distribution of the two features is exhibited in Fig.4.13,

showing that their distributions are significantly different. Fig.4.14 gives the detection outcomes of using different probability-based distances to measure the dissimilarity of LMD features. Notes that the first 100 realizations are healthy samples, while the last 100 are faulty. In this example, all the probability-based distances allow a correct detection of a fault since their detection outcomes for the faulty samples are significantly larger than for healthy ones. However, their sensitivities are different, which implies different detection capabilities and robustness to noise. By using the Wasserstein distance, the change of the detection outcomes after a fault occurs is the most obvious. It can be expected that this distance may still maintain a high detection accuracy if the noise power increases or the fault severity becomes small. Conversely, the Jensen-Shannon divergence has relatively close detection outcomes for healthy and faulty samples, indicating a high risk of false detection if fault severity is small or the environment is noisy. A deeper study of these detection strategies' sensitivity will be considered in the next chapter.

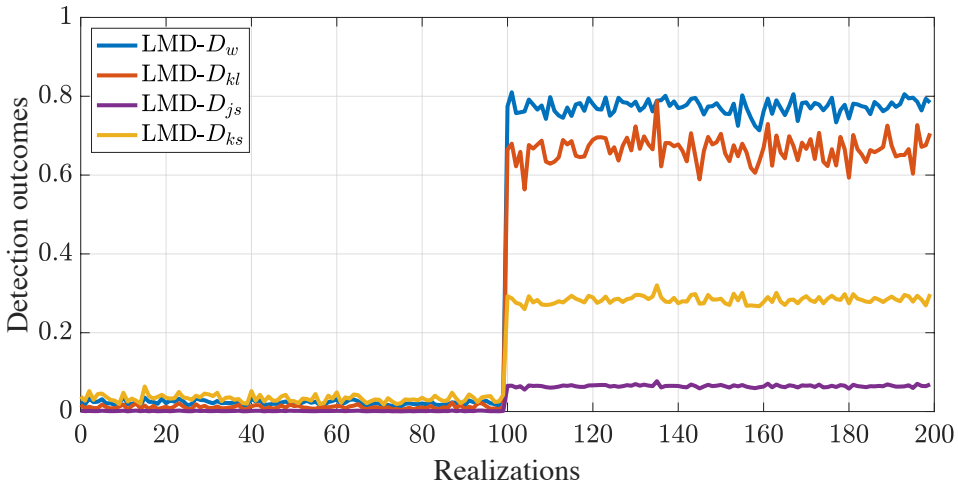

FIGURE 4.14 – Detection outcomes of LMD feature using different probability-based distances

Generally, this toy example shows a real benefit of using the LMD feature for fault detection. Its efficiency will be further studied in the next section. Besides, we will also deeply discuss the detection capability of different probability-based distances.

4.4 Contribution plot for faulty variables isolation

Let's assume that a fault has been detected by the previous detecting process. Subsequently, in this section, we first consider a single faulty variable denoted as ι and propose a LMD-based fault isolation method to identify the faulty source. In the LMD-based fault detection scheme, the corresponding anchor of a sample is first located, serving as a reference position for fault analysis. When a fault occurs, the faulty samples obviously deviate from their references, while healthy samples are close to the corresponding anchors. These two cases can be distinguished by calculating the Mahalanobis distance between samples and anchors (LMD). In other words, the distance information is used for the fault detection task. Similarly, we consider the change of samples' relative position for the fault isolation task. The position's change is based on observed samples and their corresponding

anchors, where the latter has been obtained in the previous detection step.

Let the local anchor of a sample be $\mathcal{A}(\mathrm{x})$, which is written as :

$$\mathcal{A}(\mathrm{x}_i) = \arg\min_k \{d_M(\mathrm{x}_i, y_k) | y_k \in S\} \tag{4.17}$$

Given a sample vector x_i and its corresponding anchor vector $\mathcal{A}(\mathrm{x}_i)$, the position's change with normalization weight is presented as :

$$\mathcal{G}_i = (\mathrm{x}_i - \mathcal{A}(\mathrm{x}_i))^T \mathcal{W} \tag{4.18}$$

where $\mathcal{W}$ is the diagonal matrix consisting of the standard deviation σ_j of each healthy variable, such that :

$$\mathcal{W} = \mathrm{diag}\{\sigma_1^{-1}, \cdots, \sigma_m^{-1}\} \tag{4.19}$$

Note that the weight is used here to unify the scale of different variables.

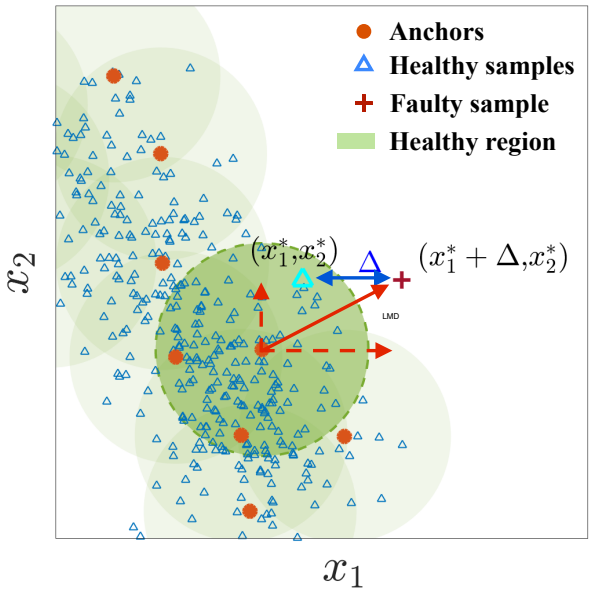

FIGURE 4.15 – Two-dimensional example for LMD based fault isolation

As illustrated in Fig.4.15, we use a toy example with two variables to show how the position's change information contributes to faulty source isolation. In fact, when a variable is affected by a fault, the faulty samples will deviate in a certain direction. In this example, the first variable x_1 is affected, and thus the first element of the sample vector increases from x_1^* to $(x_1^* + \Delta)$. We show the faulty sample and its healthy version to highlight this change. However, the increment Δ in x_1 direction is invisible for the unknown healthy version until introducing a reference point, the anchor. For healthy cases, since samples are close to their anchors, the deviation of samples with respect to their anchors in different directions is small. In contrast, faulty samples are far away from their anchors. With the single fault assumption, there is always a distinct shift of faulty samples in one direction. In other words, the single faulty variable can effectively be isolated by observing the increment of faulty samples in different directions, which is reflected by the change vector $\mathcal{G}$.

Regarding the incipient faults cases, the reliability of sample-based results $\mathcal{G}_i$ is easily impacted by noise, especially at the fault early stage. Therefore, a number of sequential

faulty samples are used for the final decision to increase the reliability and isolation accuracy. Suppose a fault is detected at the b_dth observation, and then the average position change vector can be calculated as :

$$\bar{\mathcal{G}} = \sum_{i=b_d}^{N} \mathcal{G}_i \tag{4.20}$$

Let denote $\bar{\mathcal{G}} = [\mathrm{g}_1, \cdots, \mathrm{g}_j, \cdots, \mathrm{g}_m]$. The Softmax function is further applied to the elements of $\bar{\mathcal{G}}$ for the unifying purpose. Then the fault contribution Con_j of each variable is calculated as :

$$\mathrm{Con}_j = \frac{e^{|\mathrm{g}_j|}}{\sum_{l=1}^{m} e^{|\mathrm{g}_l|}} \tag{4.21}$$

Finally, for a single faulty variable case, the faulty variable index number is recognized as :

$$\hat{\imath} = \arg\max_j \mathrm{Con}_j \tag{4.22}$$

The corresponding $\boldsymbol{x}_{\hat{\imath}}$ faulty variable is then isolated.

As an example, we apply the proposed isolation approach to the simulation data defined in section 4.2, where a fault with 20dB FNR was introduced in the 3rd signal $x_3(t)$. Fig.4.16 illustrates the fault contribution plot of each variable. The 3rd contribution value is the largest, and thus the 3rd signal is recognized as faulty. In this example, the 3rd contribution value is significantly larger than the rest, indicating that the isolation result given by this approach is highly reliable. Indeed, this approach still suffers from the effect of heavy noise and small fault amplitude. Detailed performance will be evaluated in the next chapter to reveal how these factors affect this approach. Remarkably, the proposed isolation method is dedicated to the single faulty source situation. To deal with the multiple fault cases, an improved approach will be proposed in the following section.

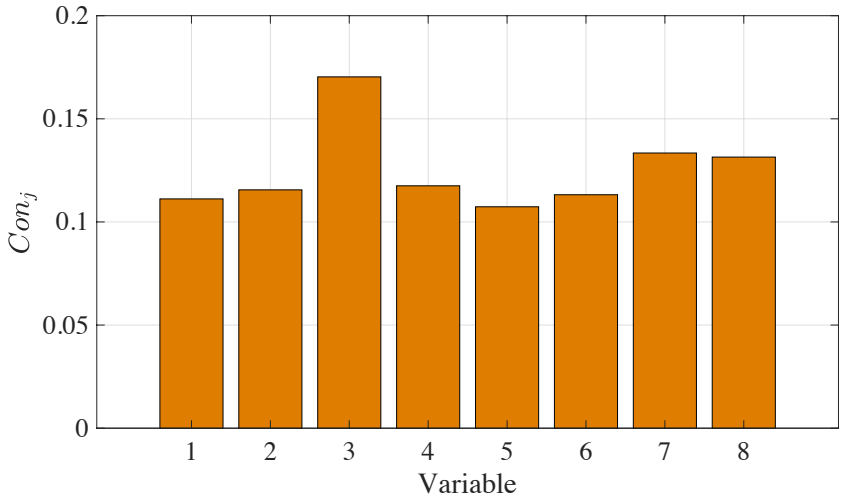

FIGURE 4.16 – Contribution plot for simulation signals

4.5 Fault increasing rate estimation based on first order model

Once the faulty variable is isolated, the fault severity estimation procedure is activated. This section is dedicated to estimate the fault's increasing rate based on first order model. More precisely, we estimate the coefficient δ of the first-order approximation written in Eq.(4.3) by explicitly establishing its expression based on the LMD index. To that end, we develop the LMD calculation equation by substituting Eq.(4.17) and Eq.(4.1) into Eq.(3.8), such that

$$\begin{aligned} D_L(\mathrm{x}_i) &= d_M(\mathrm{x}_i, \mathcal{A}(\mathrm{x}_i)) \\ &= \sqrt{[\mathrm{x}_i - \mathcal{A}(\mathrm{x}_i)]^T \Sigma^{-1} [\mathrm{x}_i - \mathcal{A}(\mathrm{x}_i)]} \\ &= \| \, [\mathrm{x}_i - \mathcal{A}(\mathrm{x}_i)]^T \Sigma^{-\frac{1}{2}} \, \|_2 \\ &= \| \, [\mathrm{x}_i^* + \mathrm{x}_i^f - \mathcal{A}(\mathrm{x}_i)]^T \Sigma^{-\frac{1}{2}} \, \|_2 \end{aligned} \tag{4.23}$$

where $\| \cdot \|_2$ corresponds to the ℓ^2 norm.

For simplicity, we define the following three notations :

$$\boldsymbol{\zeta}_i = \mathrm{x}_i^* - \mathcal{A}(\mathrm{x}_i) \tag{4.24}$$

$$\boldsymbol{\zeta}_i^{'} = \boldsymbol{\zeta}_i \Sigma^{-\frac{1}{2}} = [\zeta_{i1}^{'}, \cdots, \zeta_{ij}^{'}, \cdots, \zeta_{im}^{'}] \tag{4.25}$$

$$\Sigma^{-\frac{1}{2}} = \begin{bmatrix} \varsigma_{11} & \cdots & \varsigma_{1j} & \cdots & \varsigma_{1m} \\ \vdots & & \vdots & & \vdots \\ \varsigma_{j1} & \cdots & \varsigma_{jj} & \cdots & \varsigma_{jm} \\ \vdots & & \vdots & & \vdots \\ \varsigma_{m1} & \cdots & \varsigma_{mj} & \cdots & \varsigma_{mm} \end{bmatrix} \tag{4.26}$$

With the above notations and first-order approximation model Eq.(4.3), we further develop Eq.(4.23) to Eq.(4.27).

$$\begin{aligned} D_L^2(\mathrm{x}_i; S) =& \| \, \zeta_i^{'} \, \|_2^2 + 2\Delta_i \sum_{j=1}^{m} \zeta_{ij}^{'} \varsigma_{cj} + \Delta_i^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \\ =& \| \, \zeta_i^{'} \, \|_2^2 + 2\delta \left(i - b_d + 1\right) T_e \sum_{j=1}^{m} \zeta_{ij}^{'} \varsigma_{cj} + \left[\delta \left(i - b_d + 1\right) T_e\right]^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \\ =& \, i^2 \left(\delta^2 T_e^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \right) + 2i \left[\delta T_e \sum_{j=1}^{m} \zeta_{ij}^{'} \varsigma_{cj} + \left(1 - b_d\right) \delta^2 T_e^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \right] \\ &+ 2\left(1 - b_d\right) \delta T_e \sum_{j=1}^{m} \left(\varsigma_{cj}^2 + \zeta_{ij} \varsigma_{cj} \right) + \| \, \zeta_i^{'} \, \|_2^2 \end{aligned} \tag{4.27}$$

Although Eq.(4.27) establishes the link between δ and the LMD index, δ can not be calculated directly through this equation for the unknown $\zeta_{ij}^{'}$. However, Eq.(4.27) is a quadratic function of sample index i, which can be simplified as :

$$D_L^2(\mathrm{x}_i; S) = \varrho_2 i^2 + \varrho_1 i + \varrho_0 \tag{4.28}$$

where

$$\varrho_2 = \delta^2 T_e^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \tag{4.29}$$

$$\varrho_1 = 2\delta T_e \sum_{j=1}^{m} \zeta_{ij}^{'} \varsigma_{cj} + 2(1-b)\delta^2 T_e^2 \sum_{j=1}^{m} \varsigma_{cj}^2 \tag{4.30}$$

$$\varrho_0 = 2(1-b_d)\delta T_e \sum_{j=1}^{m} (\varsigma_{cj}^2 + \zeta_{ij}\varsigma_{cj}) + \parallel \zeta_i^{'} \parallel_2^2 \tag{4.31}$$

In fact, $D_L^2(\mathrm{x}_i)$ has been calculated for fault detection, and i is known. Therefore, the coefficients $\varrho_0, \varrho_1, \varrho_2$ can be easily estimated by the polynomial curves fitting technique. More accurately, the solution for Eq.(4.28) is obtained by minimizing the sum of squared errors between the true and estimated D_L^2 values based on samples from b_d to N. In order to derive the calculation process, we let :

$$\Upsilon = [\varrho_0, \varrho_1, \varrho_2] \tag{4.32}$$

$$\Psi^T = \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ b_d & \cdots & i & \cdots & N \\ b_d^2 & \cdots & i^2 & \cdots & N^2 \end{bmatrix} \tag{4.33}$$

and

$$Y = [D_L^2(\mathrm{x}_{b_d}), \cdots, D_L^2(\mathrm{x}_i), \cdots, D_L^2(\mathrm{x}_N)]^T \tag{4.34}$$

The coefficients vector can be calculated as :

$$\Upsilon = (\Psi^T \Psi)^{-1} \Psi^T Y \tag{4.35}$$

Finally, the fault severity represented by δ is estimated as $\hat{\delta}$ as follows :

$$\hat{\delta} = \sqrt{\frac{\varrho_2}{T_e^2 \sum_{j=1}^{m} \varsigma_{cj}^2}} \tag{4.36}$$

As an example, we simulate multivariate data using Eq.4.6 and introduce a fault with different δ values in the 3rd signal.We apply the proposed estimation approach to the simulation data and obtain the result as shown in Fig.4.17. As δ increases, the estimated error gets closer to 0, meaning that the estimated value is close to the actual one. For extremely small values of δ ($\delta < 0.3$), the tiny change is easily confused with noise, resulting in the weak estimating capability (the relative error is larger than 0.1). It is noticed that the relative error is always negative, i.e., the result is underestimated, which should be avoided. Nevertheless, as the error is so small, this will not lead to a major issue. To overcome this limitation, we will improve the fault severity estimation approach in the next section. In summary, Fig.4.17 shows that the proposed method is still effective and can obtain accurate results for tiny deviations.

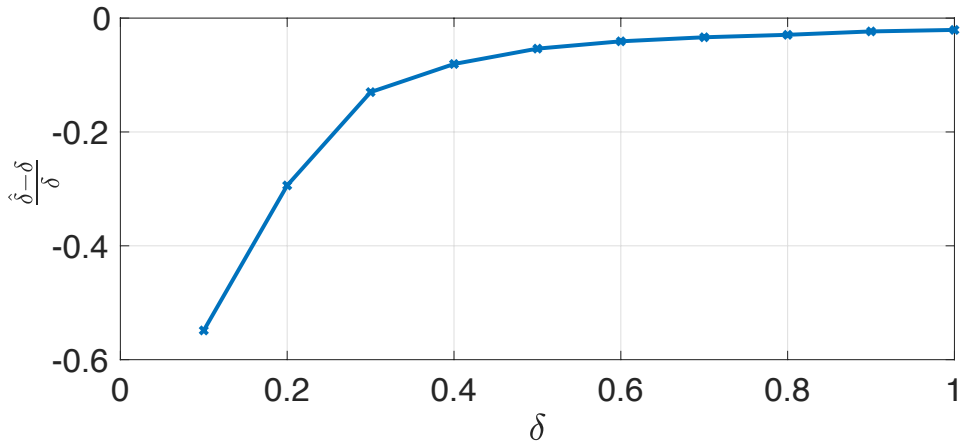

FIGURE 4.17 – Relative error of the proposed method for different increasing rates

4.6 Reconstruction based contribution for faulty variables isolation and amplitude estimation

Although the proposed isolation approach and fault increasing rate estimation method are efficient for simulation data, they have numerous limitations. First, the LMD-based contribution approach is mainly developed for single fault cases. This method may yield an unexpected isolation result when multiple faults occur simultaneously. Then isolation approach should be improved for multiple faults cases. On the other hand, the previously proposed LMD-based estimation method assesses the severity degree by estimating the fault's increasing rate, which is simple but lacks information on the deviation degree from healthy behavior. A more effective way to assess fault severity is to estimate fault amplitude. The performance of the LMD-based increasing-rate estimation approach usually degenerates if the fault does not satisfy the first-order model assumption. Therefore, to overcome these limitations, we propose an improved solution based on the LMD technique and the reconstruction-based contribution (RBC) framework for faulty variable isolation and fault amplitude estimation.

The RBC combines the fault reconstruction idea and contribution plot idea. The goal is to isolate multiple faulty variables and estimate the fault amplitude more accurately. Fig.4.18 shows the flow diagram of the RBC framework based on LMD. In this framework, a signal is reconstructed by removing its faulty component if the fault direction is known. Therefore, given a fault direction $\boldsymbol{\Theta}^l$ from a candidate set of fault directions $\boldsymbol{\Xi} = \{\boldsymbol{\Theta}^1, \boldsymbol{\Theta}^2, \cdots, \}$, the reconstructed signal of the ith sample $\boldsymbol{x}_i$ is obtained as

$$\boldsymbol{x}_i^l = \boldsymbol{x}_i - \boldsymbol{\Delta}_i^l \boldsymbol{\Theta}^l \tag{4.37}$$

For practical applications, candidate fault directions are finite and easily determined according to the known number of faulty situations. When the true fault direction is selected, the fault detection index of the reconstructed-healthy signal will be lower than the detection threshold. Otherwise, the detection index of a reconstructed signal is larger than the threshold.

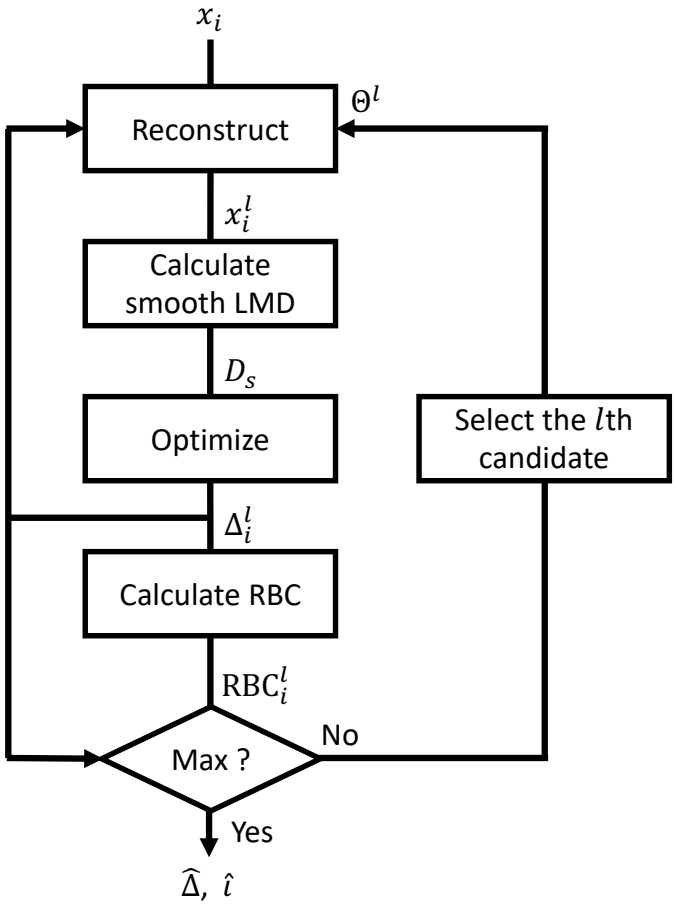

FIGURE 4.18 – Flow diagram of RBC framework based on LMD

The LMD index of the reconstructed signal is calculated as

$$\begin{aligned} D_L(\boldsymbol{x}_i^l) &= (\boldsymbol{x}_i^l - \mathcal{A}(\boldsymbol{x}_i^l))^T \boldsymbol{M} (\boldsymbol{x}_i^l - \mathcal{A}(\boldsymbol{x}_i^l)) \\ &= \|\boldsymbol{M}^{\frac{1}{2}} (\boldsymbol{x}_i - \boldsymbol{\Delta}_i^l \boldsymbol{\Theta}^l - \mathcal{A}(\boldsymbol{x}_i^l))\|^2 \end{aligned} \tag{4.38}$$

where $\boldsymbol{M} = \Sigma^{-1}$. The fault direction $\boldsymbol{\Theta}^l$ is constant during a faulty situation, and the fault amplitude $\boldsymbol{\Delta}_i^l$ develops slowly according to the definition of incipient faults. However, the item $\boldsymbol{x}_i - \mathcal{A}(\boldsymbol{x}_i^l)$ in Eq.4.38 changes dynamically due to noise disturbance and the intrinsic approximation error resulting from anchor calculation, which hinders extract faulty information and leads to unstable results of the isolation procedure. Therefore, to inhibit the high-frequency change of this item, we define an exponential smoothing index ξ_i^l as

$$\xi_i^l = (1 - \zeta_s)\xi_{i-1}^l + \zeta_s(\boldsymbol{x}_i - \mathcal{A}(\boldsymbol{x}_i^l)), \quad (\xi_0 = 0) \tag{4.39}$$

where $0 \leq \zeta_s \leq 1$ is the smoothing factor. Then the smoothing LMD index for reconstructed signals is derived as

$$D_s(\boldsymbol{x}_i^l) = \|\boldsymbol{M}^{\frac{1}{2}} (\xi_i^l - \boldsymbol{\Delta}_i^l \boldsymbol{\Theta}^l)\|^2 \tag{4.40}$$

Generally, a small value of ζ leads to smooth LMD results but a slow response to signals' changes.

Given a fault direction $\boldsymbol{\Theta}^l$, the objective of the reconstruction is to determine the fault component $\boldsymbol{\Delta}_i^l$ minimizing Eq.(4.40), which is a quadratic optimization problem without

constraint. However, since the anchor calculation $\mathcal{A}(\cdot)$ is not an analytical function of $\boldsymbol{\Delta}_i^l$, we can not directly obtain the optimal solution of Eq.(4.40). As a result, we propose the iterative method alternating between performing the following two steps to obtain an approximated solution.

1. Anchor updating step : determine the anchor of reconstructed signal $\boldsymbol{x}_i$ using the optimal solution of $\boldsymbol{\Delta}_i^l$

$$\mathcal{A}'(\boldsymbol{x}_i^l) = \arg\min_{\boldsymbol{y}\in S} d_M(\boldsymbol{x}_i - \boldsymbol{\Delta}_i^l\boldsymbol{\Theta}^l, \boldsymbol{y}) \tag{4.41}$$

   Initially, we let $\boldsymbol{\Delta}_i^l\boldsymbol{\Theta}^l = \boldsymbol{0}$.

2. Optimization solving step : calculate the optimal solution minimizing Eq.(4.40) as

$$\boldsymbol{\Delta}_i^l = (\boldsymbol{\Theta}^{lT}\boldsymbol{M}\boldsymbol{\Theta}^l)^{-1}\boldsymbol{\Theta}^{lT}\boldsymbol{M}\xi_i^l \tag{4.42}$$

   where ξ_i^l is updated by Eq.(4.39) with the result $\mathcal{A}'(\boldsymbol{x}_i^l)$ obtained in the previous step.

The two steps are executed till the solution converges.

