Outlier-Tolerant Identification and Abrupt Change Detection for Stationary Multi-dimensional Process

Hu Jingzhi

Abstract—Covariance matrix and location vector \((\Sigma, \mu)\) are widely used to describe the multi-dimensional stationary process. A lot of practice in data analysis show that classical approaches, such as the mean estimator, the least squared estimator and the maximum likelihood estimator, are sensitive to outliers and impurities in sampling series. In other words, these classical estimators stated above will possibly break down if there exists a few outliers in sampling series, which may result in incorrect evaluation and improper judgment on states of the processes. Therefore, this paper adopts methods such as the robust statistical and outliers-tolerant theories and techniques to build a series of new optimal outliers-tolerant algorithms to estimate the parameters \((\Sigma, \mu)\) of stationary multi-dimensional processes. These new estimation algorithms are not only affine equivariant with data transformation but also are well tolerant to outliers and impurities in sampling data series. Based on these new estimation algorithms, a series of detecting methods are proposed in this paper to detect as well as to monitor abnormal changes in process. Simulation results given in this paper validate the efficiency and outliers-tolerance of these estimating algorithms and detecting methods.

Keywords—Parameter Identification, Abrupt Change, Outlier-tolerant Identification, Multi-dimensional Process, Detection.

I. INTRODUCTION

States analysis and faults diagnosis (SAFD) are very important in terms of guaranteeing the safety of spacecraft. Telemetry data flow from sensors fixed in spacecraft are paramount information resources to judge and to diagnose the status of spacecraft in orbit and substantial of them flow can be expressed as multi-dimensional time series samples. Theoretically, the key residuals analysis of model-based fault diagnosis and abnormalities detection in complicated-structure dynamic system with complicated-structured can be converted into the problems on statistical inferences of multi-dimensional time series samples.

In mathematical statistics fields, the major characteristic parameters of a multi-dimensional stationary time series are the covariance matrix \(\Sigma\) and the location vector \(\mu\), which are expressed conjunctively as \((\Sigma, \mu)\).

A lot of theoretical analysis and practical applications show that the least squared estimator and the maximum likelihood estimation algorithm of \((\Sigma, \mu)\) are sensitive to outliers in time series samples\(^{[1-4]} \). If there are outliers as well as impurities in the telemetry data flow, the classical estimation algorithms of \((\Sigma, \mu)\) stated above will be fallacious and the safety monitoring results will be incorrect.

In order to overcome the hypersensitivity of ordinary estimators to outliers in multi-dimensional time series samples and to reduce erroneous alarm ratio as well as illusive alarm ratio, a series of outliers-tolerant and affine equivariant identification algorithms are built for the joint parameter \((\Sigma, \mu)\). And through combining the new identification algorithms and the T-statistics detection approaches, we modified the accuracy in detecting and diagnosing abnormalities. Simulation results given at the end of this paper validate that these modified estimation and detection algorithms are effective.

The aim of this paper is to set up a practical algorithm to identify model parameters and to detect abrupt changes in a stationary multi-dimensional process. The outlier-tolerant identifications of covariance and location vector are built in Section 2, an convergent iterative computing algorithm is presented in Section 3. In section4, a new kind of detection algorithm is set up for pulse-type changes. In section 5, some simulations are done. The simulation results show that the algorithms given above are effective.

II. OUTLIERS-TOLERANT IDENTIFICATION ALGORITHMS OF PROCESS PARAMETER \((\Sigma, \mu)\)

In this paper, we will all of our attention to the weak stationary process \(\{Y(t), t \in T\}\), the slices of which is denoted as \(Y(t)\). The stationary process are iid (independent and identical distribution) and satisfies function \(F\), mean and covariance of which is \((\mu, \Sigma)\).

It is well-known that the optimal estimators of location vectors \(\mu \in \mathbb{R}^m\) and covariance matrix \(\Sigma \in \mathbb{R}^{m \times m}\) in a multidimensional stationary Gaussian process are as follows
\[ \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\mu})(Y_i - \hat{\mu})^T \] (1)

A lot of research results in recent 30 years discovered that the so-called optimality is correct only in the case that the stationary process is not contaminated and the sampling series are normal and contains no outliers. In other words, if the normal processes are contaminated by abnormal disturbance and there are outliers in the sampling series, the estimators (1) are fallible and identification results are not credible. Because of this, it is necessary to modify the estimators (1) and to make sure the modified estimators are insensitive to outliers.