Substitute the solution Eq.(4.42) into Eq.(4.40), we have

$$\begin{aligned} D_s(\boldsymbol{x}_i^l) &= \|(\boldsymbol{I} - \boldsymbol{M}^{\frac{1}{2}}\boldsymbol{\Theta}^l(\boldsymbol{\Theta}^{lT}\boldsymbol{M}\boldsymbol{\Theta}^l)^{-1}\boldsymbol{\Theta}^{lT}\boldsymbol{M}^{\frac{1}{2}})\boldsymbol{M}^{\frac{1}{2}}\xi_i^l\|^2 \\ &= \|\boldsymbol{R}^l\boldsymbol{M}^{\frac{1}{2}}\xi_i^l\|^2 \end{aligned} \tag{4.43}$$

where $\boldsymbol{R}^l \triangleq \boldsymbol{I} - \boldsymbol{M}^{\frac{1}{2}}\boldsymbol{\Theta}^l(\boldsymbol{\Theta}^{lT}\boldsymbol{M}\boldsymbol{\Theta}^l)^{-1}\boldsymbol{\Theta}^{lT}\boldsymbol{M}^{\frac{1}{2}}$. When the correct fault direction is selected, i.e., $l = \iota$, the isolation index $D_s(\boldsymbol{x}_i^\iota)$ will be lower than the threshold r_L, while the isolation index of other cases will be higher than the threshold, such that

$$D_s(\boldsymbol{x}_i^\iota) < r_L \leq D_s(\boldsymbol{x}_i^l), \quad (l \neq \iota) \tag{4.44}$$

Therefore, the fault direction is identified by searching the minimal isolation index $D_s(\boldsymbol{x}_i^l)$. Further, according to the traditional RBC idea [49, 85], the reconstruction-based contribution of a candidate fault direction is defined as

$$\begin{aligned} \mathrm{RBC}_i^l &\triangleq \|\boldsymbol{M}^{1/2}\boldsymbol{\Delta}_i^l\boldsymbol{\Theta}^l\|^2 \\ &= \xi_i^{lT}\boldsymbol{M}\boldsymbol{\Theta}^{lT}(\boldsymbol{\Theta}^{lT}\boldsymbol{M}\boldsymbol{\Theta}^l)^{-1}\boldsymbol{\Theta}^{lT}\boldsymbol{M}\xi_i^l \end{aligned} \tag{4.45}$$

Equivalently, the fault direction is identified by searching the maximal reconstruction-based contribution

$$\hat{\iota}_i = \arg\max_l \mathrm{RBC}_i^l \tag{4.46}$$

Similar to the LMD-based contribution plot method, the Softmax function (Eq.4.21) can be further applied to the above reconstruction-based contribution result for the unifying purpose. With the fault direction being isolated, its corresponding fault amplitude is simultaneously obtained as

$$\hat{\boldsymbol{\Delta}}_i = (\boldsymbol{\Theta}^{\hat{\iota}_i T}\boldsymbol{M}\boldsymbol{\Theta}^{\hat{\iota}_i})^{-1}\boldsymbol{\Theta}^{\hat{\iota}_i T}\boldsymbol{M}\xi_i^{\hat{\iota}_i} \tag{4.47}$$

We generate simulation data using Eq.4.6 and introduce faults with 20dB FNR in the 2nd and 3rd variables. We apply the proposed RBC approach to the data for faulty variable isolation and fault amplitude estimation. Fig.4.19 and Fig.4.20 show the obtained

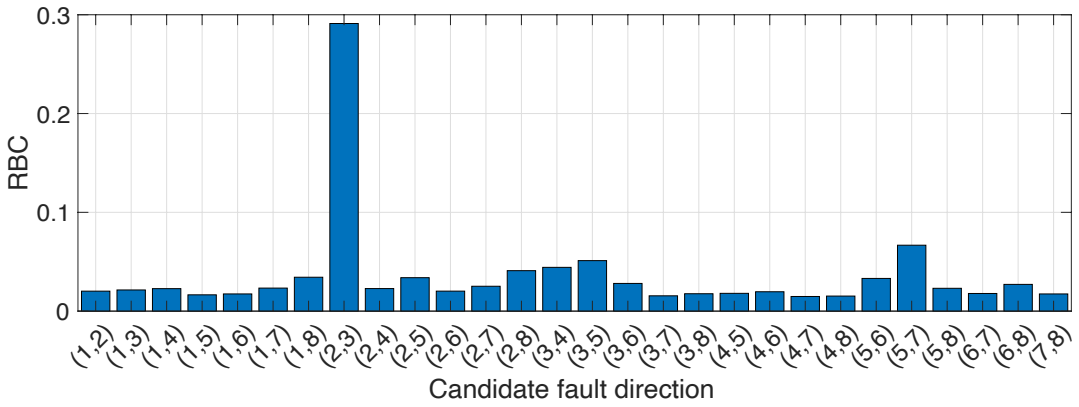

FIGURE 4.19 – Reconstruction-based contributions of each candidate direction

isolation and estimation results, respectively. For the isolation task, the contribution value of the fault direction $(2, 3)$ is the largest, and thus its corresponding variables (2 and 3) are recognized as faulty. Concerning the estimation performance of the RBC approach, Fig.4.20 shows that the relative errors of the estimation for the two faulty variables are large at the beginning but converge with time to a small value close to 0. Indeed, when faults just occur, their amplitudes are too small to estimate, which thereby leads to large errors in the results. However, the estimation results converge with time, and their relative errors are stable between -0.1 and 0.1, meaning their absolute percentage errors are around 10% of the original values. Compared to the increasing rate estimation, this approach does not always underestimate the fault severity. Therefore the proposed method is effective and practical in the fault amplitude estimation task. In the next chapter, we will further investigate the estimation performance of this approach concerning different noise levels and fault severities. Then its superiority will be highlighted by comparing it with other estimation methods.

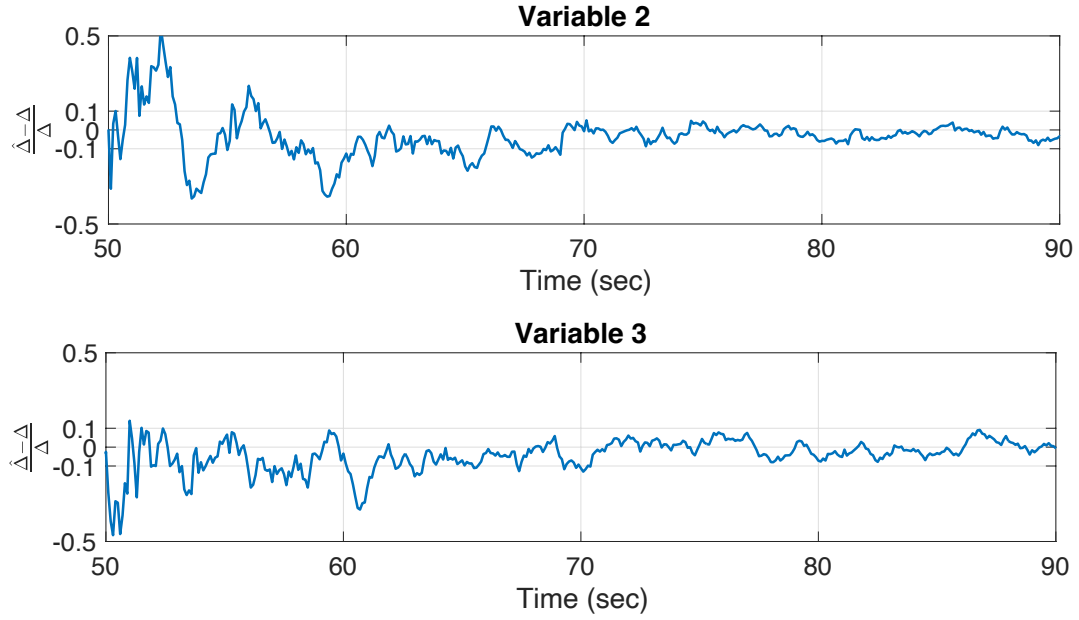

FIGURE 4.20 – Relative error of the approach estimation approach based on RBC for simulation data

4.7 Conclusion

This chapter proposes a specific incipient fault diagnosis methodology for incipient faults in a noisy environment. It is based on a procedure composed of two main parts : the training process and the fault diagnosis process. In the training process, we propose a specific unsupervised healthy domain approximation method using the Mahalanobis distance in a particular way. It is composed of an optimized anchor-generation algorithm and a special margin selection approach. Compared with PCA-based Hotelling's T^2 and SPE (Q) statistics, the proposed healthy domain approximation method can effectively handle non-elliptically distributed data. Based on the approximated healthy domain characteristics, the detecting process is derived and the local Mahalanobis distance is defined and serves as a preliminary monitoring index for fault detection. Subsequently, we derive this sensitive incipient fault detection framework by combining Local Mahalanobis Distance with the improved Empirical Probability Density Cumulative Sum method. Besides, we also show that LMD can be used as an effective feature for fault detection. Probability-based distance is then used to decide if test samples deviate from the healthy pattern.

After detecting a fault, we consider the faulty variable isolation problem and propose an approach based on the contribution plot idea. This approach is developed on the LMD index and its anchor generation algorithm. By analyzing the relative position between faulty samples and their corresponding anchors, we can intuitively discover which variable is relative to the fault. The proposed fault isolation method can recognize the single faulty variable, even for tiny faults. Subsequently, we derive the LMD calculation equation by substituting the first-order fault model. Thanks to a theoretical study, the fault severity estimation approach analytically establishes the relation between the increasing rate of a fault and the LMD index. This offers an effective way to estimate the severity of incipient faults. These two methods preserve the intrinsic advantages of LMD, such as robustness to outliers, distribution-free assumption, and high sensitivity for incipient faults. Together with the LMD-based fault detection method, they compose the LMD-based incipient fault diagnosis framework. In order to improve the isolation and fault severity estimation performance, we propose a novel reconstruction-based method using the LMD index for incipient faults. This method combines the high sensitivity characteristics of the LMD methodology for incipient faults with the RBC technique's good properties, which has a solid theoretical foundation for faulty variables isolation and fault amplitude estimation.

This chapter uses the simulation data with 8 signals as an example to validate the effectiveness of all the proposed approaches. The calculation results in each key step of the approaches are given to clearly show how our proposal works and how effective they are. The primary results indicate that the contribution-plot-based isolation method can accurately recognize a single faulty variable. The fault increasing rate estimation approach based on the first-order fault model effectively estimates the increasing rate of a fault. As the fault increase, the error of fault estimation approach significantly reduces. Finally, the simulation results show that the RBC-based approach can accurately identify faulty variables and estimate the their amplitude simultaneously. These results based on the simulation data validate the effectiveness of our proposal. In the next chapter, we will validate the proposed methodology using 2 types of industrial application data. The detailed performance evaluation will be given and will highlight the limits of the proposal.

5

Applications to industrial engineering

5.1 Introduction

The last chapter presented a series of effective methodologies for different fault diagnosis tasks based on LMD techniques and validated their effectiveness based on simulation data. However, there is always a gap between simulation and practical industrial applications. The industrial data usually follows a non-Gaussian complex distribution and is contaminated by strong noise. Therefore, this chapter aims to comprehensively evaluate the efficiency of the proposed methods using industrial data. Prior to the efficiency evaluation, performance criteria are first given. Then, based on the Continuous-flow Stirred Tank Reactor (CSTR) process data, we investigate the efficiency of the proposed fault diagnosis solutions developed on LMD for different diagnosis tasks. The performance analysis concerning different fault severity and noise level is performed to show the comprehensive diagnostic capability of our proposals. We also compare our proposals with state-of-the-art approaches, such as methods based on Kullback-Leibler divergence and Jensen-Shannon divergence, RBC methods based on principal component analysis, and Mahalanobis distance. Moreover, we evaluate the feature extraction efficiency of the LMD technique by using bearing data. In this study, LMD is used as a feature extraction tool, and the probability-based distance is subsequently used to detect a fault. By combining different features and probability-based distances, we then show the benefit of using the LMD feature.

5.2 Evaluation criteria

Fault severity and noise level are two major external factors affecting the performance of fault diagnosis approaches, which should be studied carefully in performance evaluation. To take into account the effect of fault severity and noise strength simultaneously, the Fault-to-Noise Ratio (FNR) is used as the quantitative metric of the fault severity regarding

the noise strength [38]. It is defined as :

$$\mathrm{FNR} = 10 \log \frac{p_f}{p_v} \tag{5.1}$$

where p_f is the fault power and p_v is the noise power. Note that incipient faults are defined for FNR values lower than 5dB. In that range, these faults can be partially or totally masked by the noise. Similarly, the Signal-to-Fault Ratio (SFR) is defined to quantify fault severity regarding the signal power :

$$\mathrm{SFR} = 10 \log \frac{p_s}{p_f} \tag{5.2}$$

where p_s is the signal power. In the particular case of incipient fault, the SFR values are large. The above two ratios are related to the common metrics SNR by the following equation :

$$\mathrm{SNR} = \mathrm{SFR} + \mathrm{FNR} \tag{5.3}$$

The detection performance evaluation based on samples considers three criteria : the detection probability (P_d), the false alarm one (P_{FA}), and the accuracy (ACC_D) which are defined as :

$$P_d = \frac{\text{No. of samples}\{\text{detected as faulty}|\text{faulty}\}}{\text{No. of }\{\text{All the faulty samples}\}} \tag{5.4}$$

$$P_{FA} = \frac{\text{No. of samples}\{\text{detected as faulty}|\text{healthy}\}}{\text{No. of }\{\text{All the healthy samples}\}} \tag{5.5}$$

$$\mathrm{ACC}_D = \frac{\text{No. of }\{\text{collectly detected samples}\}}{\text{No. of }\{\text{All the samples}\}} \times 100\% \tag{5.6}$$

Additionally, the Receiving Operating Characteristics (ROC) curve [51] and its corresponding Area Under Curve (AUC) value are also used to assess the detection performance in terms of robustness and efficiency. The evaluation of AUC is independent of the threshold selection but focuses on the global detection capability of an index. It allows judging the performance of the two methodologies conveniently. Usually, AUC value is between 0 and 1. The value 1 indicates perfect detection performance ; value 0.5 means the diagnosis index cannot distinguish between healthy and faulty samples ; and the value 0 shows inverse detection results. Moreover, in the early detection context focusing on incipient faults, an overview of the reaction speed of a fault detection method is also an important criterion. Suppose that a fault occurs at b_o and the detection time is b_d, the detection delay DD calculated as the difference of these two values will allow reflecting the method response speed. It is written as :

$$\mathrm{DD} = b_o - b_d \tag{5.7}$$

The synthetical performance evaluation of fault detection approaches is performed by using the above criteria while changing the fault severity and noise level. However, if a data set does not contain sufficient faulty samples, such as smaller faulty severity, methods' detection capability for unknown faulty situations cannot be evaluated. Therefore, we use the detection sensitivity criterion to evaluate the potential detection capability of a method. The detection sensitivity criterion is defined as

$$\mathrm{Sen} = \frac{\bar{D}_f - \bar{D}_h}{\max(D_h) - \bar{D}_h} \tag{5.8}$$

where $\bar{D}_h$ is the average value of any diagnosis index in a healthy condition, and $\bar{D}_f$ is the average value in faulty conditions. When the detection sensitivity criterion value is less than 1, the miss detection rate will be too high and unacceptable for the fault detection task. The detection sensitivity value should be larger than 2 to achieve a reliable detection performance [69]. With higher value, it usually indicates a more powerful detection capability.

In the faulty variables isolation task, performance can be simply evaluated by the isolation accuracy criterion calculated as

$$\mathrm{ACC}_{Iso} = \frac{\text{No. of samples}\{\text{correct isolation}\}}{N_f} \times 100\% \tag{5.9}$$

Further, we calculate the confusion matrix to allow the visualization of the isolation performance. It is a table whose row instances indicate the predicted results while each column is the actual one. Each element of the confusion matrix is the probability that an actual faulty variable is predicted to be a particular variable. The confusion matrix makes it easy to conclude if an isolation approach confuses two classes.

Concerning the fault amplitude estimation performance, we can easily assess the mean square error (MSE) between the actual and estimated fault amplitude, such that

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N_f}(\Delta_i - \hat{\Delta}_i)^2}{N_f} \tag{5.10}$$

The relative error is used to evaluate the estimation performance, and it can unify the estimation error of the different studied faulty cases. It is defined as

$$\mathrm{RE} = \frac{\sum_{i=1}^{N_f} \frac{\hat{\Delta}_i - \Delta_i}{\Delta_i}}{N_f} \times 100\% \tag{5.11}$$

Generally, smaller MSE or RE value indicates better estimation performance. In the following, we further analyze our proposal performance using these criteria in different diagnosis tasks.

5.3 Performance evaluation of fault diagnosis approaches based on process data

This section focuses on the performance of the proposed fault diagnosis approaches based on LMD. The CSTR process is first introduced, providing basic information on data and faults. We discuss fault diagnosis performance in three tasks : fault detection, faulty variable isolation, and fault severit estimation. For each task, the proposed methods are evaluated by the performance criteria introduced in the last section and compared with the state-of-the-art methods. Different fault severities and noise levels will be changed to test methods' diagnosis capability.

5.3.1 Continuous-flow stirred tank reactor process

As a benchmark case, the CSTR process is frequently used to evaluate fault diagnosis methodologies [141, 194, 202]. The schematic diagram of the CSTR process is displayed in

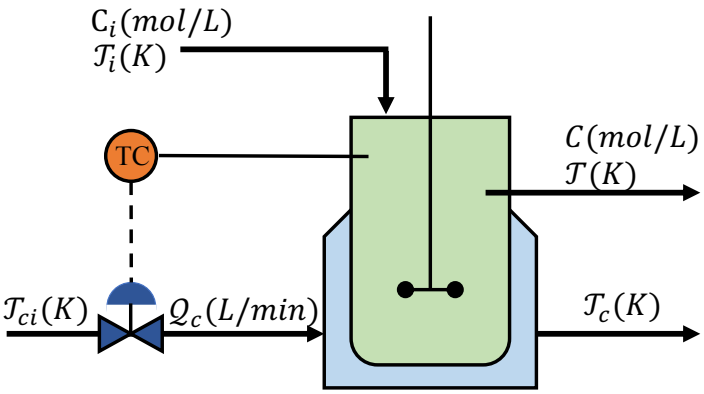

FIGURE 5.1 – Schematic diagram of the CSTR process

Fig.5.1, where an exothermic reaction takes place in a reactor surrounded by a jacketed tank. In the process, a fluid stream fed the reactor and perfectly mixed with catalysts. The reactor temperature is maintained by feeding a coolant medium at a lower temperature. The model can be described as following exothermic first-order reactions,

$$\frac{d\mathcal{C}}{dt} = \frac{Q}{V}(\mathcal{C}_i - \mathcal{C}) - \beta \mathcal{C} q_0 e^{(\frac{-E}{RT})} + \nu_1 \tag{5.12}$$

$$\frac{d\mathcal{T}}{dt} = \frac{Q}{V}(\mathcal{T}_i - \mathcal{T}) - \beta \frac{\Delta H_r q \mathcal{C}}{\varphi \mathcal{C}_p} - h\frac{\mathcal{U}\mathrm{A}}{\varphi \mathcal{C}_p V}(\mathcal{T} - \mathcal{T}_c) + \nu_2 E \tag{5.13}$$

$$\frac{d\mathcal{T}_c}{dt} = \frac{Q_c}{V_c}(\mathcal{T}_{ci} - \mathcal{T}_c) + h\frac{\mathcal{U}\mathrm{A}}{\varphi_c \mathcal{C}_{pc} V_c}(\mathcal{T} - \mathcal{T}_c) + \nu_3 \tag{5.14}$$

where the outputs of the process are the reactor temperature ($\mathcal{T}$), the concentration in the reactor ($\mathcal{C}$), the coolant temperature ($\mathcal{T}_c$), and the coolant flow rate ($\mathcal{Q}_c$). The system's inputs are the input flow concentration ($\mathcal{C}_i$), the flow temperature ($\mathcal{T}_i$), and the coolant flow temperature ($\mathcal{T}_{ci}$). Notations of the above equations are given in Table 5.1.

The corresponding simulation model introduced by Karl Ezra Pilario in [141] is one of the most influential models for its easy access and professional design. The simulation model is shown in Fig.5.2. It can simulate the healthy case and ten different faulty scenarios. All are defined in Table 5.2 : Faults 1 and 2 simulate catalyst decay and heat transfer fouling ; Fault 3 simulates these two faults' simultaneous evolution ; Faults 4 to 10 correspond to additive faults simulating sensor drifts. Note that faults 1, 2, and 3 are multiplicative faults, while all the other faults are additive ones. The subscript F indicates the single faulty variable. The process noises ν_1, ν_2, ν_3 are additive white Gaussian with zero mean and variances $\sigma_1^2 = 0.002$, $\sigma_2^2 = \sigma_3^2 = 2$, respectively, leading to $\mathrm{SNR}_1 = 27\mathrm{dB}$ and $\mathrm{SNR}_2 = \mathrm{SNR}_3 = 47.9\mathrm{dB}$. As examples, we show input signals in Fig.5.3 and output signals in Fig.5.4, respectively.

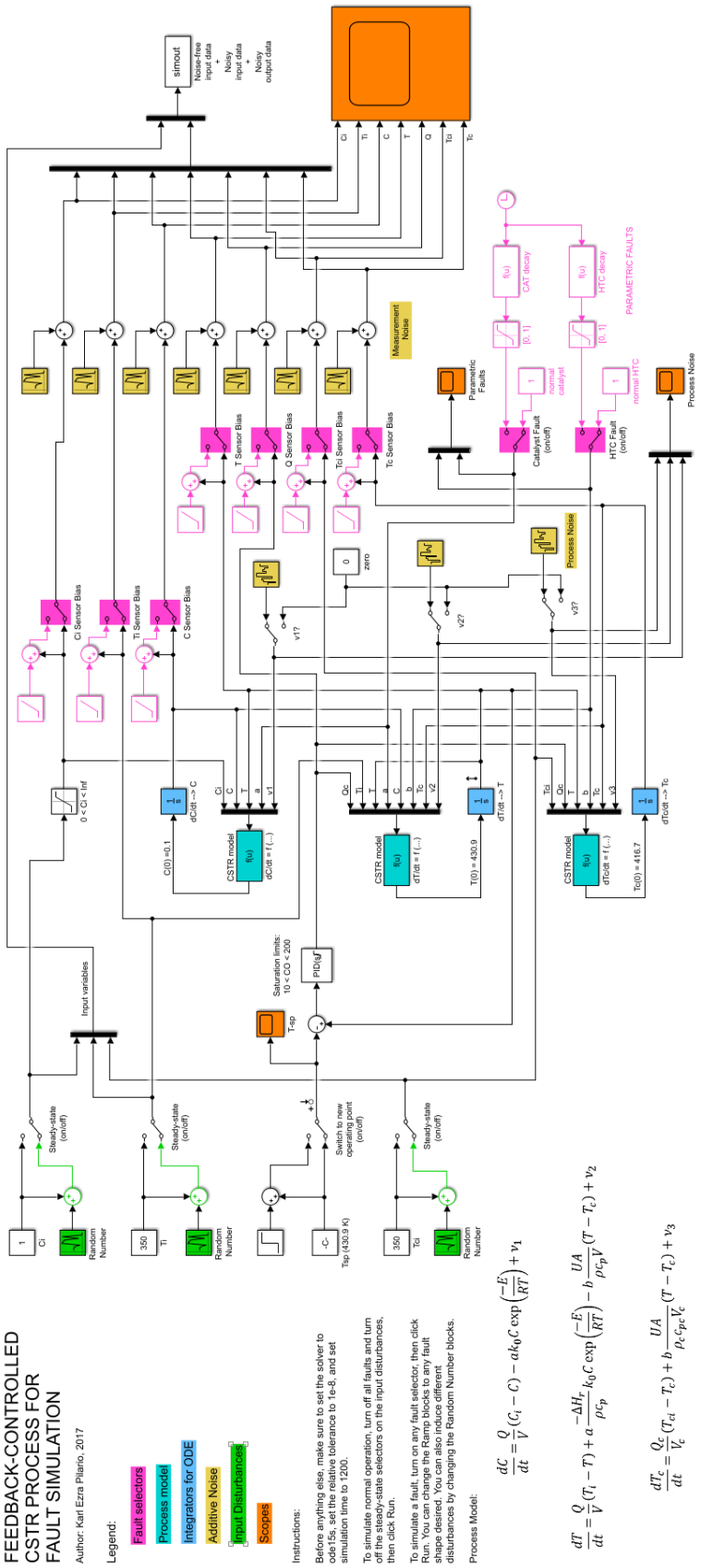

FIGURE 5.2 – Simulation model of the CSTR process

TABLE 5.1 – Notations for CSTR Process

Parameter	Description	Units
$\mathcal{T}$	Reactor temperature	K
$\mathcal{T}_c$	Coolant temperature	K
$\mathcal{T}_i$	Flow temperature	K
$\mathcal{T}_{ci}i$	Coolant flow temperature	K
$\mathcal{C}$	Concentration in the reactor	mol/L
$\mathcal{C}_i$	Input flow concentration	mol/L
$\mathcal{Q}_c$	Coolant flow rate	L/min
$\mathcal{Q}$	Inlet flow rate	L/min
V	Tank volume	L
$\mathcal{U}$A	Heat transfer coefficient	cal/in/K
V_c	Jacket volume	L
q_0	Pre-exponential factor	min^{-1}
ΔH_r	Heat of reaction	cal/mol
φ, φ_c	Fluid density	g/L
E/R	Activation energy	K
$\mathcal{C}_p$,$\mathcal{C}_{pc}$	Fluid heat capacity	acl/g/K
$\nu_{1,2,3}$	Process noise	dB

TABLE 5.2 – Description of the faulty scenarios

Fault ID	Description	Type
F_1	$\beta = \beta_0 e^{-\delta t}$	Multiplicative
F_2	$h = h_0 e^{-\delta t}$	Multiplicative
F_3	Fault 1 and 2	Multiplicative
F_4	$\mathcal{C}_i = C_{i,0} + \delta t$	Additive
F_5	$\mathcal{T}_i = \mathcal{T}_{i,0} + \delta t$	Additive
F_6	$\mathcal{T}_{ci} = \mathcal{T}_{ci,0} + \delta t$	Additive
F_7	$\mathcal{C} = \mathcal{C}_0 + \delta t$	Additive
F_8	$\mathcal{T} = \mathcal{T}_0 + \delta t$	Additive
F_9	$\mathcal{T}_c = \mathcal{T}_{c,0} + \delta t$	Additive
F_{10}	$\mathcal{Q}_c = \mathcal{Q}_{c,0} + \delta t$	Additive

5.3.2 Fault detection performances

5.3.2.1 Detecting results

We structure the process variables as $\boldsymbol{X} = [\mathcal{C}_i, \mathcal{T}_i, \mathcal{T}_{ci}, \mathcal{C}, \mathcal{T}, \mathcal{T}_c, \mathcal{Q}_c]$. For convenience, the above variables are denoted as $\boldsymbol{x}_1$ to $\boldsymbol{x}_7$, such that $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7]$. Our proposal, whose scheme is shown in Fig.4.4, is then applied to the data $\boldsymbol{X}$ for the CSTR case study. For the training procedure, we used one group of fault-free data with 1200 samples (i.e., 1200 minutes) as the fault-free data matrix $\boldsymbol{X}^*$. Based on $\boldsymbol{X}^*$, the optimal region radius was determined as $\gamma_{an}^{opt} = 1.448$ and then 224 anchors were generated. The healthy domain margin was obtained as $r_L = 2.7$ with the given significance level $\alpha = 99.9\%$. To evaluate the online monitoring performance, we generated 250 groups of

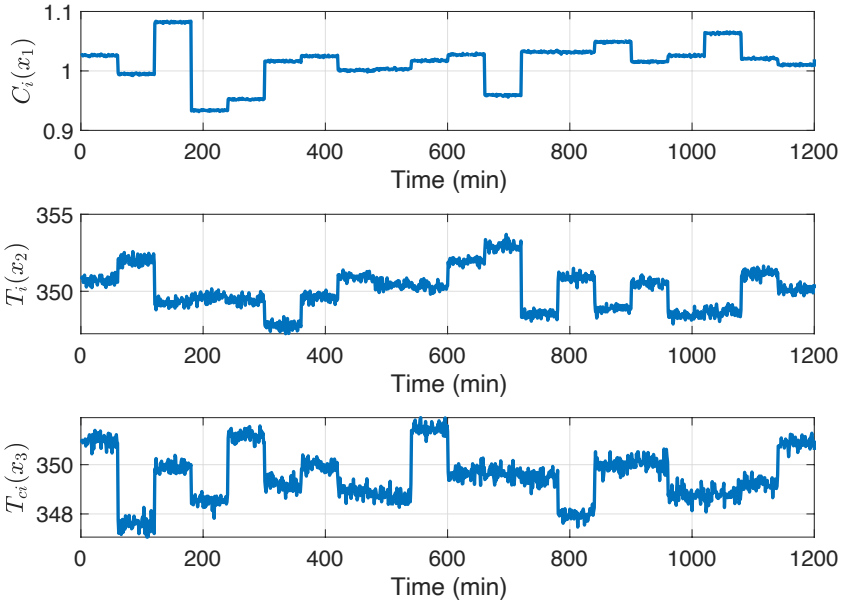

FIGURE 5.3 – Healthy input signals of CSTR process

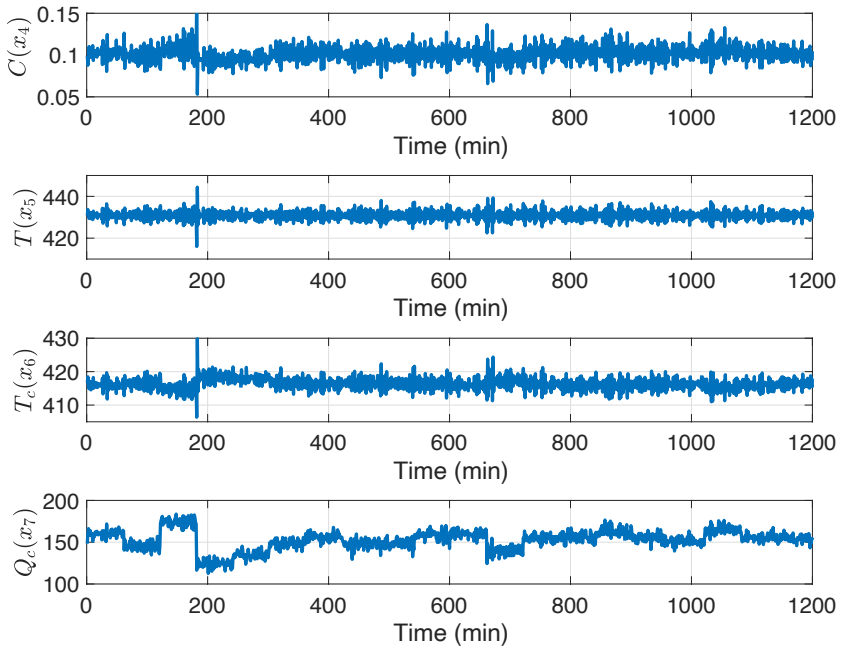

FIGURE 5.4 – Healthy output signals of the CSTR process

data for each faulty scenario, where each group contained 1200 samples, and faults were introduced at $t = 200$ minutes (the 200th sample). For the EPD-CUSUM calculation, the tuning parameters were set as $\omega = 0.99$ and $U_G = 110$.