A. M-type Estimators and Affine Equi-variant Estimators

For a symmetrical real matrix \( \Sigma \in \mathbb{R}^{m \times m} \), the elongation vector \( V(\Sigma) \in \mathbb{R}^p \) of the matrix \( \Sigma \) is defined as follows

\[ V(\Sigma) = \left( \frac{\Sigma_{11}}{\sqrt{2}}, \ldots, \frac{\Sigma_{m,m}}{\sqrt{2}}, \Sigma_{2,1}, \Sigma_{3,1}, \Sigma_{1,2}, \ldots, \Sigma_{m,m-1} \right)^T \] (2)

Where the number of dimensions \( p = m + m^2 \).

Obviously, the vector \( V(\Sigma) \) is uniquely determined by the symmetrical matrix \( \Sigma \), and vice versa.

Lemma 1. For any multivariate real function \( \phi(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times m} \), a \( m \times m \) dimensional matrix \( \Sigma^\phi(X) \) and \( m \)-dimensional vector \( \mu^\phi(X) \) can be uniquely constructed by the following formula

\[ \phi(X) = \begin{bmatrix} V(\Sigma^\phi(X)) \\ \mu^\phi(X) \end{bmatrix} \] (3)

Definition 1. For any \( m \times m \) dimensional orthogonal matrix \( Q \in \mathbb{R}^{m \times m} \), the function \( \phi(\cdot) \) is called orthogonal equivariant if there exists a function \( \phi(\cdot) \) satisfying the following relations:

\[ \begin{align*}
\mu^\phi(QX) &= Q \mu^\phi(X) \\
\Sigma^\phi(QX) &= Q \Sigma^\phi(X) Q^T
\end{align*} \] (4)

Using the set \( \Phi \) to denote all the orthogonal equivariant functions \( \phi : \mathbb{R}^m \rightarrow \mathbb{R}^{p \times m} \), it has been proved that all the functions \( \phi \in \Phi \) can be expressed by the following three functions:

\[ \phi_\mu, \phi_\sigma, \phi_\eta : \mathbb{R}^m \rightarrow \mathbb{R} \quad (R^+ = [0, +\infty)) \] (5)

and three functions satisfy the following relationships:

\[ \begin{align*}
\mu^\phi(Z) &= Z \mu^\phi(\|Z\|^2) \\
\Sigma^\phi(Z) &= ZZ^T \phi_\eta(\|Z\|^2) - I \phi_\eta(\|Z\|^2)
\end{align*} \] (6)

Conversely, all the functions satisfying the above forms are orthogonal equi-variant.

\[ j \phi_\rho \left( d^2(Y, \hat{\mu}, \hat{\Sigma}) \right) |(Y - \hat{\mu})| \theta = 0, \]

\[ j \phi_\rho \left( d^2(Y, \hat{\mu}, \hat{\Sigma}) \right) |(Y - \hat{\mu})^T - \phi_\eta \left( d^2(Y, \hat{\mu}, \hat{\Sigma}) \right) \theta = 0 \]

where \( d^2(Y, \mu, \Sigma) = (Y - \mu)^T \Sigma^{-1}(Y - \mu) \).

Lemma 2. If the three functions \( \{ \phi_\mu, \phi_\sigma, \phi_\eta \} \) are selected as \( \phi_\mu(r) = \phi_\eta(r) = \phi_\sigma(r) \equiv 1 \), then the orthogonal equivariant M-type estimators is the least squared estimators of parameters \( (\mu, \Sigma) \).

Lemma 3. If the process \{ \( Y(t), t \in T \) \} is stationary Gaussian process, density function of which is \( f((x - \mu)^T \Sigma(x - \mu)) \), and the three functions are selected as follows

\[ \phi_\mu(r) = \phi_\sigma(r) = \phi_\eta(r) = -\frac{2 r f^2(r)}{f^3(r)}, \quad \phi_\eta(r) \equiv 1 \] (8)

Then orthogonal equivariant M-type estimators is the Maximum Likelihood (ML-) estimators.

In order to conveniently monitor the spaceflight state and the health status of an in-orbit spacecraft, the coordinate are usually changed and several different coordinate systems are usually adopted. Generally, results of analysis and judgment must be accordant in different coordinate systems used in process monitoring. Because of these reasons stated above, estimators of \( (\mu, \Sigma) \) must be equivariant under different affine transformation of coordinate system.

Definition 3. For any matrix \( P \in \mathbb{R}^{m \times n} \) and vector \( Z \), the function \( \phi(\cdot) \) is called affine equivariant if \( \phi(\cdot) \) satisfies the following properties:

\[ \begin{align*}
\mu^\phi(PX + Z) &= P \mu^\phi(X) + Z \\
\Sigma^\phi(PX + Z) &= P \Sigma^\phi(X) P^T
\end{align*} \] (9)
Definition 4. For any vector $Z \in \mathbb{R}^n$ and matrix $P \in \mathbb{R}^{m\times n}$, the functional $(\hat{\mu}(\phi(Y)), \hat{\Sigma}(\phi(Y)))$ is called affine equivariant estimators of $(\mu, \Sigma)$ if the following relationships are satisfied

$$
\left(\hat{\mu}(\phi(PY + Z)), \hat{\Sigma}(\phi(PY + Z))\right) = (P\hat{\mu}(\phi(Y)) + Z, P\hat{\Sigma}(\phi(Y))P^T)
$$

(10)

Where, $\phi(Y)$ is the probability distribution of stochastic vector $Y \in \mathbb{R}^m$.

Theorem 1. For any three real functions $\{\phi_\mu, \phi_\delta, \phi_\eta\}$ valued on $[0, +\infty)$, the estimators $(\hat{\mu}(F), \hat{\Sigma}(F))$ given in formulae (7) are affine equivariant as long as all the slices are from a stationary independent Gaussian process.

Proof: Because distribution functions of stationary independent Gaussian process are ellipsoidal with absolute continuity, Theorem 1 can be directly educed with results in paper [9].

It can also be proved that all of the affine equivariant estimators can be expressed by formulae (7) and that rotation and movement of coordinate are two special illustrations of affine equivariant transformations. So, affine equivariant property of estimators of parameters $(\mu, \Sigma)$ is necessary and valuable in spaceflight engineering.

Theorem 2[7]. For any multi-dimensional spherically symmetric stochastic process $\{Y(t), t \in T\}$, distribution function of which is $\Phi$, if the function $\phi$ satisfies following assumptions (c1)–(c3):

(c1). There exists a constant $r^* \in [0, +\infty)$ which make sure $r^\gamma \phi_\delta(r) - m\phi_\delta(r^*) > 0$;

(c2). $\phi_\mu(r)$ is bounded and non-increasing, $\phi_\delta(r)$ and $\phi_\eta(r)$ are continuous; $r\phi_\mu(r)$ and $r\phi_\delta(r)$ are bounded and non-decreasing.

(c3). $\phi_\mu(r) > 0, \phi_\delta(r) > 0, \phi_\eta(r) > 0$ $\forall r \in [0, +\infty)$;

the affine equivariant M-type estimators $(\hat{\mu}(F_n), \hat{\Sigma}(F_n))$ given in formulae (8) is asymptotically Gaussian

$$
\sqrt{n}\left(\left\{\hat{\Sigma}(F_n), \hat{\mu}(F_n)\right\} - \left\{I, 0\right\}\right) \xrightarrow{w} N\left(0, \Lambda^{-1}C\Lambda^{-1}\right)
$$

(11)

Theorem 3[7]: For any affine equivariant M-type estimators $(\hat{\mu}(F_n), \hat{\Sigma}(F_n))$, $\phi_\mu(r) = r\phi_\mu(r) - m\phi_\delta(r)$ defined on $[0, +\infty)$, the asymptotic covariance matrix $\Lambda^{-1}C\Lambda^{-1}$ can be expressed as three different parts

$$
\Lambda^{-1}C\Lambda^{-1} = \text{diag}(D_1, D_2, D_3)
$$

(12)