The obtained detection results for all types of faults are illustrated in Fig.5.5. The left side of these figures are LMD results, and the margin is marked as a horizontal red dashed line. On the right side of the figures are EPD-CUSUM results, and the control limit for the decision-making is marked as a horizontal cyan dashed line. For all these faults, LMD and EPD-CUSUM results increase with time after faults occurrence. Although false alarms exist in the healthy part of the LMD results, the right figures highlight that

EPD-CUSUM results can significantly alleviate the false alarm issue by turning incidental impulse-like detected results into slight fluctuation. Additionally, using the EPD-CUSUM technique allows to summarize those small continuous deviations in LMD results, producing growing results with time. It helps to detect incipient faults accurately and shortens the detection delay. Moreover, due to the cumulative nature of the EPD-CUSUM method, missed detection can be remarkably avoided.

Furthermore, to highlight the benefit of our proposal, EPD-CUSUM results are compared with other fair literature methods developed for incipient fault detection. These methods are reported as effective for non-Gaussian distributed data. The comparative methods are Generalized Canonical Correlation Analysis (GCCA) [27], Canonical Variate Analysis (CVA) [139], Canonical Variate Dissimilarity Analysis (CVDA) method [142], Principal component analysis (PCA) [198], Independent component analysis (ICA) [104,192], Partial least square (PLS) [145,183], One-class Support vector machine (OC-SVM) [93,138,157], k-centers [201], Auto-encoder (AE) [126, 180], isolation forest (IF) [65, 114, 115] . We evaluated the detection delay, false alarm probability, detection probability, and AUC performance for each of them. Due to the large time constant in the CSTR process, the detection delay is presented in hours. The false alarm and detection probabilities are given as ratios.

TABLE 5.3 – Performance of comparative methods for F_3 and F_5

Method	F_3 (Multiplicative)				F_5 (Additive)			
	DD (hour)	P_{FA}	P_d	AUC	DD(hour)	P_{FA}	P_d	AUC
Proposed	0.930	0.0197	**0.9354**	**0.983**	1.300	0.0103	**0.9084**	**0.974**
GCCA-T^2_{r1}	8.330	**0.0004**	0.5397	0.973	3.660	0.0005	0.7823	0.970
GCCA-T^2_{r2}	2.930	0.0009	0.8351	0.979	9.130	0.0008	0.4722	0.939
CVA-T^2	3.390	0.0046	0.7743	0.958	2.820	0.0033	0.8290	0.960
CVA-SPE	2.420	0.0053	0.8504	0.977	4.410	0.0050	0.7174	0.963
CVDA	1.540	0.0075	0.9032	0.983	2.500	0.0057	0.84	0.971
PCA-T^2	5.239	0.0025	0.4916	0.854	7.828	0.0028	0.003	0.499
PCA-SPE	3.931	0.0036	0.5062	0.882	2.746	0.0034	0.7833	0.964
ICA	5.024	0.0017	0.4370	0.848	7.955	0.0013	0.0024	0.546
PLS	7.707	0.0014	0.3999	0.881	4.077	0.0012	0.7278	0.966
OC-SVM	**0.307**	0.4708	0.8305	0.680	0.431	0.4773	0.9071	0.715
k-centers	3.362	0.0242	0.5888	0.782	2.212	0.0243	0.8271	0.901
AE	10.201	0.020	0.0024	0.814	12.930	**0**	0.0082	0.846
IF	0.382	0.2534	0.8421	0.794	**0.347**	0.2734	0.3416	0.534

As an example, we show the performance for a multiplicative fault (F_3) and additive one (F_5), respectively, in Table 5.3. Note that F_3 is the combination of F_1 and F_2, and F_5 is a challenging faulty case in the CSTR case study for detection. The best values of the reported techniques are highlighted in the table in bold font. In terms of detection delay, our proposal efficiently detects a fault for F_3 in less than 0.93 hours of detection delay (no more than 56 additional sample points). This result is not the best but is acceptable. Similarly, the detection delay of the proposed method for F_5 is 1.3 hours, which is also prominent among the reported techniques. GCCA, CVA, PCA, ICA, PLS, k-centers, and AE are not sensitive enough to incipient fault detection as their detection delay is quite large. Although OC-SVM and IF methods seem effective in the early detection of incipient faults, they either suffer from large false alarms probability or low detection probability.

FIGURE 5.5 – LMD and EPD-CUSUM results for 10 faults

TABLE 5.4 – Average detection performance of comparative methods

Method	DD (hours)	P_{FA}	P_d	AUC
Proposed	0.88	0.0128	**0.936**	**0.985**
GCCA-T^2_{r1}	6.34	0.0006	0.6403	0.971
GCCA-T^2_{r2}	3.733	0.0011	0.798	0.975
CVA-T^2	2.904	0.0051	0.819	0.961
CVA-SPE	2.239	0.0057	0.86	0.977
CVDA	1.414	0.0081	0.9118	0.983
PCA-T^2	6.516	0.0026	0.2315	0.675
PCA-SPE	5.114	0.0032	0.4438	0.819
ICA	6.369	0.0017	0.3365	0.767
PLS	7.973	0.0017	0.4539	0.885
OC-SVM	**0.344**	0.484	0.792	0.654
k-centers	4.006	0.0257	0.5303	0.752
AE	12.307	**0.0002**	0.0958	0.762
IF	0.354	0.2562	0.6697	0.707

CVDA shows similar performance as the proposed method but is not as efficient as the one we proposed.

Regarding false alarms, the GCCA technique is the most reliable for F_3 and AE is the best one for F_5. However, these two approaches have low detection probability, meaning they can not detect incipient faults. As for our proposal, the false alarm probability is not perfect but acceptable with values lower than 0.0197. Even with the not perfect false alarm probability, our proposal has the best performance in terms of detection probability for the two faulty cases. Concerning the AUC criterion, our proposal offers the largest value for the two cases, which indicates that the proposed method globally outperforms the other approaches. Moreover, GCCA, CVA, and CVDA also achieve good performance in these two cases since their AUC values are larger than 0.9.

We also take into account the performance of these reported approaches for other faulty cases. The detailed results of these approaches are given in Appendix A.2 (Table A.1 for detection delay, Table A.2 for false alarm probability, Table A.3 for detection probability, and Table A.4 for AUC). In summary, we calculated the average performance of all the faulty cases and illustrated them in Table 5.4. The result indicates that our proposal outperforms the other approaches in terms of global performance. Its detection speed is not the best but also acceptable among the reported approaches. Therefore our proposal can offer the best trade-off performance for the CSTR case study compared to the other techniques.

5.3.2.2 Training efficiency

The evaluation of training efficiency discusses the influence of available training samples on the method's performance. In the procedure of healthy domain approximation, two aspects are potentially affected by the number of training samples : anchor generation and domain margin selection. In the anchor generation part, the anchor-generation algorithm is proposed to reduce the redundancy information and extract crucial spatial information

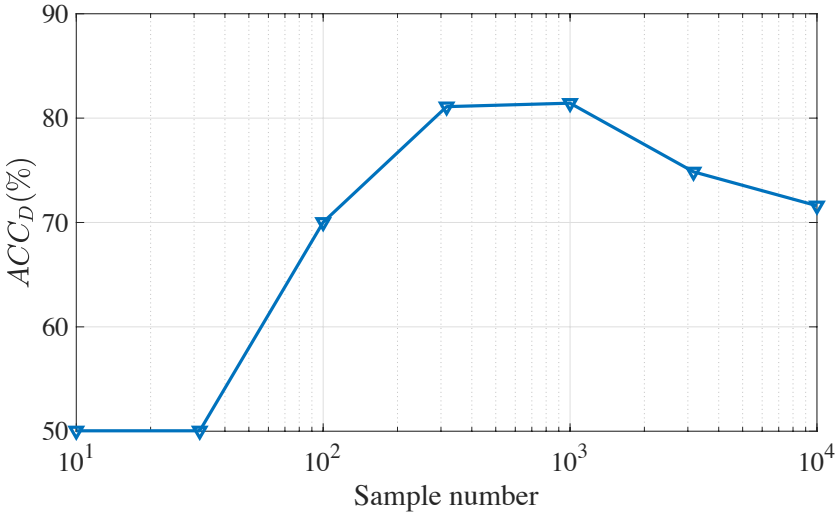

FIGURE 5.6 – The accuracy results for different numbers of training samples

of the healthy domain. Therefore, training samples are expected to include as much useful information as possible instead of redundancy and noise. While in the domain margin selection, sufficient training samples are necessary to estimate the parameters of the density model. Based on these two considerations, the relation between the training samples' number and the method's performance is then investigated.

Fig.5.6 exhibits the evolution of accuracy performance along with the change of sample number. When training samples are insufficient, such as the sample's number is less than 50, the accuracy performance is poor, lower than 50%. Consequently, when the sample number increases to 100, the result notably increases and almost reaches 70% accuracy. The accuracy further increases when 1000 samples are used for training. However, with the training samples number continually growing from 1000 to 10000, one can notice the sight performance degeneration, which may result from the introduction of a large number of irrelevant samples. As a result, sufficient training samples (> 500) are required, but excess samples (> 1000) will degrade the performance.

5.3.2.3 Parameter tuning

This part addresses the effect of parameters U_g and ω on the detection performance of the proposed approach. As introduced in section 4, the control limit U_g is selected as the maximum D_e value of the healthy samples to eliminate the false alarm. It means that the control limit U_g is settled when the parameter ω is selected, and we can investigate the effect of the parameters by only changing the parameter ω.

Fig.5.7 demonstrates the four performance criteria (AUC, detection delay, false alarm probability, and detection probability) of the proposed method with different ω values and FNR conditions. The results show that when FNR is large, e.g., FNR$>$ 15dB, selecting a large ω value is more likely to have high AUC performance, low false alarm probability, short detection delay, and high detection probability. On the contrary, with noise increase or fault decrease, e.g., FNR drops to 15dB, small ω values can achieve better AUC and detection probability performance. In other words, a larger ω value should be considered for smaller fault severity or stronger noise power. However, if the FNR condition is unknown,

which is more common in practice, a middle value of ω is a good choice to achieve excellent global performance.

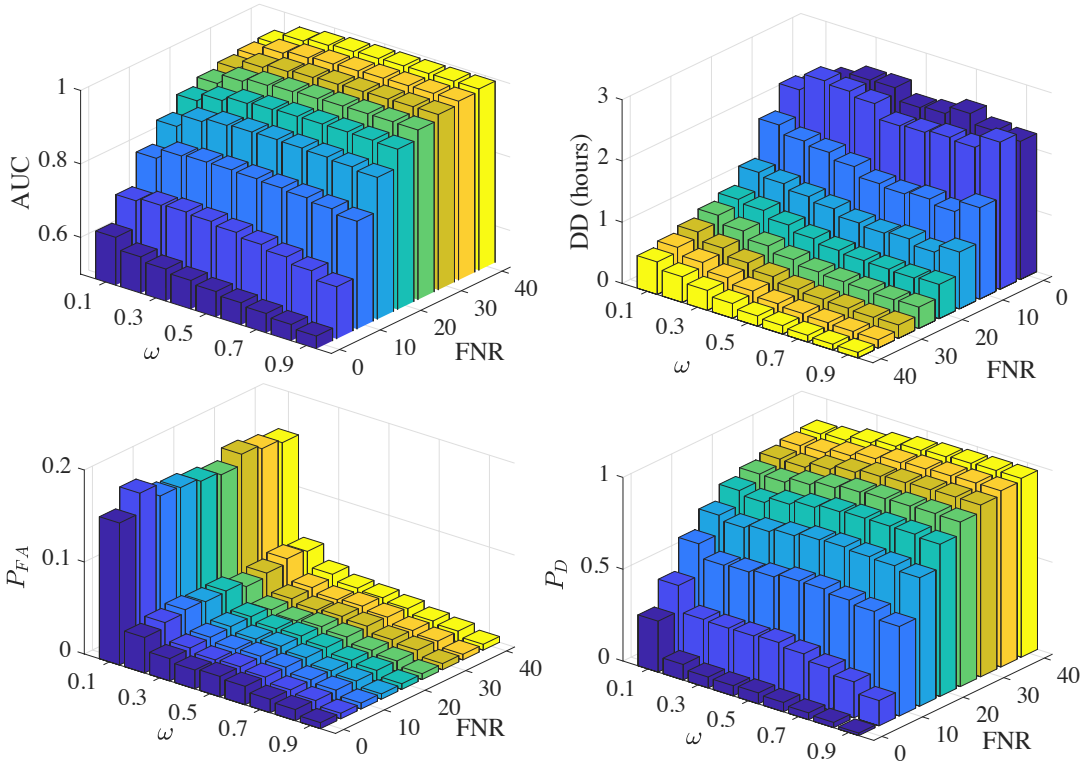

FIGURE 5.7 – Effect of parameter ω on the performance of the proposed method along with different FNR conditions

5.3.2.4 Detection capability

This performance analysis focuses on the detection capability for different fault severity. Therefore, we considered varying faults severity of F_5 by changing FNR from -10dB to 20dB and kept all the other operating conditions the same as previously defined. In Fig.5.8, ROC curves of the EPD-CUSUM index for different FNR settings are displayed to show the fault detection capability. It can be noticed that once FNR decreases from 20dB to -10dB (fault severity decreases), the corresponding AUC value decreases and reaches a low value close to 0.5, leading to the worst detection performance. When FNR is lower than 5dB, meaning that the fault power is much lower than the noise power, the fault detection capability of the proposed diagnosis index is weak.

When we focus on the detection probability performance with a low P_{FA} value, corresponding to a practical condition for fault detection, the improvement from our proposal is the most significant. To further show the benefit of our proposal, we compare the detection probability of the detection approaches whose average AUC performance is larger than 0.9 in Table 5.4. Our proposal, GCCA with two indexes, CVA with two indexes, and CVDA are then compared. We consider the case of $P_{FA} = 0.01$ and show their detection probability results in Fig.5.9. The results highlight that the proposed method notably outperforms the other approaches, particularly for incipient faults detection (like FNR < 5dB). Our proposed methodology offers a high detection sensitivity and efficiency for incipient faults.

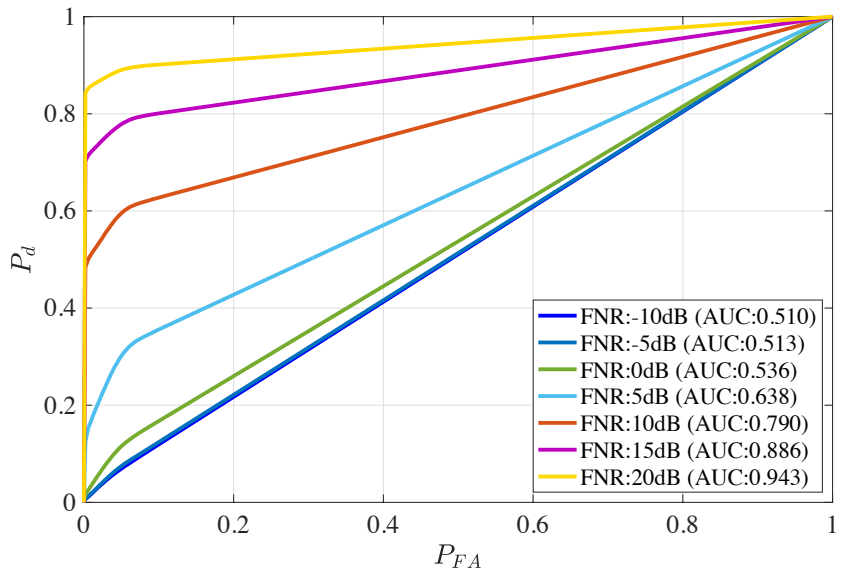

FIGURE 5.8 – ROC curves of EPD-CUSUM results for different FNR settings

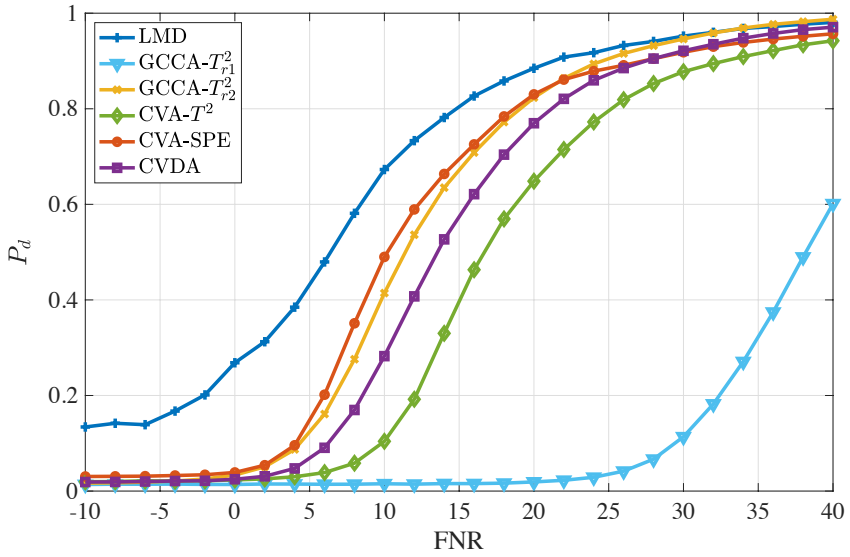

FIGURE 5.9 – Detection probability performance for different detection approaches, where $P_{FA} = 0.01$

5.3.2.5 Detection robustness

Subsequently, different noise and fault severity levels are taken into account in the simulation to evaluate the performance of our proposal deeply. The goal is then to qualify the robustness of the proposed fault detection methodology regarding noise. Note that Gaussian white noise is added to both the training and testing data. With SNR varying from 0dB to 50dB and SFR varying from −40dB to 40dB, the resulting AUC values are plotted in Fig.5.10. The results highlight that both factors affect the performance of the proposed detection method : either the increase of noise power or the decrease of fault severity leads to low detection performance. However, for the common range of noise levels, such as from 15dB to 40dB SNR, when the SFR correspondingly decreases from 0dB to −40dB, the detection performance of the proposed framework is effective since AUC values

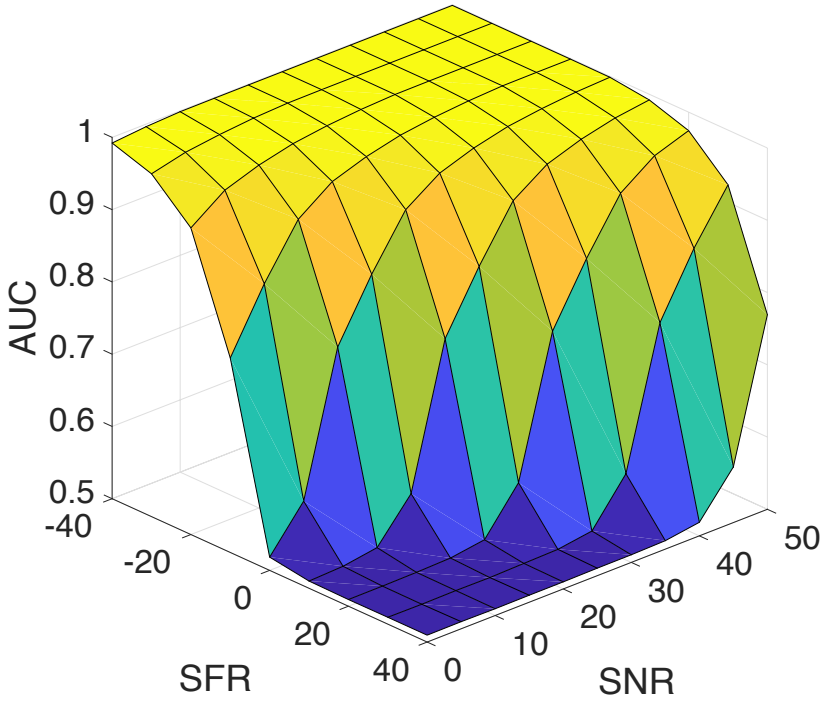

FIGURE 5.10 – AUC performance for different noise strength and fault severity

in this area are larger than 0.9. In that range, the proposed methodology remains robust enough.

5.3.2.6 Detection time occurrence efficiency

The detection time of the fault occurrence is important to qualify the efficiency of the proposal properly. Therefore, this part considers the detection time occurrence efficiency for varying fault severity. We set the noise with SNR = 20dB and changed the fault severity (FNR). In a low false alarm condition ($P_{FA} = 0.01$), this study compares four approaches : GCCA, CVA, CVDA, and our proposal, as in subsection 5.3.2.4, and evaluates their detection delay performance. As demonstrated in Fig.5.11, the result illustrates that our proposal is the most efficient one for its shortest detection delay. The progress of our approach in terms of detection delay is significant, especially for incipient faults (FNR< 5dB). The detection delay of the proposed method is almost one hour smaller than other methods when FNR< 0dB. It indicates that our approach is the only one that can detect faults in low FNR conditions. With FNR increase (fault severity increase), all these approaches have shorter detection delay times.

Then, we changed the noise power and fault severity to evaluate the detection delay efficiency of the proposed method. As in subsection 5.3.2.4, SNR varies from 0dB to 50dB and SFR vary from −40dB to 20dB. The detection delay is evaluated for each condition, and the result is illustrated in Fig.5.12. Note that the detection delay is evaluated in hours since the signals are collected each minute and incipient faults change slowly. We can notice that the detection delay remains at a low value when SNR is a very high positive value (*i.e.,* very low noise level) and SFR is a negative very low value (*i.e.,* very high fault severity). But if the noise dramatically increases (SNR close to zero) and the fault severity

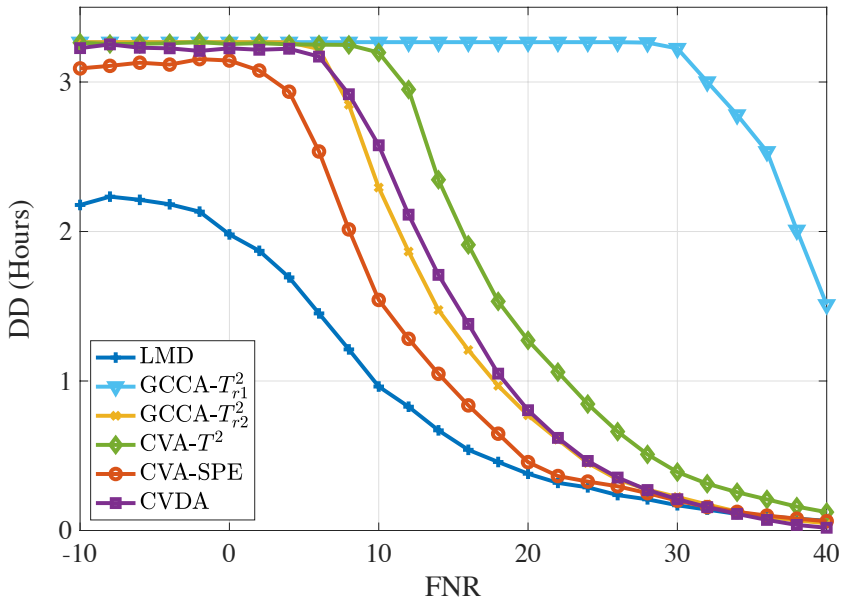

FIGURE 5.11 – Detection delay performance for different fault severity

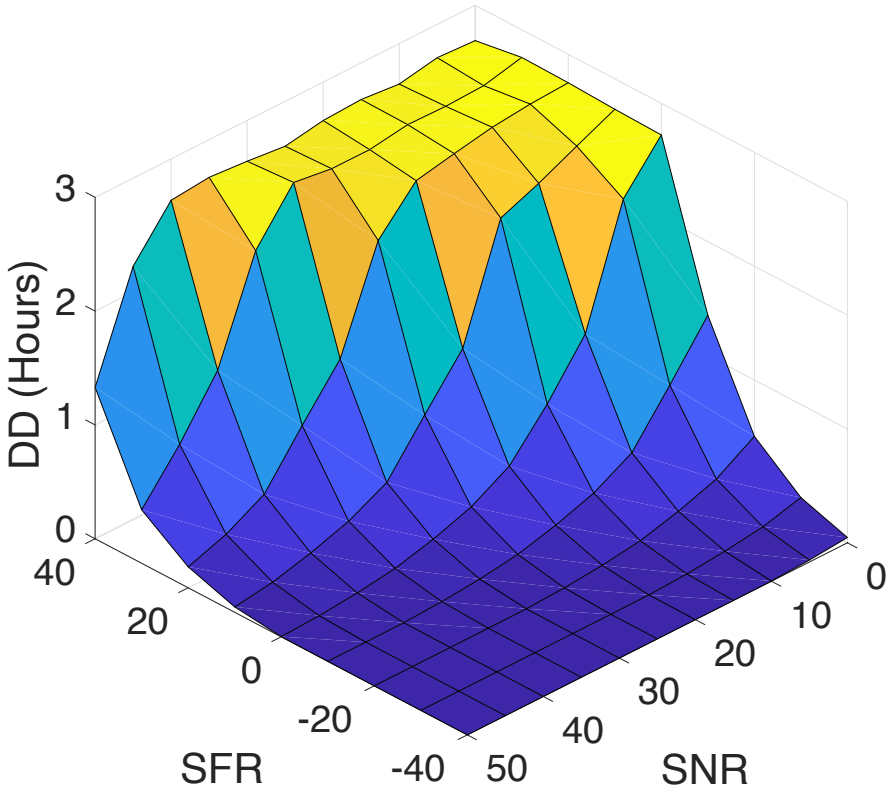

FIGURE 5.12 – Detection delay performance for different noise strengths and fault severities

deeply decreases (SFR positive), then the detection delay will exponentially increase and its value will be very high. Indeed, the detection occurrence is efficient with our proposal. The noise influence has been widely reduced by using our proposal, but it is still remaining and cannot be completely removed. This reduction seems sufficient enough to consider this proposal for an engineering application study.