Where, $D_1 = d_\mu I_m$, $D_2 = d_\phi I_m$ and

$$
D_2 = \begin{bmatrix}
\frac{1}{m}d_\phi + \frac{m-1}{m}d_\eta & \cdots & \frac{1}{m}(d_\phi - d_\eta) \\
\vdots & \ddots & \vdots \\
\frac{1}{m}(d_\phi - d_\eta) & \cdots & \frac{1}{m}d_\phi + \frac{m-1}{m}d_\eta
\end{bmatrix}
$$

(13)

Where

$$
\begin{align*}
d_\mu &= \frac{1}{m} E_{\phi_\mu}\left(\frac{1}{m}\right)^2 + \frac{m-1}{m^2} E_{\phi_\mu}\left(\frac{1}{m}\right)^2 \\
d_\phi &= \frac{2}{m(m+1)} E_{\phi_\phi}\left(\frac{1}{m}\right)^2 + E_{\phi_\phi}\left(\frac{1}{m}\right)^2 \\
d_\eta &= \frac{2}{m(m+1)} E_{\phi_\eta}\left(\frac{1}{m}\right)^2 + E_{\phi_\eta}\left(\frac{1}{m}\right)^2
\end{align*}
$$

(14)

Theorem 2 and theorem 3 show that it is reasonable to separately design estimators for location vector $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma \in \mathbb{R}^{m\times m}$ if the asymptotic covariance matrices are used as the judgment indices to identify a multi-dimensional stationary ergodic process.

B. Optimal Outliers-tolerant Estimator of $\mu$

Considering that the estimators of location vector $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma \in \mathbb{R}^{m\times m}$ can be designed separately, it is acceptable to suppose that the covariance $\Sigma$ is determinate. Without loss of generality, we assume that the covariance matrix of the multi-dimensional stationary Gaussian process $\Sigma = I_m$ and define the functional $\mu_\phi(\phi)$ in formulae (14) as a M-type estimator [4,5] of location vector $\mu$:

$$
\int_{\mathbb{R}^m} \phi(Y - \hat{\mu}_\phi(\phi))dF(Y) = 0
$$

(15)

Where the function $\phi(\cdot)$ is defined on $[0, +\infty)$ and smooth in every subsection of $[0, +\infty)$, and $F(\cdot)$ is the distribution function of the $m$-dimensional stochastic vector $Y$.  


For multi-dimensional stationary ergodic process, if the empirical distribution function $F_n$ is used to substitute for the probability distribution $F(*)$, then the formulae (15) can be simplified as follows

$$
\frac{1}{n} \sum_{i=1}^{n} \phi[(Y_i - \hat{\mu}_n(\varphi))^2] = 0
$$

(16)

Obviously, the identification algorithm given in formulae (16) constitutes a large set. This set is called M-type estimator set of the location vector $\mu$ and denoted with $\Theta_F$. Generally, properties and characters of M-type estimators changes along with function $\phi(*)$. In order to overcome bad influences from faults in processes and outliers in samples, much attention will be paid to a so-called redescending (RD-) subset of $\Theta_F$, which is called the re-descending subset of M-type estimators.

**B.1 Rd-Identification of $\mu$**

Redescending (Rd-) estimator of location vector $\mu$ is an outliers-tolerant estimator which is determined by the formula (14) or (15) where the function $\phi_\mu(*)$ satisfies the following conditions (d1)–(d3) as follows:

(d1). $\phi_\mu(*)$ is a bounded continuous function which is smooth in every subsection of $[0, +\infty)$;

(d2). there exists a non-negative constant $c > 0$, so as for every $r > c$ satisfying $\phi_\mu(r) = 0$;

(d3). $\phi_\mu(\|Y\|^2)Y$ is measurable about $Y$ and Doob decomposable, the distribution $F(Y)$ is centrally symmetrical on $R_m$ and satisfies following formulae

$$
\int_{R_m} \phi_\mu(Y^r)YdF(Y) = 0
$$

(17)

The set $\Phi_\mu$ denotes all of the functions which satisfy the conditions (d1)–(d3). Collins[4] has proved that the Rd-type estimator $\mu_\phi$ is coincident and asymptotically Gaussian. Furthermore, it can be proved[6] that, if $\phi_\mu(*) \in \Phi_\mu$, then the estimator $\hat{\mu}_n(\varphi)$ defined in formulae (16) is asymptotically Gaussian, asymptotical covariance matrix of which is as follows

$$
\lim_{n \to \infty} Cov(\hat{\mu}_n(\varphi), \hat{\mu}_n(\varphi)) = d(\phi_\mu, F)I_m
$$