5.3.3 Fault isolation performance

5.3.3.1 Isolation efficiency of the LMD-based contribution plot method

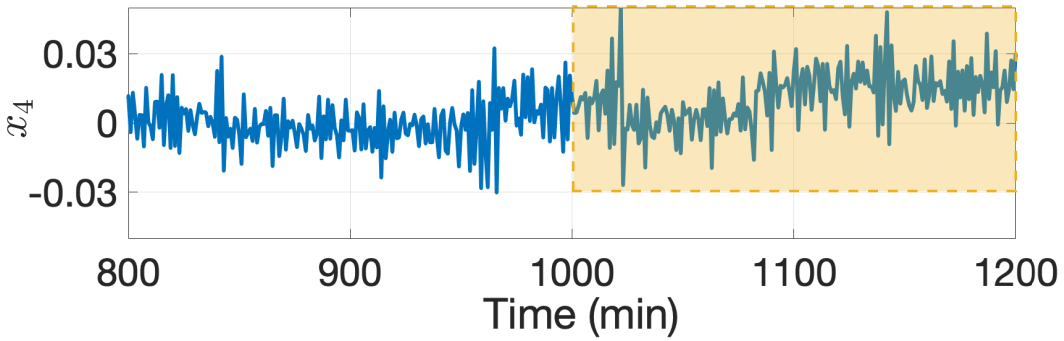

FIGURE 5.13 – Faulty signal of 4th sensor with 10dB FNR.

In order to properly evaluate the isolation efficiency of the LMD-based contribution plot method (presented in section 4.4), this subsection only considers the CSTR data with addictive faults (F_4 to F_{10}) to meet the addictive assumption of this approach. Therefore, we generated 1 healthy and 7 faulty cases with SNR=30dB and FNR=20dB., i.e., F_4 to F_{10}. As introduced in Table. 5.2, each faulty case contains a single faulty variable, e.g., F_4 indicates a fault occurring at x_1, and so on. Fig.5.13 shows an example of a faulty signal at variable x_4 (F_7) with a fault introduced from the 1000th sample. As marked by the dotted yellow box, the elevated tendency of the faulty signal is slight and almost invisible at its early stage. The fault contribution of each variable for different cases was calculated and shown in Fig. 5.14. The results are divided into 8 groups, the healthy one and 7 faulty ones named F_4 to F_{10}. For each group, the first bar stands for the fault contribution of the first variable, and so on. In the healthy group, all contribution values are close and lower than 0.15, while the contribution of the true faulty variable is always the largest in each faulty group. For example, in group F_4, whose true faulty variable is the first one, the fault contribution of the first variable has the largest value of 0.22. Therefore, the faulty variable x_1 can be correctly determined by applying the proposed fault isolation method.

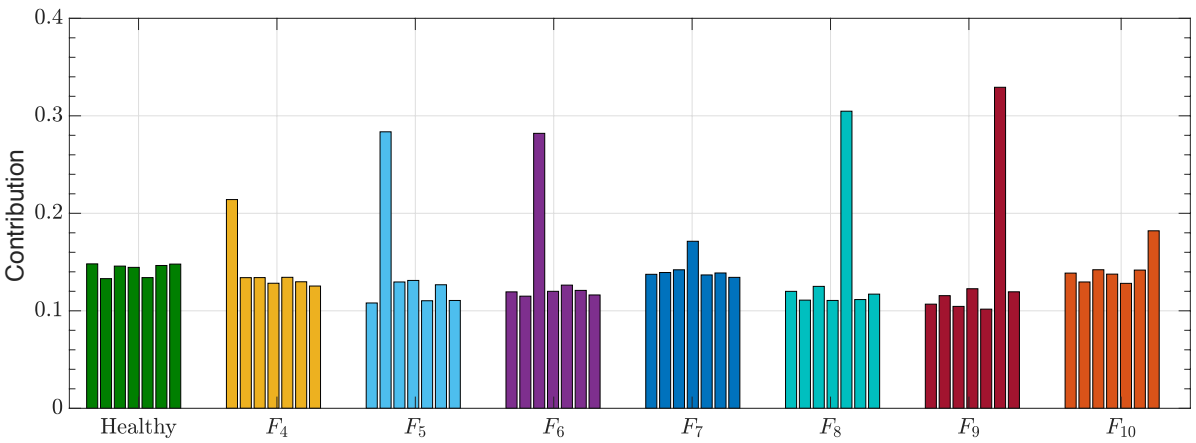

FIGURE 5.14 – Fault contribution result of each variable for 1 healthy and 7 faulty cases

To evaluate the method's reliability and discrimination ability among different faulty variables, we repeated the fault isolation experiment 1000 times for each faulty case and calculated the confusion matrix, which is shown in Fig.5.15. The result highlights that our isolation method achieves 100% accuracy for all faulty cases under the 20dB FNR condition. It means that if the noise power is not so close to the fault severity, we can accurately perform the fault isolation with excellent reliability.

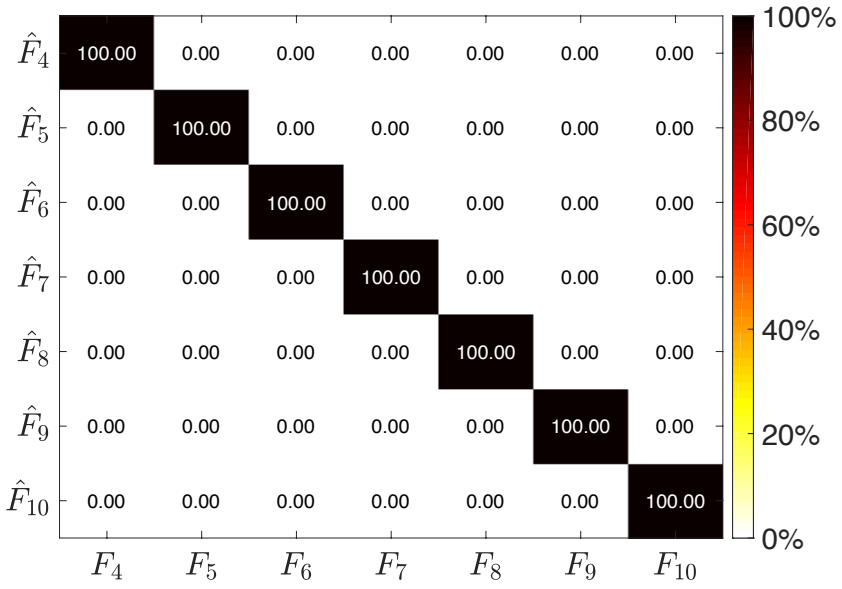

FIGURE 5.15 – Confusion matrix of the proposed method for FNR=20dB

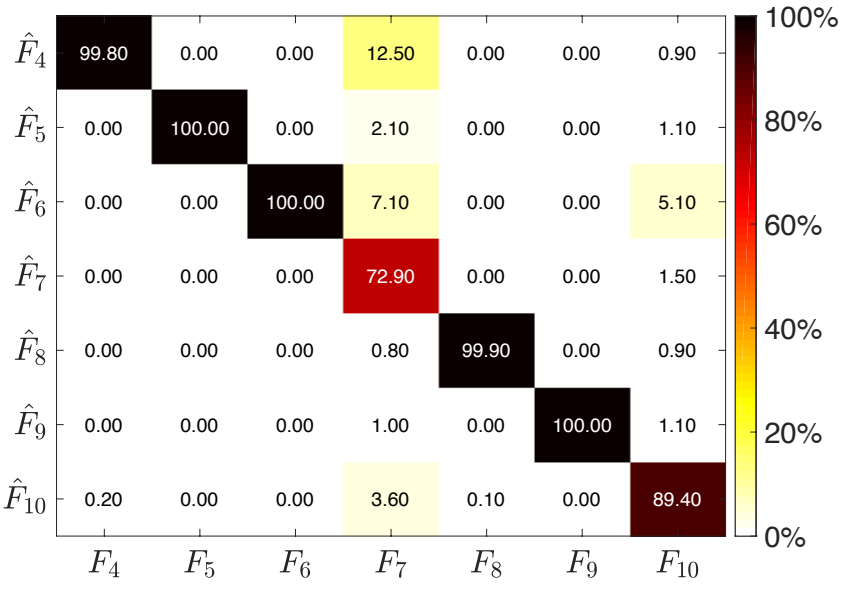

FIGURE 5.16 – Confusion matrix of the proposed method for FNR=0dB

Despite the above promising isolation result, observing the contribution values of all faulty cases, we notice that the results of F_7 and F_{10} seem less significant than others, which implies that these two cases are most challenging for the isolation procedure. Therefore, to study the impact of fault severity on the isolation procedure, the accuracy performance was further evaluated for FNR=0dB. Fig.5.16 exhibits the confusion matrix under this

condition, showing that the accuracy of F_7 and F_{10} cases significantly decreases. Case F_7 is confused with F_4 in 12.5% of realizations and confused with F_6 in 7.1% of realizations. Although the accuracy for F_{10} is slightly better than for F_7, there are 5.1% of confusing results with F_6. Besides, the accuracy for F_4 and F_9 also slightly decrease (at least 99.8% of accuracy), but they are still at a very high level. Our isolation procedure can then be considered accurate enough to identify faulty variable correctly.

5.3.3.2 Single fault isolation efficiency of the reconstruction based contribution using LMD

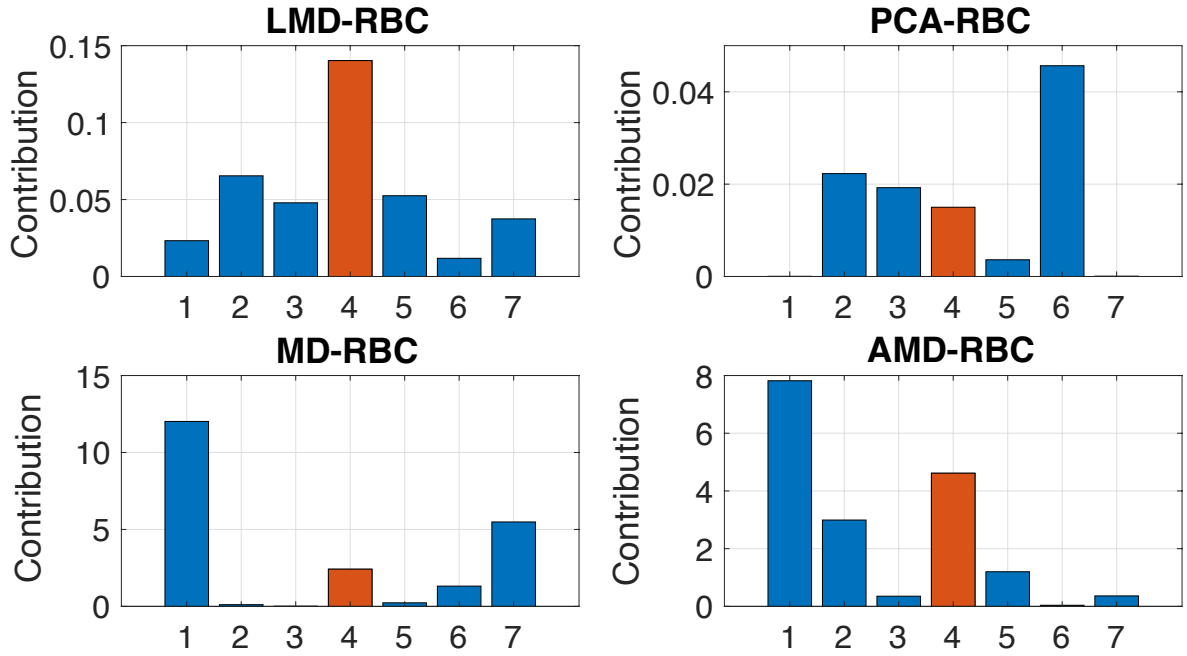

FIGURE 5.17 – Reconstruction-based contribution of the four RBC approaches for $\boldsymbol{x}_4$ faulty case in 10dB FNR condition.

This subsection investigates the isolation performance of the proposed LMD-based RBC method (see Fig.4.18) for single fault situations. The fault number is 1, and there are 7 candidates in the fault direction set. Let us consider an incipient fault with 10dB FNR occurring at the 4th sensor ($\boldsymbol{x}_4$). To highlight the advantage of using LMD in the RBC framework, we compare it with other traditional RBC approaches based on the combined index of PCA (PCA-RBC) [3], conventional Mahalanobis distance (MD-RBC) [85], and augmented Mahalanobis distance (AMD-RBC) [85]. As an example, Fig.5.17 exhibits the reconstruction-based contribution results of the four methods for the 1100th sample. In this example, the LMD-RBC method yields the largest contribution result corresponding to the variable $\boldsymbol{x}_4$. However, the largest contribution result of PCA-RBC, MD-RBC, and AMD-RBC approaches correspond to variables $\boldsymbol{x}_6$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_1$, respectively. According to the isolation strategy of the RBC methods, the fault variable is identified as the one having the largest reconstruction contribution value. Therefore, only the LMD-RBC method successfully isolates the faulty variable.

Then, we investigate the discrimination ability of the LMD-RBC approach among different faulty variables. The evaluated confusion matrix under the 20dB FNR condition

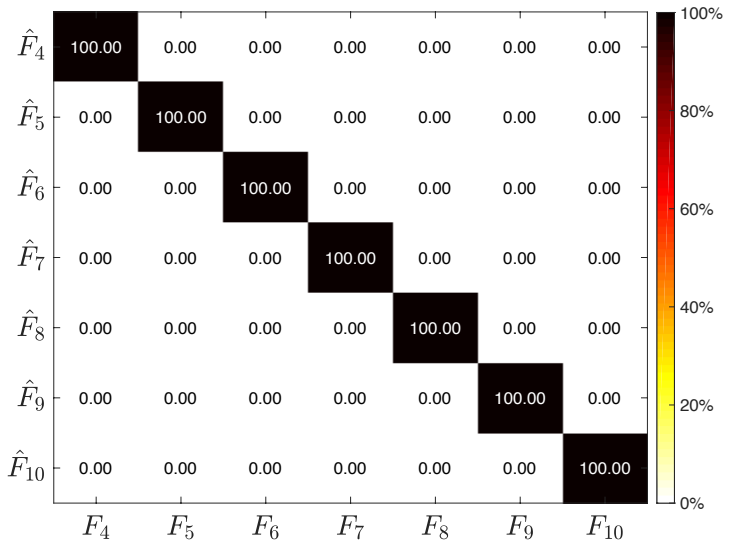

FIGURE 5.18 – Confusion matrix of the LMD-RBC method for FNR=20dB

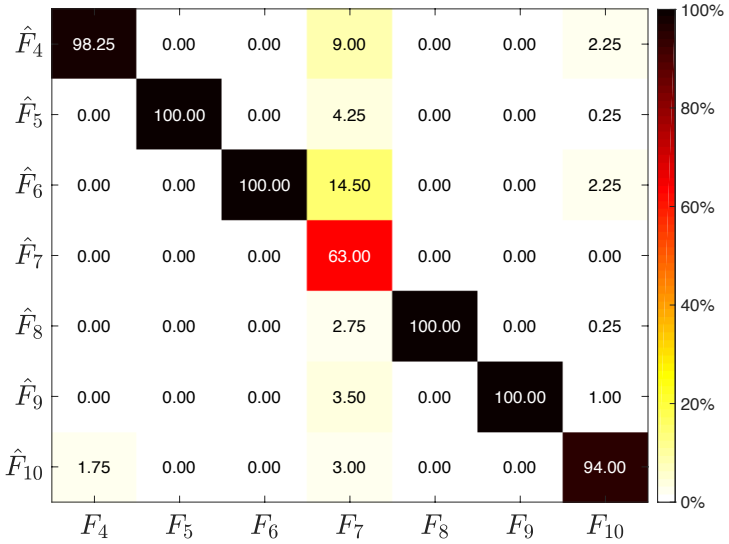

FIGURE 5.19 – Confusion matrix of the LMD-RBC method for FNR=0dB

is shown in Fig.5.18. It indicates that the LMD-RBC approach can distinguish all faulty cases in 20dB FNR. We further decreased the fault severity to 0db FNR and evaluated the confusion matrix of this approach. As illustrated in Fig.5.19, the accuracy for case F_7 significantly decreases to 63%. This approach confuses case F_7 with F_6 in 14.5% of realizations and confuses it with F_4 in 9% of realizations. Besides, the accuracy for the cases F_4 and F_{10} also decrease slightly but maintains a high level (larger than 94%). Although fault isolation of low fault severity is still a great challenge for the LMD-RBC approach, it is generally effective in identifying single fault variables.

The previous performance evaluation shows the LMD-based contribution plot and LMD-RBC methods are effective for single fault isolation. However, the comprehensive evaluation of fault isolation approaches concerning varying fault severity is necessary to reveal methods' isolation efficiency. We compare different fault isolation approaches to highlight the advantage of our proposal using LMD in the fault isolation task. More

specifically, we consider two divergence-based methods (KLD and JSD), three RBC methods (MD-RBC, AMD-RBC, and PCA-RBC), and our two proposals : LMD-based contribution plot (LMD-Con), and RBC method based on LMD (LMD-RBC). In this test, a fault occurred in an arbitrary single variable with uniform probability, i.e., 1/7 for each faulty case. For different FNR conditions, the test was repeated 500 times to obtain stable results.

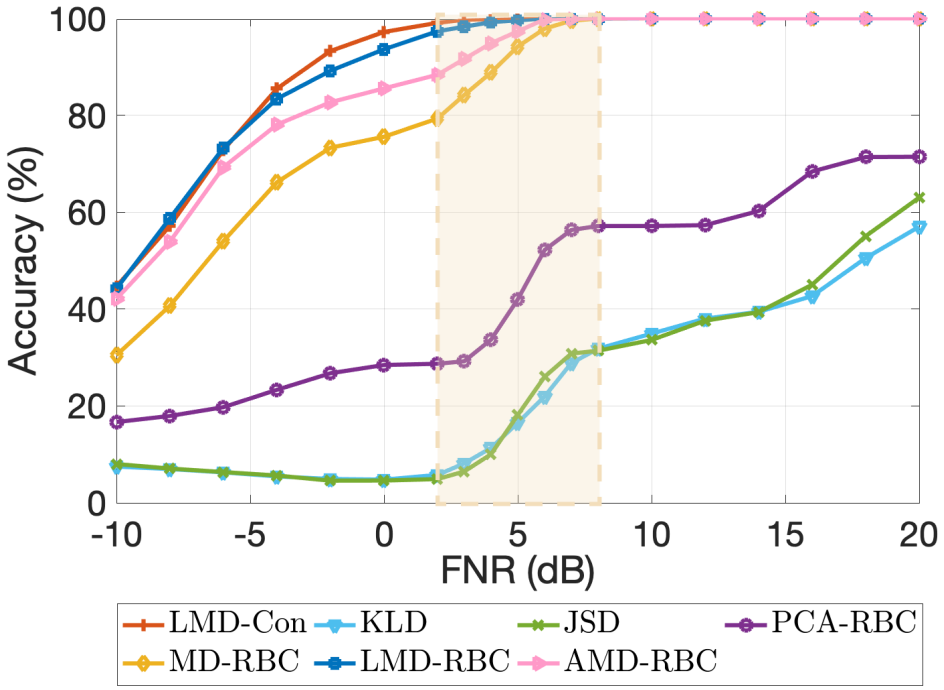

FIGURE 5.20 – Total isolation accuracy performance of different methods along with varying FNR values

Fig.5.20 displays the total accuracy of 7 methods for the 7 considered faulty cases along with varying FNR values. The result shows that the two LMD-based approaches have similar performance, and they outperform others. For FNR $>$ 3dB, the LMD-Con method can achieve 100% accuracy, and the accuracy of LMD-RBC approach is round 99%. Although the accuracy of the two LMD-based methods starts to decrease when FNR$<$ 3dB, their high accuracy performance is satisfying for FNR larger than $-$5dB. AMD-RBC approach also has prominent performance close to that of the two LMD-based approaches. MD-RBC method seems not sensitive enough to low fault severity, resulting in slightly worse performance for FNR$<$ 5dB. It can be noticed that the AMD-RBC and MD-RBC methods have the same high accuracy as the LMD-based methods when FNR$>$ 8dB, while their performance significantly decline if tiny fault severities are considered like for FNR$<$ 5dB. Although both LMD and AMD techniques originate from the Mahalanobis distance, LMD's superior performance indicates that the local change of the signal features seems more helpful for analyzing the characteristic of incipient faults. According to these comparison results, the PCA-RBC, JSD, and KLD approaches are all ineffective for fault isolation of the CSTR process. The better performance of the MD-RBC method than that of the PCA-RBC method implies that the space partition of PCA leads to the loss of faulty information and, therefore, worse performance for fault isolation. The poor performance of divergence-based approaches may result from the inaccuracy estimation of samples' probability density.

The results of Fig.5.20 also indicate that 2dB to 8dB FNR is a critical range of fault amplitude change. Below 2dB, the fault's symptom is too slight and thus is tricky to diagnose, while above 8dB, a fault is recognizable. In this critical range, the accuracy performances of all the considered methods are all degraded with FNR decreasing, where the PCA-RBC, KLD, and JSD methods have the most conspicuous performance degeneration (about 30%). Although the two LMD-based approaches, AMD-RBC, and MD-RBC methods have good performance in large FNR conditions, the accuracy of the AMD-RBC and MD-RBC methods is reduced by 10% and 20% in this critical range, respectively, while the LMD methods are almost unchanged. The significant performance difference in the critical range shows that the two proposed LMD-based methods are more sensitive to incipient faults than other approaches.

To appreciate the limitation of the mentioned methods, we display the obtained accuracy results of each faulty case for FNR=0dB in Fig.5.21. Consistent with the result of Figure 5.16, F_7 and F_10 are the two most difficult cases to isolate for all methods. In other words, the main limitation of these methods lies in the relatively weak isolation ability for F_7 and F_10 cases. For example, the worse total performance of the AMD-RBC and MD-RBC methods compared to LMD-based methods is mainly caused by their low accuracy performance for the two challenging cases. Conversely, although the total performance of the PCA-RBC method is poor, its accuracy for F_5 and F_6 cases remains promising.

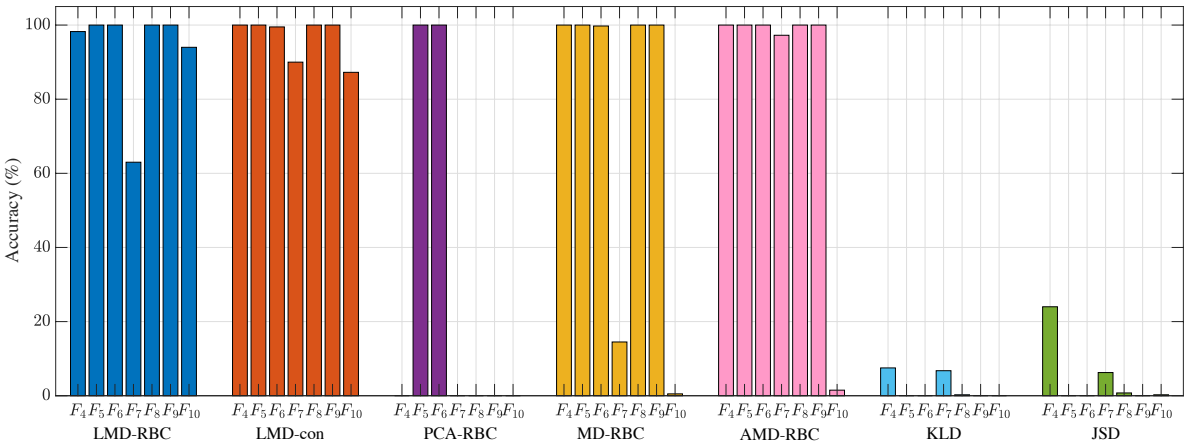

FIGURE 5.21 – Accuracy of different methods for each faulty case with FNR=0dB

5.3.3.3 Multiple fault isolation efficiency of the reconstruction based contribution using LMD

Multiple faults are common but challenging for diagnosis, which needs careful investigation. To validate the effectiveness of the LMD-RBC method for multiple faults, we consider two faulty variables, $\boldsymbol{x}_4$ and $\boldsymbol{x}_7$, occurring in the CSTR system. In this case, the total number of candidate fault directions is 21 (the combinations of any two variables from seven). After applying the proposed and comparative methods, their reconstruction contribution values for the 1100th sample were obtained and illustrated in Fig.5.22. In the figure, the elements of a tuple indicate that the two faulty variables, e.g., $(4, 7)$ means the 4th and 7th variables are faulty. For the LMD-RBC method, the largest contribution value is located at the situation $(4, 7)$, indicating that this method correctly identifies the

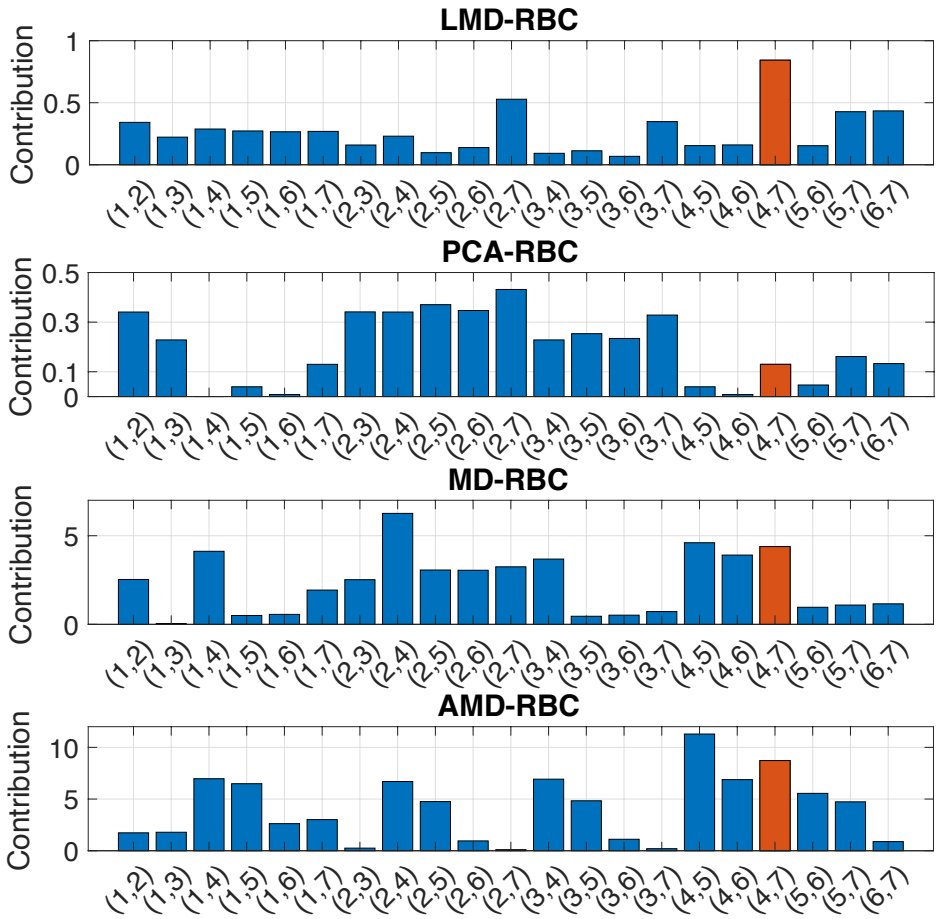

FIGURE 5.22 – Reconstruction based contribution of the four RBC approaches for $\boldsymbol{x}_4$ and $\boldsymbol{x}_7$ faulty case in 10dB FNR condition.

two faulty variables $\boldsymbol{x}_4$ and $\boldsymbol{x}_7$. Furthermore, the largest contribution value is distinct compared to other contribution values, meaning the risk of false isolating is low. As for other diagnosis methods, none of them can identify the true faulty variables in this case, where the contribution values of the PCA-RBC method are nearly random values and helpless for fault isolation.

Subsequently, we comprehensively evaluate the performance by considering multiple faulty variables and different FNR conditions. Since the LMD base contribution plot method, KLD, and JSD methods are not available for multiple faulty cases, we only consider four RBC-based approaches : LMD-RBC, MD-RBC, AMD-RBC, and PCA-RBC. The experiment was repeated 500 times to obtain a reliable result. As shown in Fig.5.23,the LMD-RBC method outperforms the other approaches for low fault severity, and its accuracy result is about 20% higher than the second good one for -10dB FNR. Similar to single fault cases, AMD-RBC and MD-RBC methods are worse than the LMD-RBC method in low fault severity conditions, but they are all accurate if a fault is large (FNR$>$ 5dB). The PCA-RBC is still unsatisfying for multiple faults isolation. The comparison between Fig.5.20 and Fig.5.23 shows that the accuracy performance for multiple faults is a little worse than that for single fault, which indicates that the multiple faults cases are more challenging. Obviously, the difficulty of the isolation task increases with the number of faulty variables, and the accuracy performance dramatically degenerates in this case because of the expanding candidate number of fault directions. However, the LMD-RBC

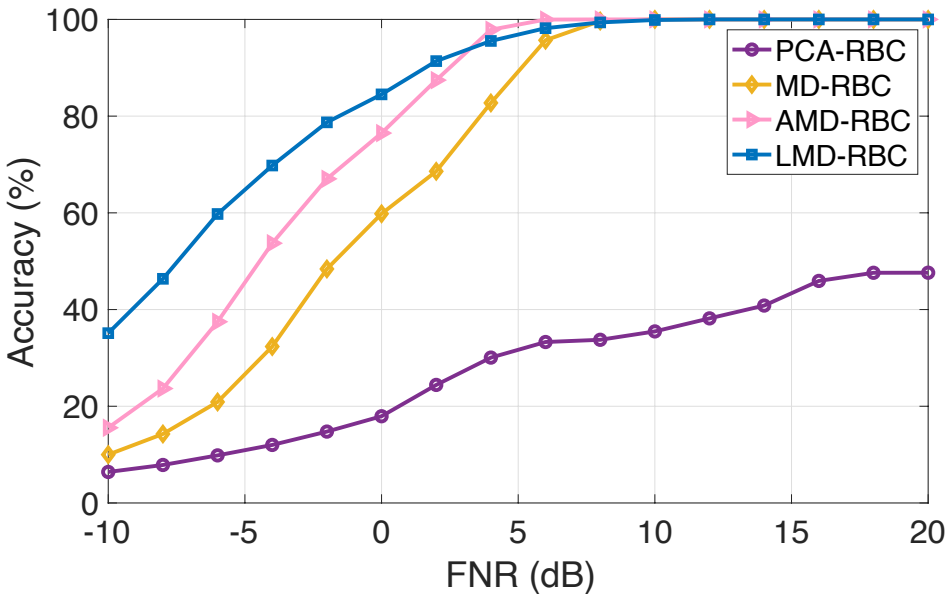

FIGURE 5.23 – Average isolation accuracy performance of the four RBC approaches for two sensors faults

approach has higher accuracy for single and multiple faults cases than the others. There is only a 5% accuracy decrease of the LMD-RBC approach for multiple faults cases compared to the single fault cases. Therefore, the proposed LMD-RBC approach is more powerful and more accurate for faulty variable isolation tasks.

5.3.4 Fault severity estimation performances

5.3.4.1 Fault's increasing rate estimation efficiency

After the isolation of the faulty variable, the fault severity estimation procedure can start based on the fault detection and isolation results obtained from the previous steps. As introduced in the last section, there are two factors that can be considered to estimate fault severity : fault's increasing rate and fault amplitude. This subsection first investigates the developed method that aims to estimate the fault's increasing rate (presented in section 4.5). We consider a fault introduced in the 4th variable with SNR=30dB (given noise) and a constant increasing rate $\delta = 0.15$. The fault occurs at 1000 min (1000th sample). Since the estimation performance depends on the isolation accuracy, to individually discuss the performance of estimation procedures, we use the true faulty variable for fault severity estimation. This operation is not necessary for practical applications but can reveal the ideal estimation performance of the proposed methods and offer a uniform condition for comparison.

Fig.5.24 shows the LMD results of observation samples. It can be noticed that LMD results grow with time after the fault occurrence. The increasing trend of the LMD results not only indicates the occurrence of a fault but also contains its evolution information.Based on the fault growing tendency, we apply our method to faulty samples to estimate the value of increasing rate δ. The fault severity can then be assessed according to the value of δ.

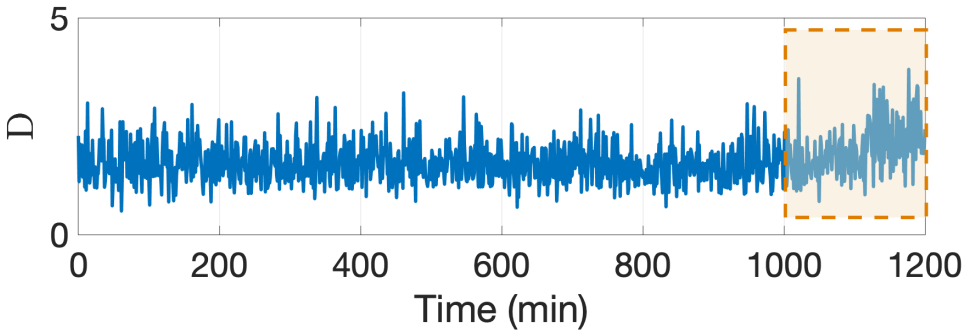

FIGURE 5.24 – Example of LMD result with SNR=30dB and FNR=20dB. The fault occurs at 1000 min and affects the 4th variable

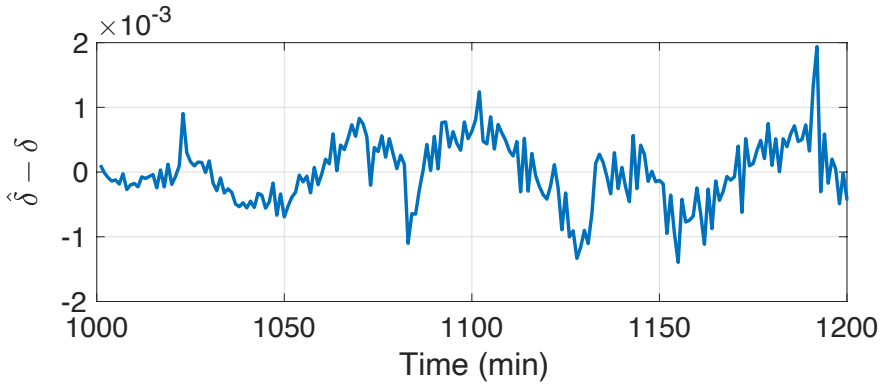

FIGURE 5.25 – Estimation error of the true increasing rate and the estimated value along with time when the increasing rate is constant

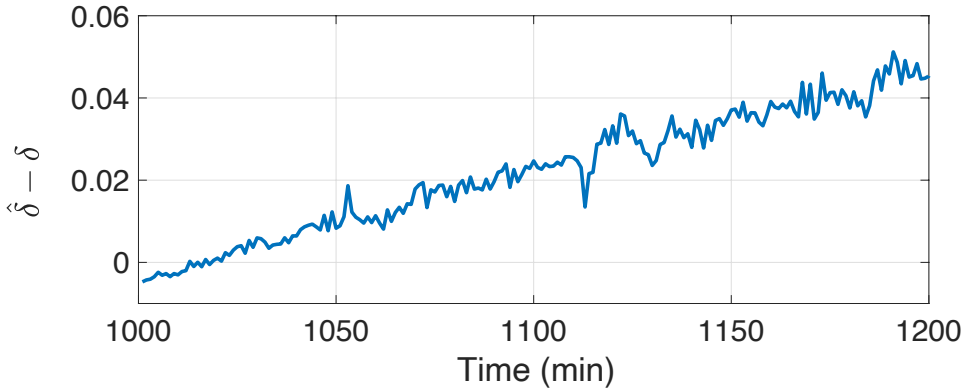

FIGURE 5.26 – Estimation error of the true increasing rate and the estimated value along with time when the increasing rate is not constant

Fig.5.25 exhibits the error of the true increasing rate and the estimated value along with time. The result shows that the estimated value is close to the true value with a small fluctuation. In this case, the proposed method based on LMD can accurately estimate the increasing rate. The estimated error mainly caused by noise is small. When the fault assumption is satisfied, i.e., the fault's increasing rate is a constant, the proposed method estimates the fault severity accurately. Then, we consider an increasing rate that grows with time, such that $\delta = 0.05 + 0.001t$. The estimated error of the proposed method for this test is illustrated in Fig.5.26. It shows that when the assumption is not satisfied, the

estimated error is significantly large and grows with time. Indeed, the increasing rate grows faster and faster over time, which notably deviates from our assumption. However, at the first beginning of the fault, such as from 0 to 30 minutes, the error is smaller than 0.005, and the estimated result is still accurate enough. When the true situation of increasing rate deviates significantly from the assumption, the estimated value is always larger than the true value, which is called over-estimation. For the industrial scenario, slightly over-estimated results are more meaningful than under-estimated results (estimated value smaller than the true one) since slightly exaggerated estimated results allow operators to avoid accidents in advance.

Next, we investigate the estimation performance of the proposed method for different increasing rate. In this test, the increasing rate is constant but was given different values. Fig.5.27 demonstrates the relative error of the proposed method for varying δ. Basically, as δ value increases from 4×10^{-4}, the relative error gradually decreases and stabilizes at a small value close to 1%. It means that the error is approximately equal to 1% of the true increasing rate. When δ decreases to an extremely small value, such as $2 \times 10^{-4} < \delta < 8 \times 10^{-4}$, the result is under-estimated. While for a larger δ value, the result is over-estimated. In summary, the proposed method has an accurate estimation performance even for tiny fault severity.

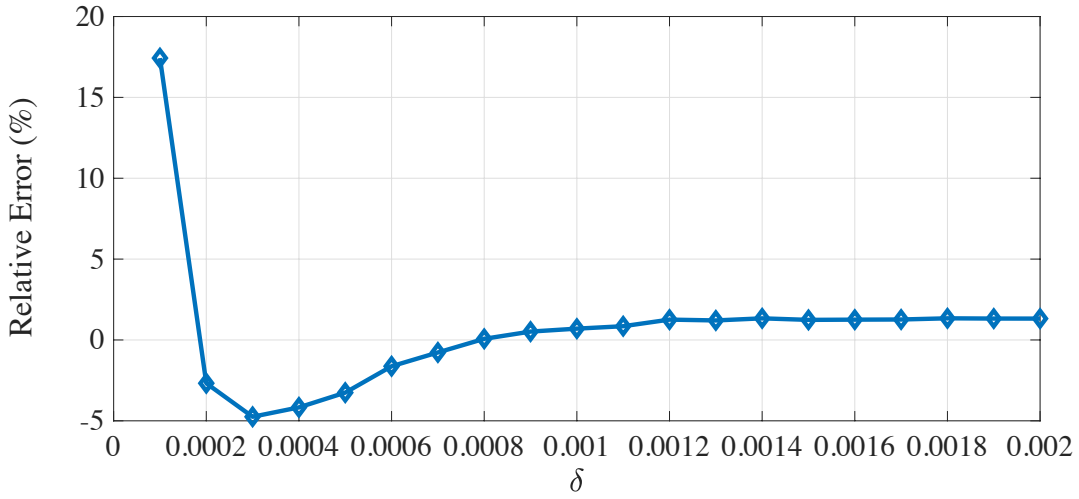

FIGURE 5.27 – Relative error versus the true value δ

Further, we study the methods' estimation performance for faults occurring at different variables and varying fault severity. Fig.5.28 shows relative errors for all faulty cases with SNR=30dB and varying FNR from −10dB to 20dB. It can be noticed that all faulty cases have similar evolution tendencies : as fault severity increase (FNR increase), the relative error decrease from a positive (over-estimation) and then increase to a negative value (under-estimation). Remarkably, consistent with fault isolation, F_7 and F_{10} are also the two most challenging cases for fault severity estimation : they have the worst estimation performance than the other faulty cases.

Finally, we compare the estimation performance of the proposed method with other approaches by evaluating the average relative error over all faulty cases. This comparative study considers the LMD-RBC, PCA-RBC, MD-RBC, AMD-RBC, KLD and JSD approaches. Since these approaches aim to estimate the fault amplitude, we convert the es-

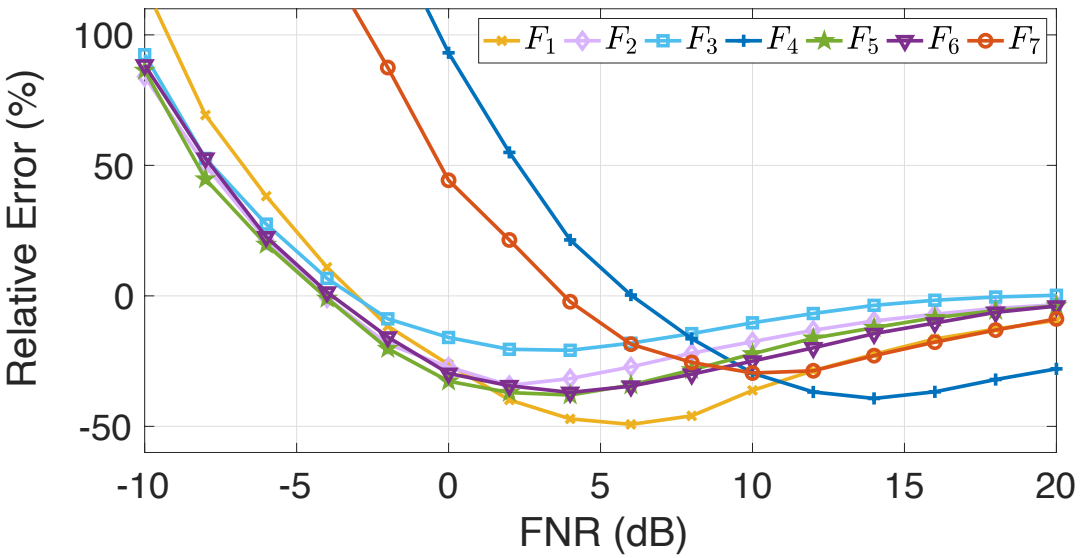

FIGURE 5.28 – Relative errors of the proposed method for the 7 faulty cases

timated fault amplitude into the corresponding fault increasing rate. Then the performance comparison can be performed fairly. The relative error result is illustrated in Fig.5.29. It indicates that the proposed method achieves the best estimation performance among these methods. The average relative error of the proposed method is less than 2%, and it decreases as FNR increases. Four RBC-based approaches have outstanding estimation performance, where the RBC method using the LMD index outperforms the other RBC approaches. However, KLD and JSD methods seem not accurate enough. Their average relative error is very large, especially for small fault severity.

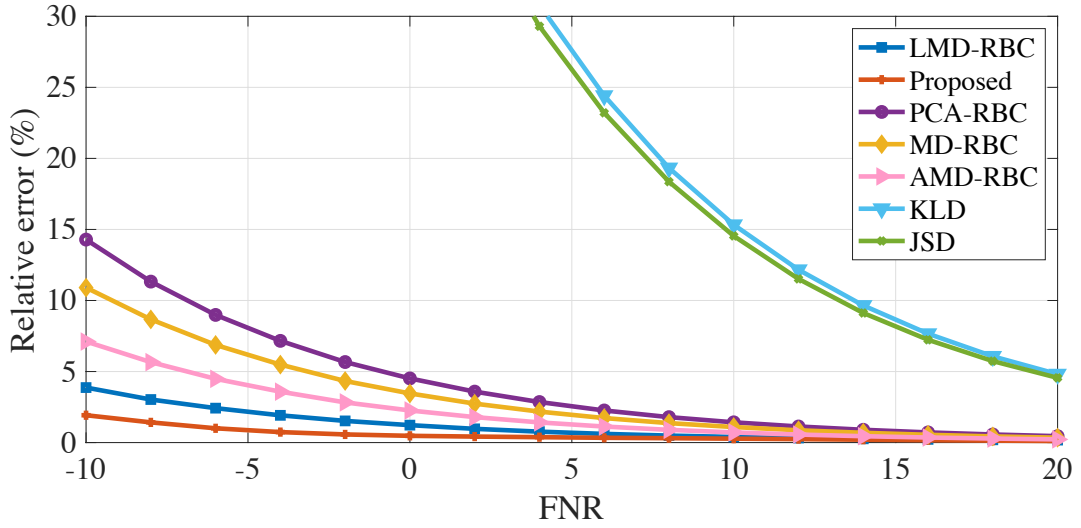

FIGURE 5.29 – Average relative errors of all the reported method.

5.3.4.2 Fault amplitude estimation efficiency

As the second way to assess the fault severity, fault amplitude is usually estimated to show faults' evolution tendency. Fault amplitude is more meaningful in assessing fault

severity than fault increasing rate since it reflects not only the increasing rate δ but also the deviation level. Unlike the increasing rate estimation, the estimation of fault amplitudes does not require the assumption of fault behavior, but it is usually a more challenging task. Previously, we proposed the LMD-RBC method that can simultaneously isolate the faulty variable and estimates their amplitude. Therefore, this subsection studies the fault amplitude estimation efficiency of the LMD-RBC method (introduced in section 4.6) for both single and multiple faults cases. We also compare the LMD-RBC method with other approaches based on the RBC framework to highlight the advantages of the proposed method. KLD and JSD approaches are not taken into account for their poor estimation performance as shown in the last subsection.

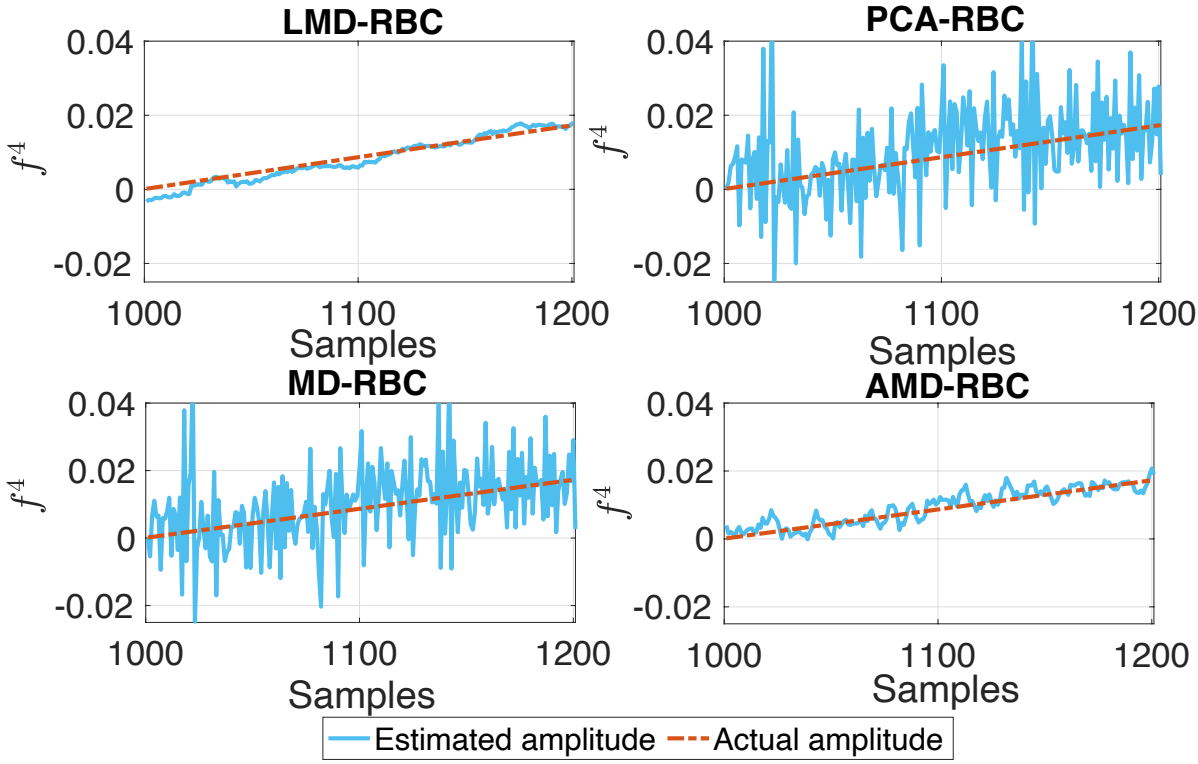

FIGURE 5.30 – Actual fault component and the estimated fault amplitude of the four RBC approaches for F_7 faulty case (a fault occurs in variable $\boldsymbol{x}_4$) in 10dB FNR condition.

For the single fault situation, we consider the most challenging faulty case F_7. The estimation results of all the methods based on the RBC framework for this case are shown in Fig.5.30. It highlights that the proposed LMD-RBC method outperforms others since its estimated results are closest to the actual value with only small variations. The evolution of the LMD-RBC result is slower than other methods, which mainly results from the exponential smoothing operation. Similarly, the AMD-RBC method using augmentation technique on adjacent samples also obtains a stable estimation result, even though its variations are slightly larger than LMD-RBC's. The error of PCA-RBC and MD-RBC methods are dramatic, and therefore their fault estimation performances are disappointing.

In order to deeply discuss the estimation performance for case F_7 with different fault severity, we calculated another typical error criterion, the mean square error (MSE), and compared the result of the proposed method with other RBC-based ones. For different FNR conditions, we repeated the test 500 times to obtain stable results. Fig.5.31 displays the MSE results in the case of F_7 for all the considered techniques. It highlights that

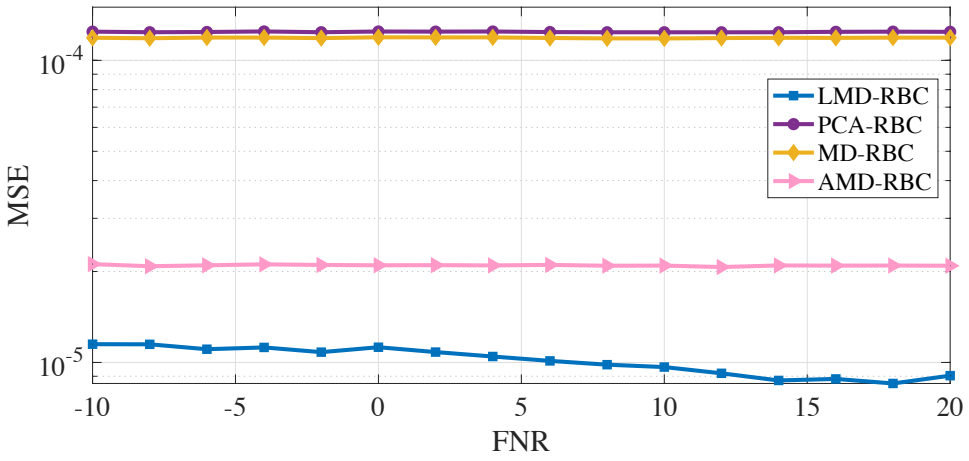

FIGURE 5.31 – Mean squared error for different methods in the case of F_7 (a fault occurs in variable $\boldsymbol{x}_4$)

the LMD-RBC method always achieves minimal error for all FNR conditions. Although the AMD-RBC method also has pronounced estimation performance, its fault severity estimation capability is significantly weaker than LMD-RBC. MD-RBC and PCA-RBC methods are similar and much worse than LMD-RBC approach. Therefore, the LMD-RBC method offers an accurate fault amplitude estimation solution for incipient faults.

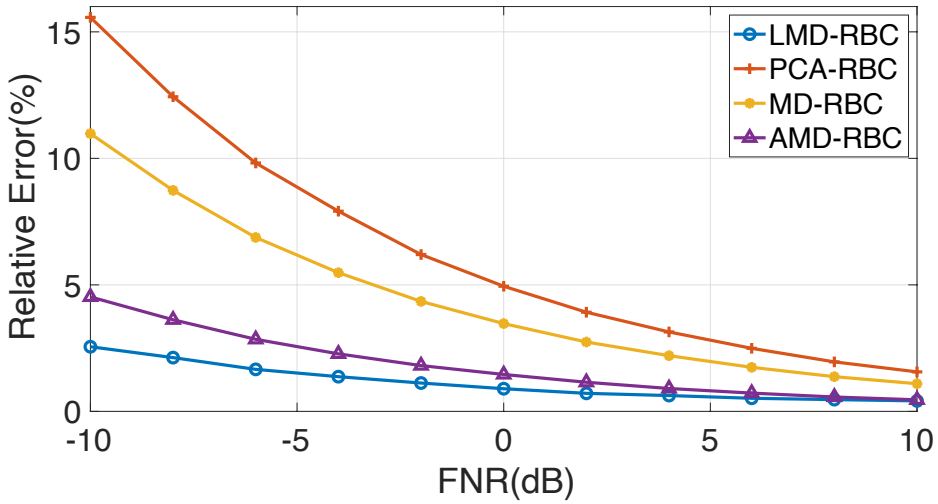

FIGURE 5.32 – Average relative error of the four RBC approaches for single fault

The comprehensive performance evaluation for different fault severity is then considered. We calculated the average relative error for all the estimation approaches and shown in Fig.5.32. The proposed method also achieves the best performance in the fault amplitude estimation task for its lowest relative error. According to the results, the absolute estimation error of the LMD-RBC method is less than 3% of the actual fault amplitude, which means that the estimated value is very close to the actual value, and the bias is insignificant. Indeed, for severe faults, such as 5 to 10dB FNR, the performances of LMD-RBC and AMD-RBC methods are nearly equal. While, for extremely tiny faults, such as -10 to -5dB FNR, the LMD method is the most reliable among the four approaches.

Multiple faults are common but challenging for diagnosis, which needs careful inves-

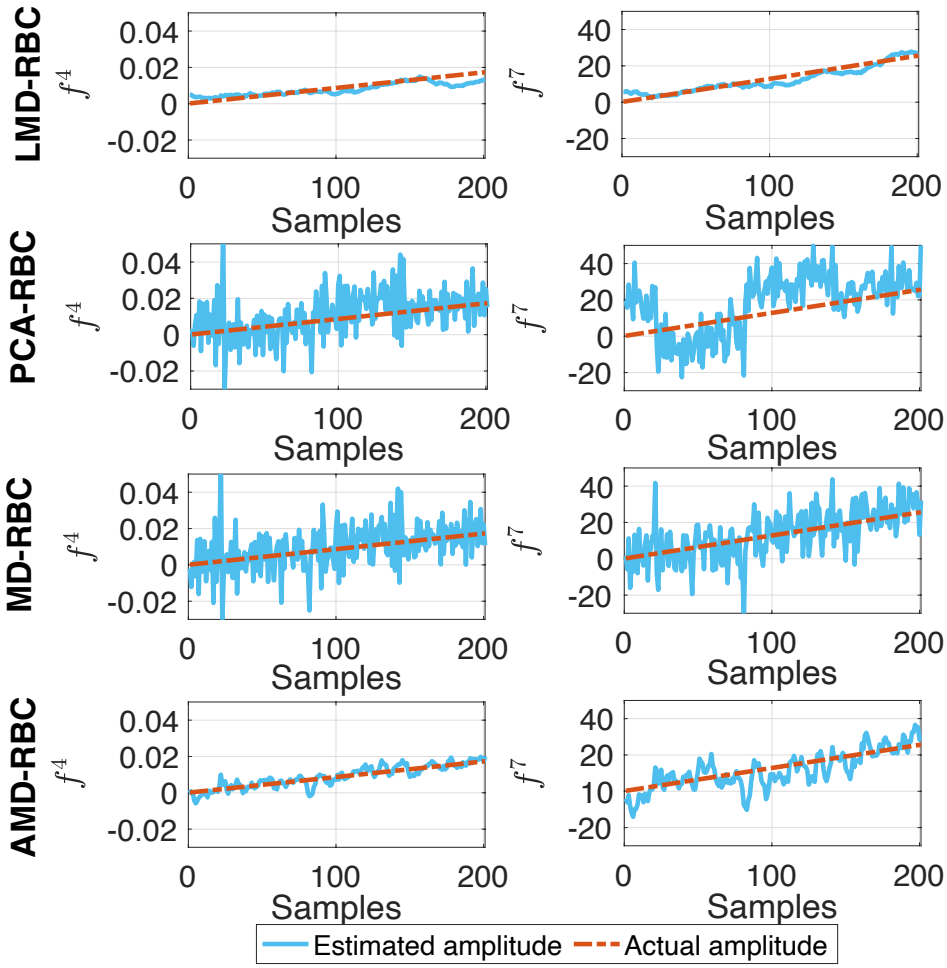

FIGURE 5.33 – Actual fault component and the estimated fault amplitude of the four RBC approaches for faults occurring in variables $\boldsymbol{x}_4$ and $\boldsymbol{x}_7$ in 10dB FNR conditions

tigation. The proposed LMD-RBC solution can be used for that purpose. With no loss of generality and in order to highlight the efficiency of the LMD-RBC approach with a reduced degree of complexity, we show this effectiveness for two faulty variables. The following example considers faults occurring at $\boldsymbol{x}_4$ and $\boldsymbol{x}_7$ in the CSTR system. In this case, the total number of candidate fault directions is 21. After applying the proposed and comparative methods, The estimated results are demonstrated in Fig.5.33, where the LMD-RBC method accurately estimated the amplitudes of the two faults with a minor bias. Although the estimated result of the AMD-RBC method is promising, the results' fluctuations are significant, especially for variable $\boldsymbol{x}_7$. The PCA-RBC and MD-RBC methods are inaccurate for fault amplitude estimation since their estimation error are enormous.