(18)

where

$$
d(\phi_\mu, F) = \frac{1}{2} \frac{E(\phi_\mu(Y^r)Y^r)}{E(\phi_\mu(Y^r)Y^r)^2}
$$

(19)

Assume that distribution function $F$ of the slice $Y(t)$ in a stationary process is absolutely continuous and spherically symmetrical. Using the redescending property of function $\phi_\mu(*)$ and the following spherical coordinate transformation

$$
\begin{align*}
y_1 &= r \cos \varphi_1 & 0 \leq r < +\infty \\
y_2 &= r \sin \varphi_1 \cos \varphi_2 & 0 \leq \varphi_2 \leq \pi \\
\vdots & \quad \vdots & (i = 1, 2, \cdots m - 1) \\
y_m &= r \sin \varphi_1 \cdots \sin \varphi_{m-1} & 0 \leq \varphi_{m-1} \leq 2\pi
\end{align*}
$$

(20)

The functional $d(\phi_\mu, F)$ can be simply expressed as follows

$$
d(\phi_\mu, F) = \frac{1}{2} \frac{\int C_m \int \phi_\mu(r)rf(r)dr}{(\int C_m \int \phi_\mu(r) + \phi_\mu(r) r^2 f(r)^2 dr)^2}
$$

(21)

Where, $f(r) = \frac{1}{2} C_m f^{m-1}$, $f^2(r)$ is probability function of stochastic variable $Y(t)$ and

$$
C_m = 2\pi \int_0^{\pi} \sin^{m-1} \varphi_1 d\varphi_1 \int_0^{\pi} \sin^{m-2} \varphi_2 d\varphi_2 \cdots \int_0^{\pi} \sin^{m-m-2} \varphi_{m-2} d\varphi_{m-2}
$$

**B.2 Optimal Outliers-tolerant Rd-type Estimator of $\mu$ under Bounded BCS constraint**

In order to improve robustness of statistics, Hampel et al put forward an important infinitesimal influence function and set up Bias Change Sensitivity (BCS) optimality approach[3]. Generally, this new idea can be used to design the optimal estimator of location vector for reference. The infinitesimal influence function of M-type functional $\mu_F(\phi)$ and sensitivity index are defined as follows

$$
\begin{align*}
IF(X, \mu_F(\phi), F) &= \lim_{\epsilon \to 0} \frac{\mu_{\phi_\epsilon}(\phi) - \mu_F(\phi)}{\epsilon} \\
B_F(\phi) &= \sup_{X \in H^m} IF(X, \mu_F(\phi), F)
\end{align*}
$$

(22)

Where $F_\epsilon = (1-\epsilon)F + \epsilon \Delta_X$ and $\Delta_X$ is a distribution function, the mass of which is integrated at X. Obviously, the distribution function $F_\epsilon$ is suitable to describe processes with pulse-type faults with $\epsilon$ depicting the possibility of faults.
International Journal of Emerging Technology and Advanced Engineering

If the distribution function $F_0$ of the principal part is $N(0, I_m)$, for any Rd-type function $\phi_{\mu,2} \in \Phi_c$, it will be easy to use spherical coordinate transformation formulae (19) to prove that the Bias Change Sensitivity (BCS) of the identifier functional $\mu_c(\phi_{\mu,2})$ is as follows\[5

$$B_{\mu}(\phi_{\mu,2}) = \sup_{r \in [0, c]} \left\{ m \int_0^r \phi_{\mu,2}(r) f(r) dr + 2 \int_0^r \phi_{\mu,2}(r) f(r) dr \right\} \tag{22}$$

The expression (20) can be simply expressed as follows

$$B_{\mu}(\phi_{\mu,2}) = \sup_{r \in [0, c]} \left\{ \sqrt{m} \phi_{\mu,2}(r) \right\} = \frac{1}{M_c(\phi_{\mu,2})} \sup_{r \in [0, c]} \left\{ \sqrt{r} \phi_{\mu,2}(r) \right\} \tag{23}$$

Where the notation

$$M_c(\phi_{\mu,2}) = \sqrt{m} \int_0^c \phi_{\mu,2}(r) f(r) dr + \frac{2}{m} \int_0^c \phi_{\mu,2}(r) f(r) dr$$