Then, we investigate the estimation efficiency of different RBC-based approaches concerning varying fault severity. To that end, we evaluate the average relative error of the four RBC approaches for two faulty situations. As shown in Fig.5.34, the results again validate the high accuracy of the proposed method for fault amplitude estimation. The similar results of the single fault and multiple faults cases given in Fig.5.34 and Fig.5.32, respectively, show that the number of faulty variables has less influence on the estimation procedure.

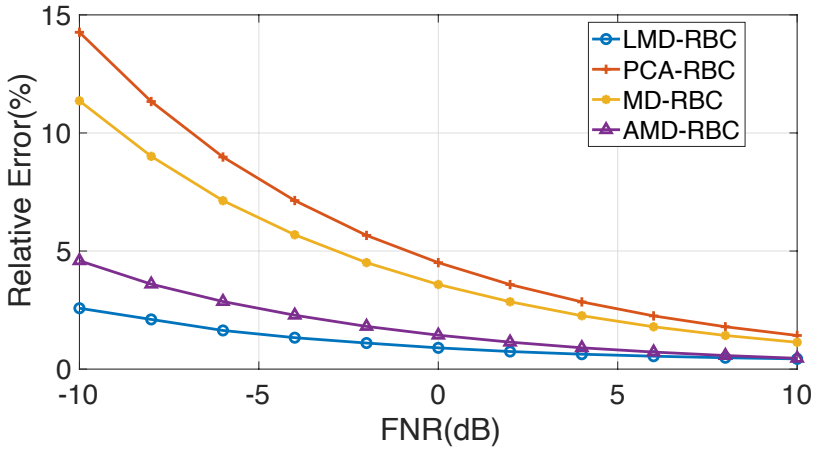

FIGURE 5.34 – Average relative error of the four RBC approaches for two faults

5.4 Feature efficiency evaluation based on bearing data

The last section demonstrated the specially designed fault diagnosis methods based on LMD. Nevertheless, LMD can also be used as a general feature for fault detection. Fault detection schemes using statistical feature extraction approaches and probability based distance are essential for fault detection. This section aims to discuss the feature extraction efficiency of LMD by considering the bearing data application. This study compares different statistical features, such as Local Mahalanobis distance, principal component analysis, kernel principal component analysis, and independent component analysis. For each, we evaluate and compare the capability of feature analysis using Kullback-Leibler divergence, Jensen-Shannon divergence, Wasserstein distance, and Kolmogorov Smirnov distance.

5.4.1 Case Western Reserve University Bearing Data

The Case Western Reserve University (CWRU) bearing data center published a standard bearing database based on a reliance electric motor to provide publicly available data for assessing the performance of bearing fault diagnosis approaches [118]. As shown in Fig.5.35, the experimental set-up of the CWRU data set consists of a reliance electric motor, a torque transducer/encoder (center), a dynamometer, and control electronics. Vibration signals were collected via accelerometers attached to the drive end and fan end of the motor housing, respectively. In the experiment, faulty bearings with different fault diameters ranging from 0.007 inches to 0.028 inches were reinstalled into the test motor for both fan and drive end. The fault types include inner race fault, ball fault, and three outer race faults located at centered, orthogonal, and opposite positions, respectively. Meanwhile, different motor loads from 0 to 3 horsepower (ph), which corresponds to the motor speeds of 1797 to 1720 RPM, were also considered in the experiment.

The comparative study uses the first 2000 healthy drive end and fan end signals under the 0 hp motor load condition as the reference data, shown in Fig.5.36 as an example, to determine the healthy models of PCA, KPCA, ICA, and LMD methods. For the KPCA technique, we used the Radial Basis Function (RBF) kernel and optimized its performance

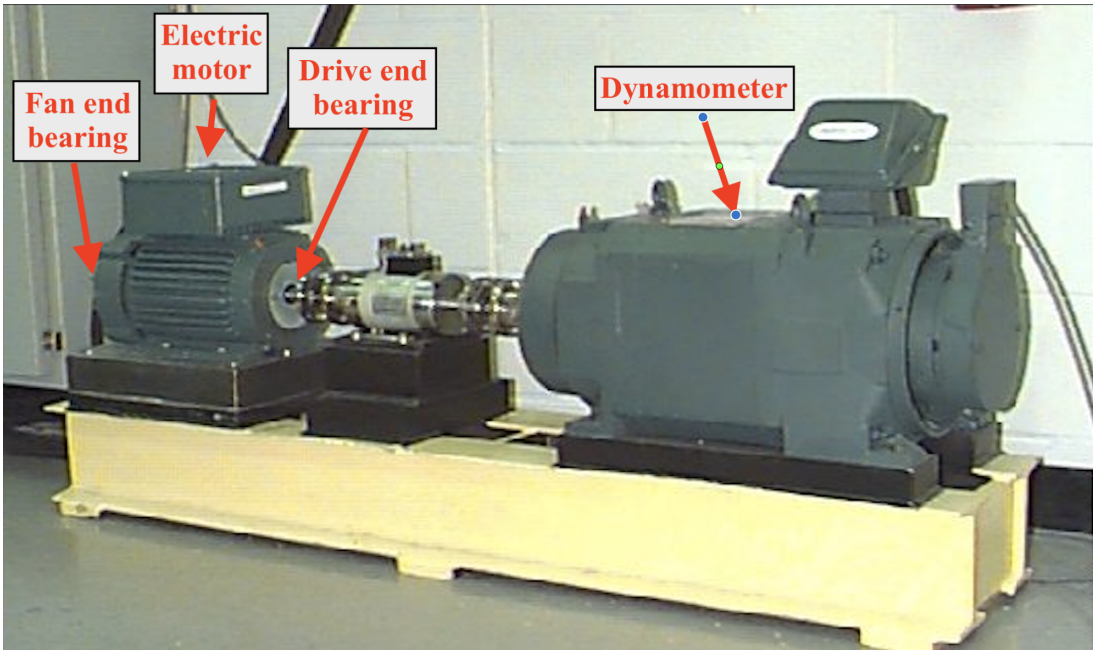

FIGURE 5.35 – CWRU bearing data experimental set-up.

using a cross-validation approach [131]. Then, the pre-fitted models of different approaches were reused to process the test data. As a reminder, the detection scheme (Fig. 4.12) was shown in subsection 4.3.3.

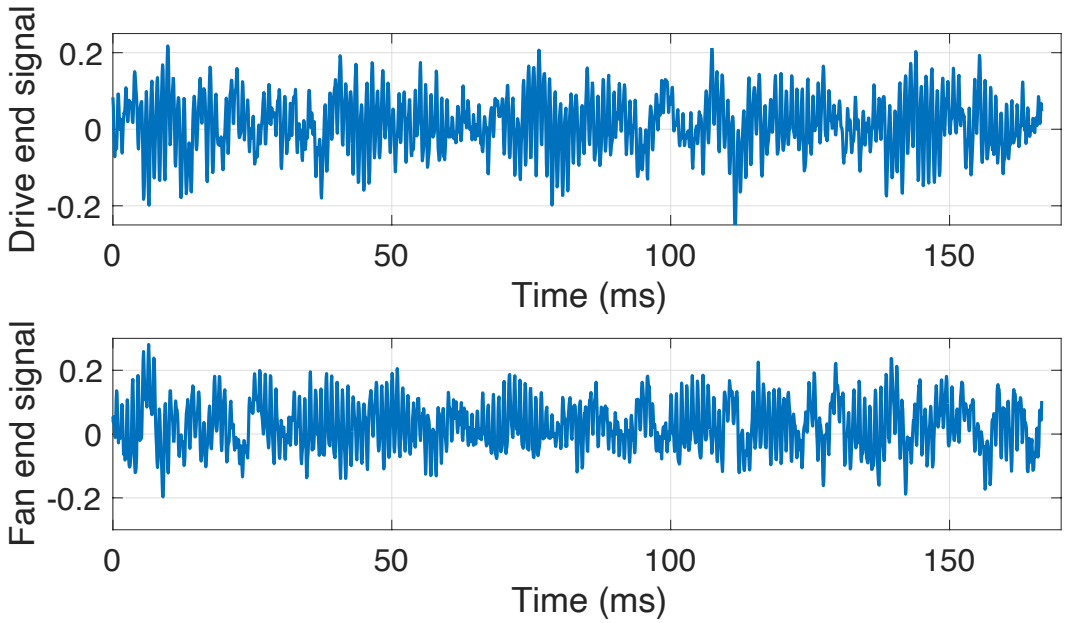

FIGURE 5.36 – The drive end (upper) and fan end (lower) healthy signals under the 0 hp motor load condition.

In the bearing fault detection procedure, we perform technical combinations of different feature extraction approaches and probability based distances on the vibration signal, where every 2000 data points were taken from the raw signal without overlapping as a testing sample. In this study, we only consider ten faulty cases with the smallest fault

diameter, detailed in Table.5.5, to evaluate the detection performance for tiny faults.

TABLE 5.5 – Detailed information of reference and ten faulty cases

ID	Fault type	Fault Location	Load (HP)	Diameter (inches)	Frequency (samples/sec)
Ref	Healthy	-	0	-	12000
F_1	Inner race	Drive	0	0.007	12000
F_2	Ball	Drive	0	0.007	12000
F_3	Outer race (centered)	Drive	0	0.007	12000
F_4	Outer race (orthogonal)	Drive	0	0.007	12000
F_5	Outer race (opposite)	Drive	0	0.007	12000
F_6	Inner race	Fan	0	0.007	12000
F_7	Ball	Fan	0	0.007	12000
F_8	Outer race (centered)	Fan	0	0.007	12000
F_9	Outer race (orthogonal)	Fan	0	0.007	12000
F_{10}	Outer race (opposite)	Fan	0	0.007	12000

5.4.2 Detection capability

In this experiment, we evaluate the detection capability of different technical combinations. Note that both the healthy reference data and faulty data are collected under the 0 hp motor load condition. The Area Under the receiver operating characteristic Curve (AUC) was used to summarize the global detection capability of different combinations since it is not affected by the choice of the threshold value and therefore is fair for performance comparison. As shown in Table.5.6, most technical combinations achieve perfect detection performance in this study since their AUC values reach the maximum. Note that the bold values in the table indicate the lowest AUC value. The unsatisfied results are the PCA-based and KPCA-based approaches for F_7 and F_9 cases corresponding to the ball fault of the fan end and the outer race fault of the fan end at the opposite position, respectively. Therefore, these two faulty cases seem more trick for detection compared to others. On the other hand, LMD and ICA approaches always have superior performance for all faulty cases indicating that they effectively extract the crucial features from vibration signals.

TABLE 5.6 – The AUC values of different combinations for ten faulty cases

ID	LMD				PCA				KPCA				ICA			
	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}
F_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_7	1.00	1.00	1.00	1.00	**0.98**	**0.99**	**0.99**	**0.93**	**0.98**	**0.98**	**0.98**	**0.93**	1.00	1.00	1.00	1.00
F_8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_9	1.00	1.00	1.00	1.00	1.00	**0.98**	**0.98**	**0.98**	1.00	**0.98**	**0.98**	**0.98**	1.00	1.00	1.00	1.00
F_{10}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.99**	1.00	1.00	1.00	**0.99**	1.00	1.00	1.00	1.00

Fig.5.37, as an example, shows the detection outcomes of using the four feature extraction techniques combined with the Wasserstein distance for the F_7 case. In this

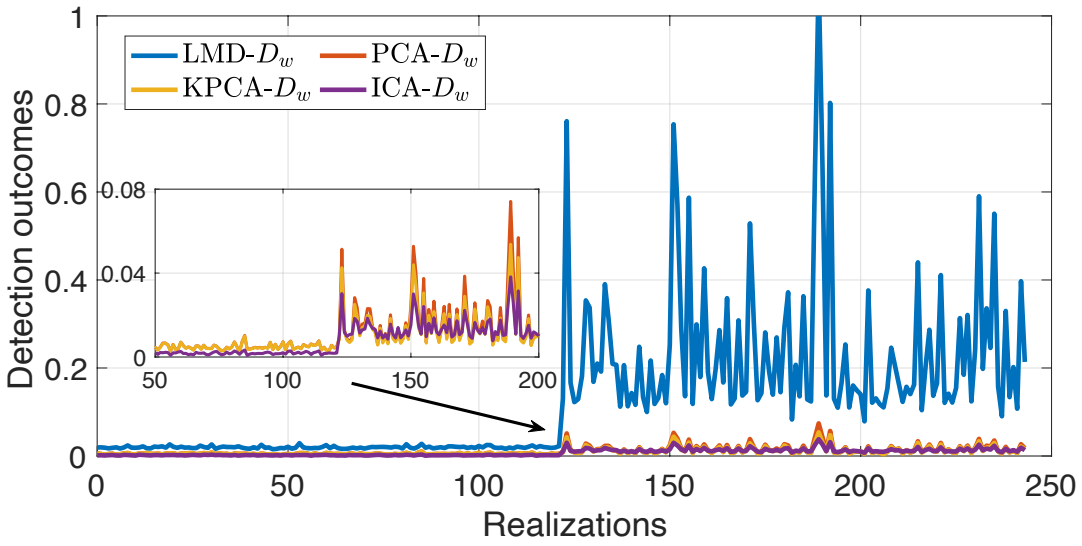

FIGURE 5.37 – The detection outcomes of the four combinations for F_7 case : LMD-D_w ; PCA-D_w ; KPCA-D_w ; ICA-D_w.

result, LMD and ICA-based combinations can totally distinguish healthy and faulty samples, while the PCA and KPCA-based combinations partly confuse the two cases. Moreover, the outcomes of LMD-based combinations for faulty samples are much more distinct than ICA-based combinations.

5.4.3 Detection sensitivity

Although the arbitrary combination of feature extraction techniques and feature analysis tools has excellent performance for bearing fault detection, their limitation of detection capability is still unknown. The detection performance naturally degenerates with the fault severity decrease, such as a smaller fault diameter. Since the CWRU bearing data does not contain fault diameters smaller than 0.007 inches, we further evaluate the detection sensitivity of approaches to reveal their potential detection capability for tiny faults.

TABLE 5.7 – The sensitivity values of different combinations for ten faulty cases

ID	LMD				PCA				KPCA				ICA			
	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}
F_1	**282.86**	33.94	22.92	19.67	29.78	45.71	44.41	9.96	24.76	46.22	44.40	9.70	25.62	47.91	31.62	8.56
F_2	**26.68**	19.55	16.32	7.63	4.58	7.73	7.56	3.04	4.43	7.66	7.31	2.91	6.47	10.52	8.78	2.40
F_3	**476.42**	33.41	25.44	18.41	36.58	32.43	23.56	11.65	17.73	31.40	22.63	10.16	35.03	29.59	20.88	8.41
F_4	**755.94**	31.79	24.84	18.49	60.96	61.13	41.36	13.85	33.12	74.37	49.01	13.47	21.87	30.12	21.84	6.69
F_5	**115.05**	30.23	22.85	8.89	11.35	12.86	13.60	5.07	9.37	13.38	13.81	5.05	7.71	19.31	13.22	3.37
F_6	**118.62**	28.99	23.10	10.24	15.45	21.56	16.64	5.67	12.96	21.87	16.59	5.57	26.24	27.21	19.75	7.78
F_7	**24.24**	21.55	14.85	4.25	3.39	2.89	3.22	0.97	2.92	2.98	3.24	0.94	6.32	14.90	10.91	4.59
F_8	**190.67**	30.66	23.89	10.97	35.87	27.04	25.93	8.03	22.09	26.34	24.39	7.60	37.56	50.42	30.47	7.54
F_9	7.73	**10.42**	8.13	2.79	1.25	0.83	1.06	1.10	1.21	0.83	1.03	1.10	4.82	5.86	5.07	3.63
F_{10}	**23.35**	17.63	14.73	7.03	2.30	3.44	3.08	1.56	2.19	3.27	2.94	1.51	8.05	12.02	9.16	5.56

The detection sensitivity values of different combinations are shown in Table.5.7, where the bold values correspond to the highest sensitivity value leading to the best detection performance. The result shows that the sensitivity value of the combinations using the LMD feature is significantly larger than the combination using other feature extraction

tools. Healthy and faulty samples can be effectively distinguished in the LMD feature space. Despite the effectiveness of the LMD feature, different analysis tools can lead to diverse detection performance, where the Wasserstein distance has a higher sensitivity value than the other three distances. However, the superiority of Wasserstein distance does not exist when it combines with PCA, KPCA, or ICA features. In other words, a proper choice of combination of features extraction approach and feature analysis tool is essential to achieve a promising detection performance. In addition, the performances of two entropy-based methods, namely Kullback-Leibler divergence and Jensen-Shannon divergence, are similar, and they outperform the Kolmogorov Smirnov distance.

5.4.4 Robustness to non-stationary operating conditions

In the previous experiments, both the reference and test data were collected under the 0 hp motor load condition without concerning the effectiveness for different operating conditions. The change of operating conditions may lead to a false detection issue when the reference data only consists of a single operating condition. This experiment investigates the robustness of different combinations for non-stationary operating conditions.

TABLE 5.8 – The AUC values of different combinations for healthy signals under different operating conditions

Load (ph)	LMD				PCA				KPCA				ICA			
	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}	D_w	D_{kl}	D_{js}	D_{ks}
0	0.52	**0.50**	0.49	0.49	0.43	0.45	0.45	0.46	0.43	0.45	0.45	0.46	0.44	0.48	0.47	0.45
1	0.24	0.17	0.23	**0.47**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.91	0.92	0.94
2	0.16	0.12	0.18	**0.39**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.98	0.99	0.97
3	0.79	**0.55**	0.57	0.62	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.77	0.89	0.89	0.86

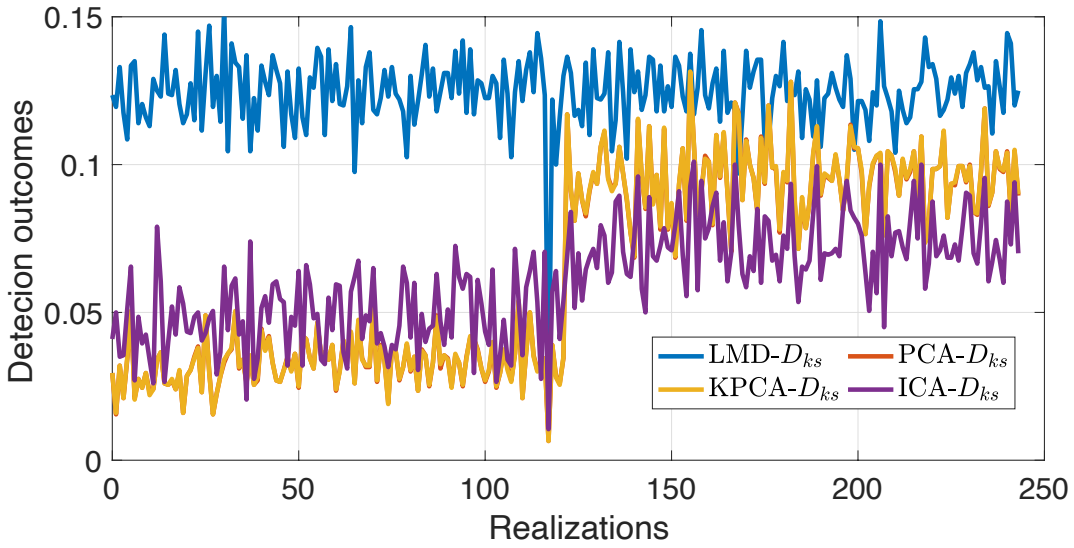

FIGURE 5.38 – The detection outcomes of the four combinations for healthy signals with 1 ph motor load : LMD-D_{ks} ; PCA-D_{ks} ; KPCA-D_{ks} ; ICA-D_{ks}.

To that end, we used healthy data under the 0 hp motor load condition as the reference samples and performed fault detection approaches on other healthy data with different motor loads. A robust fault detection approach is expected to produce consistent outcomes

for healthy data regardless of operating conditions. Accordingly, the AUC value between the reference and testing samples should be close to 0.5. Table.5.8 gives the AUC values of different combinations for varying motor loads. In this table, we highlight the results closet to 0.5 in bold. The result demonstrates that approaches using PCA and KPCA features have higher AUC values, inferring that these approaches easily cause a false detection issue when the operating condition changes. Similarly, fault detection methods based on the ICA feature also suffers from potential false detection issue though their AUC values are slightly smaller than PCA and KPCA. Conversely, the AUC values of the approaches using the LMD feature are much close to 0.5, which illustrates that their detection outcomes are consistent for healthy data coming from diverse conditions.

This performance is further confirmed in Fig.5.38, which illustrates the detection outcomes of the four feature extraction techniques combined with the Kolmogorov Smirnov distance for the healthy signal with 1 ph motor load. Therefore the robustness of the LMD technique remarkably outperforms other feature extraction approaches.

5.5 Conclusion

This chapter focus on the efficiency of the proposed fault diagnosis approaches based on local Mahalanobis distance and the efficiency of this diagnosis index as a feature. To comprehensively evaluate methods' efficiency, we first introduce different kinds of evaluation criteria. Then based on the CSTR process data from industrial scenarios, we evaluate the efficiency of our fault diagnosis proposal in three diagnosis tasks : fault detection, faulty variable isolation, and fault severity estimation.

The performance analysis shows that the proposed fault detection method based on LMD is effective for data without any distribution-type assumption. The proposal is robust to noise influence and achieves pronounced performances in incipient fault detection (high sensitivity to incipient faults). Comparing the two diagnostic strategies using LMD alone and combining LMD with the EPD-CUSUM technique, we found that EPD-CUSUM helps to improve incipient fault detection capability significantly. Compared to other well-tuned and efficient state-of-the-art methods, the detection capability of our proposal outperforms these techniques in terms of false alarm probability, detection probability, detection delay, and AUC.

Concerning the isolation efficiency, the two proposed solutions based on LMD also achieve state-of-the-art performance. For single fault cases, the LMD-based contribution plot approach outperforms the other comparative methods, including the approaches based on Kullback-Leibler divergence and Jensen-Shannon divergence ; RBC methods based on LMD ; principal component analysis ; and Mahalanobis distance. The total accuracy of this approach reaches 100% for incipient faults with FNR larger than 5 dB. The LMD-based contribution plot approach is more sensitive to incipient faults than other approaches. Although the performance of the LMD-based contribution plot approach is slightly better than that of LMD-RBC method in single fault cases, this method is not developed for multiple faults cases. In contrast, the LMD-RBC method is available for both single and multiple faults cases. When multiple faults occur simultaneously, the isolation accuracy of all the reported methods decreases compared to the single fault case. However, the comparative study indicates that the LMD-RBC approach significantly outperforms other

RBC methods, especially for incipient fault isolation.

For the fault severity estimation task, we proposed two LMD-based solutions for estimating the fault's increasing rate and fault amplitude. The first proposal makes an assumption that the fault's evolution is a first-order function of time. Therefore, we investigate the estimation performance of this approach by considering a constant and dynamic increasing rate, respectively. The result indicates that when the assumption is satisfied, the estimation error is small, even for tiny faults. If the increasing rate is not constant (assumption is not satisfied), the error is small at the beginning of a fault but increases significantly with time. The case study on the CSTR process indicates that the average relative error of the estimation method is less than 5%, and its performance outperforms other approaches.

The second solution of fault severity estimation is based on the RBC framework using LMD, and it aims to estimate fault amplitude. This approach effectively estimates single and multiple faults. The average estimation error of the proposed method is less than 3% of the actual fault amplitude, showing promising performance in the fault estimation task. Compared with the traditional RBC approaches, the new RBC-based proposal offers significant progress in both faulty variables isolation and fault amplitude estimation tasks.

This chapter also investigates the efficiency of LMD as a statistical feature extraction tool by using the bearing data. The fault detection scheme extracts the LMD feature and uses probability-based distance as an analysis tool. Overall, this fault detection scheme achieves a reliable performance for bearing fault detection tasks, even for slight fault diameters. The comparative study on the CWRU bearing data indicates that the LMD technique, as a feature extraction approach, is more effective and robust than the PCA, KPCA, and ICA techniques. Remarkably, the combination of LMD and Wasserstein distance achieves the best sensitivity performance in this study. Regardless of the probability based distance, detection approaches using the LMD feature have strong robustness to non-stationary operating conditions.

In summary, the performance analysis for the considered industrial applications indicates that LMD has numerous notable advantages, such as high sensitivity to incipient faults, robustness to noise and outliers, and distribution-free assumption. Fault diagnosis approaches developed on the LMD technique can achieve state-of-the-art performance in fault detection, faulty variable, and fault severity estimation tasks. Besides, LMD can also be used as an efficient feature for fault detection. It can provide the representative information of a faulty behavior and allows accurate detection by subsequently using probability based distance.

6

Conclusion and perspectives

6.1 Conclusion of the study

Fault diagnosis and prognosis play a key role in modern complex systems for increasing requirements on reliability, availability, maintainability, and safety. Fault diagnosis techniques can save human lives and significantly increase productivity and efficiency. Therefore, more and more fault diagnosis techniques are applied to different industrial scenarios, making it as a hot topic in the signal processing domain. Generally, fault diagnosis problem is divided into three tasks, namely fault detection, faulty sources isolation, and fault severity estimation, according to different golds of applications. In the last decade, a large number of approaches have been proposed for each task, and they have achieved significant improvement. However, these problems still face numerous challenges, especially in industrial applications, such as the lack of sufficient faulty data for training, ineffectiveness to complex distributed data, low sensitivity to incipient faults, and interference of noise and outliers. Recently, one-class classification techniques, aiming to identify one class of samples, tend to become promising solutions for fault diagnosis in industrial applications, particularly for the challenging task of incipient fault detection when low faulty information are available.

One-class classification techniques aim to determine a minimum region based on the given healthy samples to distinguish between faulty and healthy samples. From this point of view, this work shows that the two traditional statistics, namely Hotelling T^2 and SPE, are ineffective for non-Gaussian distributed data. Therefore, we propose a more effective approach for healthy region approximation by focusing on the local information of spatial distribution. The proposed method develops a healthy region by determining multiple centers and a local region margin. To that end, a robust anchor-generation algorithm and a region margin selection approach are proposed. A particular distance measurement called local Mahalanobis distance is then defined to indicate the distance between a sample and the healthy region. The proposed method can avoid the distractions of outliers in

the training process. By comparing the performance of different OCC approaches, we show that this proposal is more accurate, effective for data without any distribution-type assumption, and robust against outliers.

Based on the proposed healthy region approximation method and the LMD index, we develop specific fault diagnosis methodologies for three diagnosis tasks (fault detection, isolation and estimation). Using time-series signals, we propose the fault detection framework combining the LMD with the empirical probability density cumulative sum method to achieve high sensitivity performance. With time-independent data, LMD is used to extract features, and probability-based distance is then applied to decide if test samples deviate from the healthy pattern.

A fault isolation approach is also developed based on the LMD index and the contribution plot idea. This approach calculates the contribution value of each variable to the fault by analyzing the relative position between faulty samples and their corresponding anchors. The proposed fault isolation method can recognize a single faulty variable, even for tiny faults. Subsequently, we derive an analytical expression from the LMD index to estimate the fault increasing rate. These two methods preserve the intrinsic advantages of LMD, such as robustness to outliers, no distribution assumption, and high sensitivity for incipient faults. To enable the isolation and fault severity estimation for multiple faults cases, we further propose a novel reconstruction-based method using the LMD index for the two tasks. This method combines the high sensitivity characteristics of the LMD techniques with the reliability of the RBC technique. It can isolate multiple faulty variables and estimate their fault amplitude simultaneously. Using the simulation data, we validate the proposed methods' effectiveness.

Based on the CSTR process data, we deeply evaluate the efficiency of our fault diagnosis proposal. The full performance analysis shows that the proposed fault detection method based on LMD is effective for complex distribution and robust to noise. Our proposals outperform state-of-the-art methods in three fault diagnosis tasks. The comparative study on the CWRU bearing data indicates that the LMD technique can be used as an efficient and accurate feature extraction approach. It is more effective and robust than other transformation-based techniques like the PCA, KPCA, and ICA techniques. Therefore, this study shows that the LMD technique has significant benefits for the fault diagnosis problem.

6.2 Perspectives

The proposed healthy region approximation method, along with the local Mahalanobis distance, is the core of this study. Based on this technique, fault diagnosis solutions for fault detection, faulty variable isolation, and fault severity estimation are developed and achieve state-of-the-art performance. However, the LMD technique and the developed fault diagnosis solutions can be further improved in some aspects.

- The first perspective of this work concerns the healthy region approximation approach. This proposed method in work is implemented in two steps : anchor generation and region margin selection. The optimization procedure of these two steps, minimizing the size of the healthy region, is performed sequentially, which may not achieve the globally optimal goal. Therefore, it is expected to optimize

the two steps at the same time so as to produce an optimal healthy region. A reasonable objective function should be first determined for this purpose. An effective optimization algorithm can be then selected to solve the optimization problem, such as Stochastic Gradient Descent (SGD) algorithm, Newton's method, and Sequential Quadratic Programming (SQP).

- Secondly, we consider the performance improvement of detection delay. This work proposed the detection strategy combining the LMD with the EPD-CUSUM technique. This proposal achieves state-of-the-art performance for the fault detection task in terms of detection delay. However, this performance should be further improved to increase the security of the systems. We can achieve this goal in two ways. First, we can refine the healthy region to obtain a more accurate healthy region. Then, the combination of LMD with other highly sensitive detection techniques should be studied. For example, an improved Exponential Weighted Moving Average algorithm (EWMA) approach was proposed and shown prominent performance in detecting small changes [91]. Then, we will deeply investigate this technique and consider its combination with LMD.
- We will also consider the performance of the proposed fault diagnosis methods for dynamic systems. In this study, fault diagnosis approaches are developed for the steady state of systems, meaning that they may incorrectly identify the healthy dynamic system as faulty. To cope with this problem, we need to improve the LMD technique to adapt to the dynamic characteristic of systems. The possible improvement of the LMD technique is to consider the temporal correlation between adjacent samples in developing a healthy region. For example, the augmented matrix of adjacent samples can be generated and used for anchor generation.
- The study of combining deep learning techniques with LMD is also taken into account in our further work. Deep learning techniques can learn relevant features automatically, which has been extremely successful in many scenarios. The benefits of deep learning techniques can be transferred to one-class classification approaches. For example, the deep one-class classification approach [152] is a meaningful attempt in this direction, which inspires us to leverage the LMD technique for deep learning.
- The applications of our fault diagnosis proposals to other real industrial scenarios will be considered in our further works. This work has studied the performance of the proposed methods on CSTR and CWRU data. However, various applications have different challenges, especially in real industrial environments. The performance verification and evaluation of these approaches in different applications are necessary to help us improve the fault diagnosis methodologies. For example, the proposed LMD-based detection approach can be applied to the Tennessee Eastman Process data. The proposed RBC-LMD approach can isolate the faulty sensors and estimate their deviations in high-speed trains' air brake control system. Then, the limitation and advantages of these approaches can be further validated in these applications.

A
Appendix

A.1 Introduction to extreme value statistics

Theorem A.1.1. *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed (i.i.d) random variables with cumulative distribution function $\mathbb{F}(X)$ and consider their maximum value M_n of N samples, i.e. $M_n = \max\{X_1, \cdots, X_N\}$. The induced distribution of M_n can only take one of three forms, Gumbel, Weibull, or Frechet with ρ, β, and τ_a respectively denoted as the location, scale, and shape parameters :*

- *Gumbel*

$$\overline{F}_G(x, \rho, \beta) = \exp\left[-\exp\left(-\frac{x-\rho}{\beta}\right)\right] \qquad x \in \mathbb{R}, \qquad \beta > 0 \tag{A.1}$$

- *Weibull*

$$\overline{F}_W(x, \rho, \beta, \tau_a) = \begin{cases} 1 & x \geq \rho \\ \exp\left[-\left(\frac{\rho-x}{\beta}\right)^{\tau_a}\right] & \text{otherwise} \end{cases} \tag{A.2}$$

- *Frechet*

$$\overline{F}_F(x, \rho, \beta, \tau_a) = \begin{cases} \exp\left[-\left(\frac{\beta}{x-\rho}\right)^{\tau_a}\right] & x \geq \rho \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

Let $m_n = \min\{X_1, \cdots, X_N\}$. Similarly, the induced distribution of m_n can be one of three :

- *Gumbel*

$$\underline{F}_G(x, \rho, \beta) = 1 - \exp\left[-\exp\left(\frac{x-\rho}{\beta}\right)\right] \qquad x \in \mathbb{R}, \qquad \beta > 0 \tag{A.4}$$

- *Weibull*

$$\underline{F}_W(x,\rho,\beta,\tau_a) = \begin{cases} 0 & x \le \rho \\ 1-\exp\left[-\left(\frac{x-\rho}{\beta}\right)^{\tau_a}\right] & x > \rho \end{cases} \tag{A.5}$$

- *Frechet*

$$\underline{F}_F(x,\rho,\beta,\tau_a) = \begin{cases} 1-\exp\left[-\left(\frac{\rho}{\rho-x}\right)^{\tau_a}\right] & x \le \rho \\ 1 & \text{otherwise} \end{cases} \tag{A.6}$$

Eqs.(A.1)–(A.3) can be unified into a simple form for maxima

$$\overline{\Phi}_G(x,\rho,\beta,\tau_a) = \exp\left\{-\left[1+\tau_a\left(\frac{x-\rho}{\beta}\right)\right]^{-1/\tau_a}\right\} \tag{A.7}$$
$$-\beta-\tau_a(x-\rho)\le 0, \qquad \beta>0$$

Eqs.(A.4)–(A.6) can be unified into a simple form for minima

$$\underline{\Phi}_G(x,\rho,\beta,\tau_a) = 1-\exp\left\{-\left[1+\tau_a\left(\frac{\rho-x}{\beta}\right)\right]^{-1/\tau_a}\right\} \tag{A.8}$$
$$-\beta-\tau_a(\rho-x)\le 0, \qquad \beta>0$$

For more details about extreme value statistics, readers can refer to [9, 95].

A.2 Detailed fault detection result for CSTR data

TABLE A.1 – Detection delay time (hours) of different fault detection approaches for CSTR data

Method	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
LMD	1.500	1.110	0.930	0.740	1.300	1.040	0.170	0.370	0.340	1.310
GCCA-T^2_{r1}	11.050	9.470	8.330	2.330	3.660	8.730	0.410	2.290	6.620	10.510
GCCA-T^2_{r2}	10.540	2.950	2.930	2.300	9.130	3.070	0.380	0.880	0.800	4.350
CVA-T^2	4.180	4.820	3.390	2.000	2.820	3.900	0.360	0.870	1.480	5.220
CVA-SPE	2.740	2.630	2.420	2.350	4.410	2.900	0.230	0.530	0.800	3.380
CVDA	1.460	1.590	1.540	1.440	2.500	2.190	**0.150**	0.370	0.530	2.370
PCA-T^2	4.383	6.780	5.239	7.206	7.828	8.058	7.856	6.938	2.465	8.403
PCA-SPE	4.154	8.248	3.931	7.288	2.746	3.114	7.489	3.367	3.293	7.507
ICA	4.859	8.956	5.024	7.055	7.955	2.696	8.258	5.208	5.923	7.759
PLS	8.150	8.767	7.707	1.772	4.077	5.980	15.667	11.180	11.292	5.141
OC-SVM	**0.354**	**0.300**	**0.307**	0.377	0.431	**0.305**	0.299	0.346	**0.335**	0.381
k-centers	3.870	4.892	3.362	6.116	2.212	2.638	5.823	2.947	2.887	5.309
AE	10.338	11.977	10.201	10.345	12.930	14.072	15.683	14.265	13.657	7.639
IF	0.372	0.408	0.382	**0.372**	**0.347**	0.402	0.238	**0.320**	0.383	**0.316**

TABLE A.2 – False alarm rate of different fault detection approaches for CSTR data.

Method	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
LMD	0.0139	0.0152	0.0197	0.0103	0.0103	0.0095	0.0111	0.0160	0.0113	0.0113
GCCA-T^2_{r1}	0.0007	**0.0005**	**0.0004**	0.0007	0.0005	0.0008	0.0007	**0.0007**	0.0009	0.0006
GCCA-T^2_{r2}	0.0013	0.0012	0.0009	0.0011	0.0008	0.0012	0.0011	0.0011	0.0014	0.0009
CVA-T^2	0.0057	0.0059	0.0046	0.0052	0.0033	0.0057	0.0055	0.0051	0.0066	0.0034
CVA-SPE	0.0057	0.0064	0.0053	0.0067	0.0050	0.0054	0.0052	0.0064	0.0066	0.0050
CVDA	0.0089	0.0093	0.0075	0.0083	0.0057	0.0089	0.0083	0.0077	0.0112	0.0060
PCA-T^2	0.0029	0.0030	0.0026	0.0024	0.0028	0.0025	0.0017	0.0028	0.0023	0.0028
PCA-SPE	0.0033	0.0038	0.0037	0.0032	0.0034	0.0030	0.0023	0.0037	0.0028	0.0034
ICA	0.0022	0.0023	0.0017	0.0014	0.0014	0.0018	0.0013	0.0026	0.0014	0.0019
PLS	0.0020	0.0027	0.0014	0.0014	0.0012	0.0013	0.0014	0.0026	0.0017	0.0018
OC-SVM	0.5221	0.4974	0.4708	0.4534	0.4773	0.4799	0.4780	0.4913	0.4955	0.4744
k-centers	0.0394	0.0298	0.0242	0.0273	0.0244	0.0133	0.0263	0.0267	0.0175	0.0281
AE	**0.0001**	0.0006	0.0002	**0**	**0**	**0.0001**	**0.0001**	0.0008	**0**	**0.0003**
IF	0.2556	0.2562	0.2534	0.2418	0.2734	0.2286	0.2586	0.2725	0.2680	0.2548

TABLE A.3 – Detection rate of different fault detection approaches for CSTR data

Method	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
LMD	0.8646	**0.9249**	**0.9354**	**0.9511**	**0.9084**	**0.9273**	0.9893	0.9770	**0.9787**	**0.9039**
GCCA-T^2_{r1}	0.4822	0.3995	0.5397	0.8646	0.7823	0.4836	0.9778	0.8719	0.6227	0.3790
GCCA-T^2_{r2}	0.5088	0.8319	0.8351	0.8673	0.4722	0.8247	0.9793	0.9524	0.9562	0.7526
CVA-T^2	0.7088	0.6974	0.7743	0.8852	0.8290	0.7666	0.9804	0.9515	0.9149	0.6827
CVA-SPE	0.8235	0.8351	0.8504	0.8555	0.7174	0.8192	0.9869	0.9693	0.9542	0.7889
CVDA	**0.9105**	0.8998	0.9032	0.9138	0.8400	0.8612	**0.9918**	**0.9792**	0.9695	0.8499
PCA-T^2	0.4483	0.4815	0.4917	0.0024	0.0030	0.0028	0.0027	0.0028	0.8432	0.0371
PCA-SPE	0.4644	0.3415	0.5062	0.0031	0.7834	0.7752	0.0032	0.7782	0.7776	0.0060
ICA	0.4545	0.1605	0.4371	0.0017	0.0024	0.8243	0.0018	0.6479	0.5829	0.2528
PLS	0.3834	0.3249	0.4000	0.8854	0.7279	0.5899	0.0018	0.2905	0.2920	0.6438
OC-SVM	0.7997	0.8044	0.8305	0.4899	0.9071	0.9091	0.4878	0.9030	0.9077	0.8812
k-centers	0.5158	0.5233	0.5888	0.0215	0.8272	0.7903	0.0300	0.7940	0.8047	0.4078
AE	0.2500	0.0254	0.2430	0.0000	0.0082	0.0090	0.0001	0.0128	0.0082	0.4016
IF	0.7628	0.7957	0.8422	0.3608	0.3417	0.5606	0.7934	0.8086	0.7473	0.6846

TABLE A.4 – AUC performance of different fault detection approaches for CSTR data

Method	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
LMD	0.976	**0.982**	**0.983**	**0.987**	**0.974**	**0.981**	0.996	**0.994**	**0.994**	**0.982**
GCCA-T^2_{r1}	0.895	0.968	0.973	0.986	0.970	0.979	0.995	0.991	0.992	0.962
GCCA-T^2_{r2}	0.943	0.977	0.979	0.984	0.939	0.973	**0.997**	0.993	**0.994**	0.969
CVA-T^2	0.954	0.912	0.958	0.972	0.960	0.964	0.992	0.989	0.970	0.937
CVA-SPE	0.973	0.971	0.977	0.978	0.963	0.973	0.992	0.988	0.988	0.968
CVDA	**0.981**	0.979	**0.983**	0.982	0.971	0.977	0.994	0.992	0.992	0.974
PCA-T^2	0.810	0.882	0.854	0.500	0.499	0.504	0.494	0.504	0.966	0.740
PCA-SPE	0.811	0.867	0.882	0.511	0.964	0.957	0.509	0.957	0.961	0.769
ICA	0.813	0.807	0.848	0.506	0.546	0.965	0.509	0.925	0.911	0.836
PLS	0.853	0.892	0.881	0.985	0.966	0.949	0.501	0.933	0.935	0.954
OC-SVM	0.639	0.653	0.680	0.518	0.715	0.715	0.505	0.706	0.706	0.703
k-centers	0.738	0.747	0.782	0.497	0.901	0.888	0.502	0.884	0.894	0.690
AE	0.760	0.776	0.814	0.521	0.846	0.854	0.509	0.839	0.838	0.867
IF	0.754	0.770	0.794	0.559	0.534	0.666	0.767	0.768	0.740	0.715

Bibliography

[1] Alireza Abbaspour, Sohrab Mokhtari, Arman Sargolzaei, and Kang K Yen. A survey on active fault-tolerant control systems. *Electronics*, 9(9) :1513, 2020.

[2] Mohsen Mohammadi Alamuti, Hassan Nouri, Rade M Ciric, and Vladimir Terzija. Intermittent fault location in distribution feeders. *IEEE Transactions on Power Delivery*, 27(1) :96–103, 2011.

[3] Carlos F Alcala and S Joe Qin. Reconstruction-based contribution for process monitoring. *Automatica*, 45(7) :1593–1600, 2009.

[4] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191. American Mathematical Soc., 2000.

[5] Arslan Ahmed Amin and Khalid Mahmood Hasan. A review of fault tolerant control systems : advancements and applications. *Measurement*, 143 :58–68, 2019.

[6] Md Tanjin Amin, Faisal Khan, Syed Imtiaz, and Salim Ahmed. Robust process monitoring methodology for detection and diagnosis of unobservable faults. *Industrial & Engineering Chemistry Research*, 58(41) :19149–19165, 2019.

[7] Rajeevan Arunthavanathan, Faisal Khan, Salim Ahmed, and Syed Imtiaz. Autonomous fault diagnosis and root cause analysis for the processing system using one-class svm and nn permutation algorithm. *Industrial & Engineering Chemistry Research*, 61(3) :1408–1422, 2022.

[8] Rajeevan Arunthavanathan, Faisal Khan, Salim Ahmed, Syed Imtiaz, and Risza Rusli. Fault detection and diagnosis in process system using artificial intelligence-based cognitive technique. *Computers & Chemical Engineering*, 134 :106697, 2020.

[9] Narayanaswamy Balakrishnan and Valery B Nevzorov. *A primer on statistical distributions*. John Wiley & Sons, 2004.

[10] Bing-Kun Bao, Guangcan Liu, Changsheng Xu, and Shuicheng Yan. Inductive robust principal component analysis. *IEEE transactions on image processing*, 21(8) :3794–3800, 2012.

[11] Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2) :218–233, 2003.

[12] Michele Basseville. Distance measures for signal processing and pattern recognition. *Signal processing*, 18(4) :349–369, 1989.

[13] Michele Basseville and Igor V Nikiforov. *Detection of abrupt changes : theory and application*, volume 104. Prentice hall Englewood Cliffs, 1993.

[14] Stephen Becker, Candes Emmanuel, and C.Grant Michael. Tfocs : flexible first-order methods for rank minimization. In *Low-rank Matrix Optimization Symposium, SIAM Conference on Optimization*, 2011.

[15] Tajeddine Benbarrad, Marouane Salhaoui, Soukaina Bakhat Kenitar, and Mounir Arioua. Intelligent machine vision model for defective product inspection based on machine learning. *Journal of Sensor and Actuator Networks*, 10(1) :7, 2021.

[16] Vance W Berger Berger and YanYan Zhou. Kolmogorov–smirnov test : Overview. *Wiley statsref : Statistics reference online*, 2014.

[17] HanQin Cai, Jian-Feng Cai, and Ke Wei. Accelerated alternating projections for robust principal component analysis. *The Journal of Machine Learning Research*, 20(1) :685–717, 2019.

[18] HanQin Cai, Keaton Hamm, Longxiu Huang, Jiaqi Li, and Tao Wang. Rapid robust principal component analysis : Cur accelerated inexact low rank estimation. *IEEE Signal Processing Letters*, 28 :116–120, 2020.

[19] Emmanuel Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis ? *Journal of the ACM (JACM)*, 58(3) :1–37, 2011.

[20] Mikel Canizo, Isaac Triguero, Angel Conde, and Enrique Onieva. Multi-head cnn–rnn for multi-time series anomaly detection : An industrial case study. *Neurocomputing*, 363 :246–260, 2019.

[21] Yi Chai, Songbing Tao, Wanbiao Mao, Ke Zhang, and Zhiqin Zhu. Online incipient fault diagnosis based on kullback-leibler divergence and recursive principle component analysis. *The Canadian Journal of Chemical Engineering*, 96(2) :426–433, 2018.

[22] Bin Chen, Zhiwu Huang, Rui Zhang, Fu Jiang, Weirong Liu, Heng Li, Jing Wang, and Jun Peng. Adaptive slip ratio estimation for active braking control of high-speed trains. *ISA transactions*, 112 :302–314, 2021.

[23] Hongtian Chen, Bin Jiang, and Ningyun Lu. Data-driven incipient sensor fault estimation with application in inverter of high-speed railway. *Mathematical Problems in Engineering*, 2017, 2017.

[24] Yuejian Chen, Xihui Liang, and Mingjian Zuo. An improved singular value decomposition-based method for gear tooth crack detection and severity assessment. *Journal of Sound and Vibration*, 468 :115068, 2020.

[25] Zhaomin Chen, Chai Kiat Yeo, Bu Sung Lee, and Chiew Tong Lau. Autoencoder-based network anomaly detection. In *2018 Wireless telecommunications symposium (WTS)*, pages 1–5. IEEE, 2018.

[26] Zhiwen Chen, Yue Cao, Steven X Ding, Kai Zhang, Tim Koenings, Tao Peng, Chunhua Yang, and Weihua Gui. A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring. *IEEE Transactions on Industrial Informatics*, 15(5) :2710–2720, 2019.

[27] Zhiwen Chen, Steven X Ding, Tao Peng, Chunhua Yang, and Weihua Gui. Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms. *IEEE Transactions on Industrial Electronics*, 65(2) :1559–1567, 2017.

[28] Zhiwen Chen, Steven X Ding, Kai Zhang, Zhebin Li, and Zhikun Hu. Canonical correlation analysis-based fault detection methods with application to alumina evaporation process. *Control Engineering Practice*, 46 :51–58, 2016.

[29] Zhongxin Chen, Feng Zhao, Jun Zhou, Panling Huang, and Xutao Zhang. Fault diagnosis of loader gearbox based on an ica and svm algorithm. *International Journal of Environmental Research and Public Health*, 16(23) :4868, 2019.

[30] Zhuyun Chen, Alexandre Mauricio, Weihua Li, and Konstantinos Gryllias. A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, 140 :106683, 2020.

[31] Cheng Cheng, Beitong Zhou, Guijun Ma, Dongrui Wu, and Ye Yuan. Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data. *Neurocomputing*, 409 :35–45, 2020.

[32] Sang Wook Choi, Changkyu Lee, Jong-Min Lee, Jin Hyun Park, and In-Beum Lee. Fault detection and identification of nonlinear processes based on kernel pca. *Chemometrics and intelligent laboratory systems*, 75(1) :55–67, 2005.

[33] Marco Cocconcelli, Radoslaw Zimroz, Riccardo Rubini, and Walter Bartelmus. Stft based approach for ball bearing fault detection in a varying speed motor. In *Condition monitoring of machinery in non-stationary operations*, pages 41–50. Springer, 2012.

[34] Peiling Cui, Junhong Li, and Guizeng Wang. Improved kernel principal component analysis for fault detection. *Expert Systems with Applications*, 34(2) :1210–1219, 2008.

[35] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4) :303–314, 1989.

[36] Ido Dagan, Lillian Lee, and Fernando Pereira. Similarity-based methods for word sense disambiguation. *arXiv preprint cmp-lg/9708010*, 1997.

[37] Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1) :1–18, 2000.

[38] Claude Delpha and Demba Diallo. Incipient fault detection and diagnosis : a hidden information detection problem. In *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, pages 837–842. IEEE, 2015.

[39] Claude Delpha, Demba Diallo, Hanane Al Samrout, and Nazih Moubayed. Multiple incipient fault diagnosis in three-phase electrical systems using multivariate statistical signal processing. *Engineering Applications of Artificial Intelligence*, 73 :68–79, 2018.

[40] Claude Delpha, Demba Diallo, Tianzhen Wang, Jie Liu, and Zelig Li. Multisensor fault detection and isolation using kullback leibler divergence : Application to data vibration signals. In *2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, pages 305–310. IEEE, 2017.

[41] Claude Delpha, Demba Diallo, and Abdulrahman Youssef. Kullback-Leibler divergence for fault estimation and isolation : Application to Gamma distributed data. *Mechanical Systems and Signal Processing*, 93 :118–135, 2017.

[42] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv :1810.08693*, 2018.

[43] Xinghao Ding, Lihan He, and Lawrence Carin. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 20(12) :3419–3430, 2011.

[44] Xiaoyu Dong. Wenzhou train crash pr/social media case study. 2014.

[45] Yining Dong and S Joe Qin. A novel dynamic pca algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67 :1–11, 2018.

[46] DA Dornfeld, Y Lee, and A Chang. Monitoring of ultraprecision machining processes. *The International Journal of Advanced Manufacturing Technology*, 21(8) :571–578, 2003.

[47] Yuncheng Du and Dongping Du. Fault detection and diagnosis using empirical mode decomposition based principal component analysis. *Computers & Chemical Engineering*, 115 :1–21, 2018.

[48] Cao Vu Dung et al. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 99 :52–58, 2019.

[49] Lamiaa M Elshenawy and Tarek A Mahmoud. Fault diagnosis of time-varying processes using modified reconstruction-based contributions. *Journal of Process Control*, 70 :12–23, 2018.

[50] Shu-Kai S Fan, Chia-Yu Hsu, Chih-Hung Jen, Kuan-Lung Chen, and Li-Ting Juan. Defective wafer detection using a denoising autoencoder for semiconductor manufacturing processes. *Advanced Engineering Informatics*, 46 :101166, 2020.

[51] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8) :861–874, 2006.

[52] Diego Fernández-Francos, David Martínez-Rego, Oscar Fontenla-Romero, and Amparo Alonso-Betanzos. Automatic bearing fault diagnosis based on one-class ν-svm. *Computers & Industrial Engineering*, 64(1) :357–365, 2013.

[53] Tony Fishwick. Rupture of a heat exchanger at a refinery causes fatalities. *Loss Prevention Bulletin*, 228, 2012.

[54] Zhiwei Gao, Sing Kiong Nguang, and De-Xing Kong. Advances in modelling, monitoring, and control for complex industrial systems, 2019.

[55] Suryakant Gautam, Prakash K Tamboli, Kallol Roy, Vaibhav H Patankar, and Siddhartha P Duttagupta. Sensors incipient fault detection and isolation of nuclear power plant using extended kalman filter and kullback–leibler divergence. *ISA transactions*, 92 :180–190, 2019.

[56] Janos Gertler and Jin Cao. Pca-based fault diagnosis in the presence of control and dynamics. *AIChE Journal*, 50(2) :388–402, 2004.

[57] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[58] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[59] Vasyl Golosnoy, Sergiy Ragulin, and Wolfgang Schmid. Multivariate cusum chart : properties and enhancements. *AStA Advances in Statistical Analysis*, 93(3) :263–279, 2009.

[60] Peng Guo, Jian Fu, and XiYun Yang. Condition monitoring and fault diagnosis of wind turbines gearbox bearing temperature based on kolmogorov-smirnov test and convolutional neural network model. *Energies*, 11(9) :2248, 2018.

[61] Zhen Guo, Guofei Jiang, Haifeng Chen, and Kenji Yoshihira. Tracking probabilistic correlation of monitoring data for fault detection in complex systems. In *International Conference on Dependable Systems and Networks (DSN'06)*, pages 259–268. IEEE, 2006.

[62] Pankaj Gupta and MK Pradhan. Fault detection analysis in rolling element bearing : A review. *Materials Today : Proceedings*, 4(2) :2085–2094, 2017.

[63] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah. Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 2805–2808. IEEE, 2012.

[64] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah. Moving object detection via robust low rank matrix decomposition with irls scheme. In *International symposium on visual computing*, pages 665–674. Springer, 2012.

[65] Sahand Hariri, Matias Carrasco Kind, and Robert J Brunner. Extended isolation forest. *IEEE Transactions on Knowledge and Data Engineering*, 33(4) :1479–1489, 2019.

[66] Jinane Harmouche, Claude Delpha, and Demba Diallo. Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis : Part I. *Signal processing*, 94 :278–287, 2014.

[67] Jinane Harmouche, Claude Delpha, and Demba Diallo. Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis : Part ii. *Signal Processing*, 109 :334–344, 2015.

[68] Jinane Harmouche, Claude Delpha, and Demba Diallo. Incipient fault amplitude estimation using KL divergence with a probabilistic approach. *Signal Processing*, 120 :1–7, 2016.

[69] Jinane Harmouche, Claude Delpha, Demba Diallo, and Yann Le Bihan. Statistical approach for nondestructive incipient crack detection and characterization using kullback-leibler divergence. *IEEE Transactions on Reliability*, 65(3) :1360–1368, 2016.

[70] John A Hartigan and Manchek A Wong. Algorithm as 136 : A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1) :100–108, 1979.

[71] Ran He, Bao-Gang Hu, Wei-Shi Zheng, and Xiang-Wei Kong. Robust principal component analysis based on maximum correntropy criterion. *IEEE Transactions on Image Processing*, 20(6) :1485–1494, 2011.

[72] Yan-Lin He, Yang Zhao, Qun-Xiong Zhu, and Yuan Xu. Online distributed process monitoring and alarm analysis using novel canonical variate analysis with multicorrelation blocks and enhanced contribution plot. *Industrial & Engineering Chemistry Research*, 59(45) :20045–20057, 2020.

[73] Ernst Hellinger. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136) :210–271, 1909.

[74] Heiko Hoffmann. Kernel pca for novelty detection. *Pattern recognition*, 40(3) :863–874, 2007.

[75] Agnar Höskuldsson. Pls regression methods. *Journal of chemometrics*, 2(3) :211–228, 1988.

[76] Yung-An Hsieh and Yichang James Tsai. Machine learning for crack detection : Review and model performance comparison. *Journal of Computing in Civil Engineering*, 34(5) :04020038, 2020.

[77] Jian Huang and Xuefeng Yan. Dynamic process fault detection and diagnosis based on dynamic principal component analysis, dynamic independent component analysis and bayesian inference. *Chemometrics and Intelligent Laboratory Systems*, 148 :115–127, 2015.

[78] Nantian Huang, Huaijin Chen, Shuxin Zhang, Guowei Cai, Weiguo Li, Dianguo Xu, and Lihua Fang. Mechanical fault diagnosis of high voltage circuit breakers based on wavelet time-frequency entropy and one-class support vector machine. *Entropy*, 18(1) :7, 2015.

[79] Aapo Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3) :626–634, 1999.

[80] Aapo Hyvärinen and Erkki Oja. Independent component analysis : algorithms and applications. *Neural networks*, 13(4-5) :411–430, 2000.

[81] Rolf Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, 29(1) :71–85, 2005.

[82] J Edward Jackson and Govind S Mudholkar. Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3) :341–349, 1979.

[83] Hongquan Ji, Xiao He, Jun Shang, and Donghua Zhou. Incipient sensor fault diagnosis using moving window reconstruction-based contribution. *Industrial & Engineering Chemistry Research*, 55(10) :2746–2759, 2016.

[84] Hongquan Ji, Xiao He, Jun Shang, and Donghua Zhou. Exponential smoothing reconstruction approach for incipient fault isolation. *Industrial & Engineering Chemistry Research*, 57(18) :6353–6363, 2018.

[85] Hongquan Ji, Keke Huang, and Donghua Zhou. Incipient sensor fault isolation based on augmented Mahalanobis distance. *Control Engineering Practice*, 86 :144–154, 2019.

[86] Dongnian Jiang, Wei Li, and Fuyuan Shen. Incipient fault diagnosis and amplitude estimation based on k–l divergence with a gaussian mixture model. *Review of Scientific Instruments*, 91(5) :055103, 2020.

[87] Jiuchun Jiang, Taiyu Li, Chun Chang, Chen Yang, and Li Liao. Fault diagnosis method for lithium-ion batteries in electric vehicles based on isolated forest algorithm. *Journal of Energy Storage*, 50 :104177, 2022.

[88] Qingchao Jiang, Xuefeng Yan, and Biao Huang. Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes. *Industrial & Engineering Chemistry Research*, 58(29) :12899–12912, 2019.

[89] S Joe Qin. Statistical process monitoring : basics and beyond. *Journal of Chemometrics : A Journal of the Chemometrics Society*, 17(8-9) :480–502, 2003.

[90] Majid Karami and Liping Wang. Fault detection and diagnosis for nonlinear systems : A new adaptive gaussian mixture modeling approach. *Energy and Buildings*, 166 :477–488, 2018.

[91] Nicolas Keriven, Damien Garreau, and Iacopo Poli. Newma : a new method for scalable model-free online change-point detection. *IEEE Transactions on Signal Processing*, 68 :3515–3528, 2020.

[92] Aftab A Khan, James R Moyne, and Dawn M Tilbury. Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares. *Journal of Process Control*, 18(10) :961–974, 2008.

[93] Wael Khreich, Babak Khosravifar, Abdelwahab Hamou-Lhadj, and Chamseddine Talhi. An anomaly detection system based on variable n-gram features and one-class svm. *Information and Software Technology*, 91 :186–197, 2017.

[94] Jonghyuk Kim, Hyunwoo Hwangbo, and Soyean Kim. An empirical study on real-time data analytics for connected cars : Sensor-based applications for smart cars. *International Journal of Distributed Sensor Networks*, 14(1) :1550147718755290, 2018.

[95] Samuel Kotz and Saralees Nadarajah. *Extreme value distributions : theory and applications*. World Scientific, 2000.

[96] Theodora Kourti, Paul Nomikos, and John F MacGregor. Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway pls. *Journal of process control*, 5(4) :277–284, 1995.

[97] Wenfu Ku, Robert H Storer, and Christos Georgakis. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1) :179–196, 1995.

[98] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1) :79–86, 1951.

[99] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1) :112–147, 1998.

[100] Wallace E Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *29th IEEE Conference on Decision and control*, pages 596–604. IEEE, 1990.

[101] Wallace E Larimore. Optimal reduced rank modeling, prediction, monitoring and control using canonical variate analysis. *IFAC Proceedings Volumes*, 30(9) :61–66, 1997.

[102] Yann Le Bihan, József Pávó, and Claude Marchand. Study and experimental validation of the calculation of the ect signal induced by a minute crack using a fem–bim combination. *NDT & E International*, 39(6) :476–486, 2006.

[103] Gibaek Lee, Chonghun Han, and En Sup Yoon. Multiple-fault diagnosis of the tennessee eastman process based on system decomposition and dynamic pls. *Industrial & engineering chemistry research*, 43(25) :8037–8048, 2004.

[104] Jong-Min Lee, S Joe Qin, and In-Beum Lee. Fault detection of non-linear processes using kernel independent component analysis. *The Canadian Journal of Chemical Engineering*, 85(4) :526–536, 2007.

[105] Yaguo Lei, Feng Jia, Jing Lin, Saibo Xing, and Steven X Ding. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5) :3137–3147, 2016.

[106] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, and Asoke K Nandi. Applications of machine learning to machine fault diagnosis : A review and roadmap. *Mechanical Systems and Signal Processing*, 138 :106587, 2020.

[107] Elizaveta Levina and Peter Bickel. The earth mover's distance is the mallows distance : Some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 251–256. IEEE, 2001.

[108] Si-hui Li, Bai-gen Cai, Jiang Liu, and Jian Wang. Collision risk analysis based train collision early warning strategy. *Accident Analysis & Prevention*, 112 :94–104, 2018.

[109] Xiang Li, Wei Zhang, and Qian Ding. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal processing*, 161 :136–154, 2019.

[110] Xiaochuan Li, Xiaoyu Yang, Yingjie Yang, Ian Bennett, and David Mba. A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines. *Applied Soft Computing*, 82 :105564, 2019.

[111] Yongbo Li, Xianzhi Wang, Shubin Si, and Shiqian Huang. Entropy based fault classification using the case western reserve university data : A benchmark study. *IEEE Transactions on Reliability*, 69(2) :754–767, 2019.

[112] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv :1009.5055*, 2010.

[113] Haibin Ling and Kazunori Okada. An efficient earth mover's distance algorithm for robust histogram comparison. *IEEE transactions on pattern analysis and machine intelligence*, 29(5) :840–853, 2007.

[114] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008.

[115] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1) :1–39, 2012.

[116] Jie Liu, Yan-Fu Li, and Enrico Zio. A SVM framework for fault detection of the braking system in a high speed train. *Mechanical Systems and Signal Processing*, 87 :401–409, 2017.

[117] Yi Liu, Jiusun Zeng, Lei Xie, Shihua Luo, and Hongye Su. Structured joint sparse principal component analysis for fault detection and isolation. *IEEE Transactions on Industrial Informatics*, 15(5) :2721–2731, 2018.

[118] Kenneth Loparo. Case western reserve university bearing data center. *Bearings Vibration Data Sets, Case Western Reserve University*, pages 22–28, 2012.

[119] Raul HC Lopes, ID Reid, and Peter R Hobson. The two-dimensional kolmogorov-smirnov test. *Proceedings of Science*, 2007.

[120] Gary Lorden et al. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6) :1897–1908, 1971.

[121] Zhijiang Lou, Youqing Wang, Shan Lu, and Pei Sun. Process monitoring using a novel robust pca scheme. *Industrial & Engineering Chemistry Research*, 60(11) :4397–4404, 2021.

[122] Cynthia A Lowry, William H Woodall, Charles W Champ, and Steven E Rigdon. A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1) :46–53, 1992.

[123] Lijia Luo, Shiyi Bao, and Chudong Tong. Sparse robust principal component analysis with applications to fault detection and diagnosis. *Industrial & Engineering Chemistry Research*, 58(3) :1300–1309, 2019.

[124] Majdi Mansouri, Mohamed Nounou, Hazem Nounou, and Nazmul Karim. Kernel pca-based glrt for nonlinear fault detection of chemical processes. *Journal of Loss Prevention in the Process Industries*, 40 :334–347, 2016.

[125] Majdi Mansouri, Mohamed Numan Nounou, and Hazem Numan Nounou. Multiscale kernel PLS-based exponentially weighted-GLRT and its application to fault detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(1) :49–58, 2017.

[126] Wentao Mao, Jianliang He, Yuan Li, and Yunju Yan. Bearing fault diagnosis with auto-encoder extreme learning machine : A comparative study. *Proceedings of the Institution of Mechanical Engineers, Part C : Journal of Mechanical Engineering Science*, 231(8) :1560–1578, 2017.

[127] Elaine B. Martin and A. Julian Morris. Non-parametric confidence bounds for process performance monitoring charts. *Journal of process control*, 6(6) :349–358, 1996.

[128] Luis Mendonça, Joao Miguel C. Sousa, and José Sá da Costa. An architecture for fault detection and isolation based on fuzzy methods. *Expert Systems with Applications*, 36(2, Part 1) :1092–1104, 2009.

[129] Zong Meng, Xuyang Zhan, Jing Li, and Zuozhou Pan. An enhancement denoising autoencoder for rolling bearing fault diagnosis. *Measurement*, 130 :448–454, 2018.

[130] Baligh Mnassri, Mustapha Ouladsine, et al. Inverse-variance weighting pca-based vre criterion to select the optimal number of pcs. *IFAC Proceedings Volumes*, 44(1) :2851–2856, 2011.

[131] Syrine Neffati, Khaoula Ben Abdellafou, Okba Taouali, and Kais Bouzrara. Enhanced svm–kpca method for brain mr image classification. *The Computer Journal*, 63(3) :383–394, 2020.

[132] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4) :308–313, 1965.

[133] Praneeth Netrapalli, Niranjan UN, Sujay Sanghavi, Animashree Anandkumar, and Prateek Jain. Non-convex robust pca. *Advances in Neural Information Processing Systems*, 27, 2014.

[134] Dhiraj Neupane and Jongwon Seok. Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches : A review. *IEEE Access*, 8 :93155–93178, 2020.

[135] Viethung Nguyen, Danwei Wang, Jeevanand Seshadrinath, Abhisek Ukil, Merugu Sivarama Krishna, Sivakumar Nadarajan, and Viswanathan Vaiyapuri. A method for incipient interturn fault detection and severity estimation of induction motors under inherent asymmetry and voltage imbalance. *IEEE Transactions on Transportation Electrification*, 3(3) :703–715, 2017.

[136] Frank Nielsen. On a variational definition for the jensen-shannon symmetrization of distances based on the information radius. *Entropy*, 23(4) :464, 2021.

[137] Yichun Niu, Li Sheng, Ming Gao, and Donghua Zhou. Distributed intermittent fault detection for linear stochastic systems over sensor network. *IEEE Transactions on Cybernetics*, 2021.

[138] Zineb Noumir, Paul Honeine, and Cedue Richard. On simple one-class classification methods. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 2022–2026. IEEE, 2012.

[139] Pabara-Ebiere Patricia Odiowei and Yi Cao. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Transactions on Industrial Informatics*, 6(1) :36–45, 2009.

[140] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2) :100–115, 1954.

[141] Karl Ezra Pilario. Feedback-controlled CSTR process for fault simulation. `https://www.mathworks.com/matlabcentral/fileexchange/66189-feedback-controlled-cstr-process-for-fault-simulation`, 2021. Retrieved October 25, 2021.

[142] Karl Ezra Salgado Pilario and Yi Cao. Canonical variate dissimilarity analysis for process incipient fault detection. *IEEE Transactions on Industrial Informatics*, 14(12) :5308–5315, 2018.

[143] Robin L Plackett. Karl pearson and the chi-squared test. *International statistical review/revue internationale de statistique*, pages 59–72, 1983.

[144] IE Prank and JH Friedman. A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, 35 :109–135, 1993.

[145] S Joe Qin and Thomas J McAvoy. Nonlinear pls modeling using neural networks. *Computers & Chemical Engineering*, 16(4) :379–391, 1992.

[146] Ashikur Rahaman, Md Milon Islam, Md Rashedul Islam, Muhammad Sheikh Sadi, and Sheikh Nooruddin. Developing iot based smart health monitoring systems : A review. *Rev. d'Intelligence Artif.*, 33(6) :435–440, 2019.

[147] Prahalad Rao, Satish Bukkapatnam, Omer Beyca, Zhenyu James Kong, and Ranga Komanduri. Real-time identification of incipient surface morphology variations in ultraprecision machining process. *Journal of Manufacturing Science and Engineering*, 136(2), 2014.

[148] Tiago J Rato and Marco S Reis. Fault detection in the tennessee eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (dpca-dr). *Chemometrics and Intelligent Laboratory Systems*, 125 :101–108, 2013.

[149] SW Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 42(1) :97–101, 2000.

[150] Paul Rodriguez and Brendt Wohlberg. Fast principal component pursuit via alternating minimization. In *2013 IEEE International Conference on Image Processing*, pages 69–73. IEEE, 2013.

[151] Peter J Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3) :212–223, 1999.

[152] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.

[153] Hossein Safaeipour, Mehdi Forouzanfar, and Alessandro Casavola. A survey and classification of incipient fault diagnosis approaches. *Journal of Process Control*, 97 :1–16, 2021.

[154] Raphael T Samuel and Yi Cao. Kernel canonical variate analysis for nonlinear dynamic process monitoring. *IFAC-PapersOnLine*, 48(8) :605–610, 2015.

[155] Raphael Tari Samuel and Yi Cao. Nonlinear process fault detection and identification using kernel pca and kernel density estimation. *Systems Science & Control Engineering*, 4(1) :165–174, 2016.

[156] Makhambet Sarbayev, Ming Yang, and Haiqing Wang. Risk assessment of process systems by mapping fault tree into artificial neural network. *Journal of Loss Prevention in the Process Industries*, 60 :203–212, 2019.

[157] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7) :1443–1471, 2001.

[158] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5) :1299–1319, 1998.

[159] Fanhua Shang, Yuanyuan Liu, James Cheng, and Hong Cheng. Robust principal component analysis with missing data. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1149–1158, 2014.

[160] Haidong Shao, Hongkai Jiang, Huiwei Zhao, and Fuan Wang. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 95 :187–204, 2017.

[161] Walter Andrew Shewhart. *Economic control of quality of manufactured product*. Macmillan And Co Ltd, London, 1931.

[162] Wade A Smith and Robert B Randall. Rolling element bearing diagnostics using the case western reserve university data : A benchmark study. *Mechanical systems and signal processing*, 64 :100–131, 2015.

[163] WGRFR Spendley, George R Hext, and Francis R Himsworth. Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*, 4(4) :441–461, 1962.

[164] Weixiang Sun, Jin Chen, and Jiaqing Li. Decision tree and pca-based fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 21(3) :1300–1317, 2007.

[165] Takashi Takeuchi. Oxygen sensors. *Sensors and Actuators*, 14(2) :109–124, 1988.

[166] Masayuki Tamura and Shinsuke Tsujita. A study on the number of principal components and sensitivity of fault detection using pca. *Computers & Chemical Engineering*, 31(9) :1035–1046, 2007.

[167] Gongguo Tang and Arye Nehorai. Robust principal component analysis based on low-rank and block-sparse matrix decomposition. In *2011 45th Annual Conference on Information Sciences and Systems*, pages 1–5. IEEE, 2011.

[168] Liang Tao, Qian Siqi, Yingjuan Zhang, and Huan Shi. Abnormal detection of wind turbine based on scada data mining. *Mathematical Problems in Engineering*, 2019, 2019.

[169] Yang Tao, Hongbo Shi, Bing Song, and Shuai Tan. Parallel quality-related dynamic principal component regression method for chemical process monitoring. *Journal of Process Control*, 73 :33–45, 2019.

[170] David Martinus Johannes Tax. *One-class classification : Concept learning in the absence of counter-examples*. Print partners Ipskamp, 2002.

[171] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1) :45–66, 2004.

[172] Yvon Tharrault, Gilles Mourot, and José Ragot. Fault detection and isolation with robust principal component analysis. In *2008 16th Mediterranean Conference on Control and Automation*, pages 59–64. IEEE, 2008.

[173] Michael Tschannen, Olivier Bachem, and Mario Lucic. Recent advances in autoencoder-based representation learning. *arXiv preprint arXiv :1812.05069*, 2018.

[174] Venkat Venkatasubramanian. Prognostic and diagnostic monitoring of complex systems for product lifecycle management : Challenges and opportunities. *Computers & chemical engineering*, 29(6) :1253–1263, 2005.

[175] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N Kavuri, and Kewen Yin. A review of process fault detection and diagnosis : Part III : Process history based methods. *Computers & chemical engineering*, 27(3) :327–346, 2003.

[176] Biao Wang, Yaguo Lei, Naipeng Li, and Ningbo Li. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability*, 69(1) :401–412, 2018.

[177] Haiqing Wang, Zhihuan Song, and Ping Li. Fault detection behavior and performance analysis of principal component analysis based process monitoring methods. *Industrial & Engineering Chemistry Research*, 41(10) :2455–2464, 2002.

[178] Xiyang Wang and Viliam Makis. Autoregressive model-based gear shaft fault diagnosis using the kolmogorov–smirnov test. *Journal of Sound and Vibration*, 327(3-5) :413–423, 2009.

[179] Yalin Wang, Zhuofu Pan, Xiaofeng Yuan, Chunhua Yang, and Weihua Gui. A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA transactions*, 96 :457–467, 2020.

[180] Long Wen, Liang Gao, and Xinyu Li. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on systems, man, and cybernetics : systems*, 49(1) :136–144, 2017.

[181] Johan A Westerhuis, Stephen P Gurden, and Age K Smilde. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and intelligent laboratory systems*, 51(1) :95–114, 2000.

[182] Brendt Wohlberg, Rick Chartrand, and James Theiler. Local principal component pursuit for nonlinear datasets. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3925–3928. IEEE, 2012.

[183] Svante Wold, Nouna Kettaneh-Wold, and Bert Skagerberg. Nonlinear pls modeling. *Chemometrics and intelligent laboratory systems*, 7(1-2) :53–65, 1989.

[184] Svante Wold, Michael Sjöström, and Lennart Eriksson. Pls-regression : a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2) :109–130, 2001.

[185] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis : Exact recovery of corrupted low-rank matrices via convex optimization. *Advances in neural information processing systems*, 22, 2009.

[186] Yunkai Wu, Bin Jiang, Ningyun Lu, and Donghua Zhou. Tomfir-based incipient fault detection and estimation for high-speed rail vehicle suspension system. *Journal of the Franklin Institute*, 352(4) :1672–1692, 2015.

[187] Xiaochen Xie, Wei Sun, and Kie Chung Cheung. An advanced pls approach for key performance indicator-related prediction and diagnosis in case of outliers. *IEEE Transactions on Industrial Electronics*, 63(4) :2587–2594, 2015.

[188] Xianchao Xiu, Ying Yang, Lingchen Kong, and Wanquan Liu. Laplacian regularized robust principal component analysis for process monitoring. *Journal of Process Control*, 92 :212–219, 2020.

[189] Fan Xu, Zhou Fang, Ruoli Tang, Xin Li, and Kwok Leung Tsui. An unsupervised and enhanced deep belief network for bearing performance degradation assessment. *Measurement*, 162 :107902, 2020.

[190] Xiaoan Yan and Minping Jia. A novel optimized svm classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing*, 313 :47–64, 2018.

[191] Fan Yang, Lei Zhang, Sijia Yu, Danil Prokhorov, Xue Mei, and Haibin Ling. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*, 21(4) :1525–1535, 2019.

[192] Howard Hua Yang, Shun-ichi Amari, and Andrzej Cichocki. Information–theoretic approach to blind separation of sources in non-linear mixture. *Signal processing*, 64(3) :291–300, 1998.

[193] Junjie Yang and Claude Delpha. Open-circuit fault diagnosis for interleaved DC-DC converters. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 3982–3987. IEEE, 2020.

[194] Junjie Yang and Claude Delpha. A local mahalanobis distance analysis based methodology for incipient fault diagnosis. In *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 1–8. IEEE, 2021.

[195] Junjie Yang and Claude Delpha. Local mahalanobis distance envelope using a robust healthy domain approximation for incipient fault diagnosis. In *IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–6, 2021.

[196] Yunguang Ye, Yongxiang Zhang, Qingbo Wang, Zhiwei Wang, Zhenjie Teng, and Hougui Zhang. Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment. *Mechanical Systems and Signal Processing*, 138 :106565, 2020.

[197] Zhang Yi, Wei Xueye, and Jiang Haifeng. One-class classifier based on sbt for analog circuit fault diagnosis. *Measurement*, 41(4) :371–380, 2008.

[198] Shen Yin, Steven X Ding, Adel Haghani, Haiyang Hao, and Ping Zhang. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman Process. *Journal of process control*, 22(9) :1567–1581, 2012.

[199] Shen Yin, Steven X Ding, Xiaochen Xie, and Hao Luo. A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 61(11) :6418–6428, 2014.

[200] Abdulrahman Youssef, Claude Delpha, and Demba Diallo. An optimal fault detection threshold for early detection using Kullback–Leibler divergence for unknown distribution data. *Signal Processing*, 120 :266–279, 2016.

[201] Alexander Ypma and Robert PW Duin. Support objects for domain approximation. In *International Conference on Artificial Neural Networks*, pages 719–724. Springer, 1998.

[202] Jianbo Yu, Xing Liu, and Lyujiangnan Ye. Convolutional long short-term memory autoencoder-based feature learning for fault detection in industrial processes. *IEEE Transactions on Instrumentation and Measurement*, 70 :1–15, 2021.

[203] Jie Yu. A nonlinear kernel gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chemical Engineering Science*, 68(1) :506–519, 2012.

[204] Ming Yu, Haotian Lu, Hai Wang, Chenyu Xiao, Dun Lan, and Junjie Chen. Computational intelligence-based prognosis for hybrid mechatronic system using improved wiener process. In *Actuators*, volume 10, page 213. Multidisciplinary Digital Publishing Institute, 2021.

[205] Yaoxiang Yu, Liang Guo, Hongli Gao, Yuekai Liu, and Tingting Feng. Pareto-optimal adaptive loss residual shrinkage network for imbalanced fault diagnostics of machines. *IEEE Transactions on Industrial Informatics*, 18(4) :2233–2243, 2021.

[206] Xiaoming Yuan and Junfeng Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *preprint*, 12(2), 2009.

[207] H Henry Yue and S Joe Qin. Reconstruction-based fault identification using a combined index. *Industrial & engineering chemistry research*, 40(20) :4403–4414, 2001.

[208] Jiusun Zeng, Uwe Kruger, Jaap Geluk, Xun Wang, and Lei Xie. Detecting abnormal situations using the Kullback–Leibler divergence. *Automatica*, 50(11) :2777–2786, 2014.

[209] Yuke Zeng, Huanxin Chen, Chengliang Xu, Yahao Cheng, and Qijian Gong. A hybrid deep forest approach for outlier detection and fault diagnosis of variable refrigerant flow system. *International Journal of Refrigeration*, 120 :104–118, 2020.

[210] Cheng Zhang, Qingxiu Guo, and Yuan Li. Fault detection method based on principal component difference associated with dpca. *Journal of Chemometrics*, 33(1) :e3082, 2019.

[211] Cheng Zhang, Qingxiu Guo, and Yuan Li. Fault detection in the tennessee eastman benchmark process using principal component difference based on k-nearest neighbors. *IEEE Access*, 8 :49999–50009, 2020.

[212] Jiayuan Zhang, Wei Zhan, and Mehrdad Ehsani. Fault-tolerant control of PMSM with inter-turn short-circuit fault. *IEEE Transactions on Energy Conversion*, 34(4) :2267–2275, 2019.

[213] Kangkang Zhang, Bin Jiang, Xing-Gang Yan, and Zehui Mao. Incipient voltage sensor fault isolation for rectifier in railway electrical traction systems. *IEEE Transactions on Industrial Electronics*, 64(8) :6763–6774, 2017.

[214] Ming Zhang, Zhinong Jiang, and Kun Feng. Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. *Mechanical Systems and Signal Processing*, 93 :460–493, 2017.

[215] Xiaoxia Zhang and Claude Delpha. Detection capability for incipient faults in a noisy environment using pdf and cdf based techniques : a comparative study. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 2499–2504. IEEE, 2020.

[216] Xiaoxia Zhang, Claude Delpha, and Demba Diallo. Incipient fault detection and estimation based on Jensen–Shannon divergence in a data-driven approach. *Signal Processing*, 169 :107410, 2020.

[217] Xiaoxia Zhang, Claude Delpha, and Demba Diallo. Jensen-shannon divergence for non-destructive incipient crack detection and estimation. *IEEE Access*, 8 :116148–116162, 2020.

[218] Yingwei Zhang, Shuai Li, and Zhiyong Hu. Improved multi-scale kernel principal component analysis and its application for fault detection. *Chemical Engineering Research and Design*, 90(9) :1271–1280, 2012.

[219] Song Zhanwei and Liu Zenghui. Abnormal detection method of industrial control system based on behavior model. *Computers & Security*, 84 :166–178, 2019.

[220] Chunhui Zhao and Furong Gao. A sparse dissimilarity analysis algorithm for incipient fault isolation with no priori fault information. *Control Engineering Practice*, 65 :70–82, 2017.

[221] Ting Zhong, Jianfeng Qu, Xiaoyu Fang, Hao Li, and Zeping Wang. The intermittent fault diagnosis of analog circuits based on eemd-dbn. *Neurocomputing*, 436 :74–91, 2021.

[222] Zhiyu Zhu, Lanzhi Wang, Gaoliang Peng, and Sijue Li. Wda : an improved wasserstein distance-based transfer learning fault diagnosis method. *Sensors*, 21(13) :4394, 2021.

[223] Eric R Ziegel. The elements of statistical learning. *Technometrics*, 45(3) :267–268, 2003.

[224] Qin Zou, Zheng Zhang, Qingquan Li, Xianbiao Qi, Qian Wang, and Song Wang. Deepcrack : Learning hierarchical convolutional features for crack detection. *IEEE Transactions on Image Processing*, 28(3) :1498–1512, 2018.