**Theorem 4.** For any given non-negative constant $d$, the optimal Rd-type function $\phi$, which minimizes $d_c(\phi_{\mu,2})$ in set $\Phi_d = \{ \phi_{\mu,2} : \phi_{\mu,2} \in \Phi_c, B_{\mu}(\phi_{\mu,2}) \leq d \}$, is as follows

$$\phi_{\mu,2}(r) = \begin{cases} 1, & 0 \leq r \leq r_0 \\ \frac{c - r}{c - r_0}, & r_0 \leq r \leq c \\ 0, & r > c \end{cases} \tag{24}$$

Where the constant $r_0$ is determined by

$$\int_0^c f(r) dr + \int_{r_0}^c \left( \frac{m}{m - r} \frac{(m + 2)r}{m(c - r)} \right) f(r) dr = \frac{\sqrt{r_0}}{\sqrt{d}} \left( \frac{c}{3} \leq r_0 \leq c \right) \tag{25}$$

**Proof.** Substituting formulae (24) into $M_c(\phi_{\mu,2})$ defined above, following formula is obtained

$$M_c(\phi_{\mu,2}) = \sqrt{m} \int_0^c f(r) dr + \int_{r_0}^c \frac{mc - (m + 2)r}{m(c - r)} f(r) dr = \frac{\sqrt{r_0}}{d} \tag{26}$$

Since

$$\sup_{r \in [0, c]} \left\{ \sqrt{r} \phi_{\mu,2}(r) \right\} = \max \left\{ \sqrt{r_0}, \sup_{r \leq r_0} \left\{ \sqrt{r(c - r)} \right\} \right\} = \sqrt{r_0} \tag{27}$$

So the equation $B_{\mu}(\phi_{\mu,2}) = d$ holds and thus the constriction $B_{\mu}(\phi_{\mu,2}) \leq d$ and the relation $\phi_{\mu,2} \in \Phi_c$ are both satisfied.

On the other hand, inserting the formulae (24) into formulae (18), the following formula (28) is satisfied for all the function $\phi \in \Phi_c$

$$d_c(\phi, f) = \frac{1}{mM_c(\phi)} \left[ \left( (\phi(r) - 1) \right)^2 f(r) dr + 2\sqrt{mM_c(\phi)} \cdot 2 \int f(r) dr \right] \tag{29}$$

Select a new set

$$\tilde{\Phi}_c = \left\{ \phi : \mu_c(\phi) = M_c(\phi), \sup_{r \in [0, c]} \left\{ \sqrt{r} \phi_{\mu,2}(r) \right\} \leq \sup_{r \in [0, c]} \left\{ \sqrt{r} \phi_{\mu,2}(r) \right\} \right\}$$

It is obvious that $\tilde{\Phi}_c \subseteq \Phi_c$ and that to minimize $d_c(\phi, f)$ in set $\tilde{\Phi}_c$, $B_{\mu}(\phi_{\mu,2}) \leq d_c(\phi, f)$ is equivalent to minimize $d_c(\phi, f)$ and to minimize $S(\phi) = \int (\phi(r) - 1)^2 f(r) dr$ in set $\Phi_\phi$. By reduction to absurdity, it can be proved that $\phi_{\mu,2}$ minimizes $S(\phi)$ over the set $\Phi_\phi$ and therefore $\phi_{\mu,2}$ minimize $d_c(\phi, f)$ in the set $\Phi_\phi$.

**C. Optimal Outliers-tolerant Estimators of $\Sigma$**

In terms of designing a suitable and outliers-tolerant estimator of covariance matrix $\Sigma$, Theorem 3 provides a reasonable simple route which is to select two reasonable functions $\phi_r$ and $\phi_r$.

**C.1 Bounded Function Set of Doob Decomposable $\phi_r$**

The so-called Doob decomposable bounded function $\phi_r$ is the function which not only satisfies the Fisher coincidence condition $\int \phi_r(d(Y, \hat{\Sigma}))dF = 0$ but also possesses properties (e1)-(e5):

- (e1). function $\phi_r(r)$ is continuous and smooth in every subsections of $[0, +\infty)$; $\| \phi_r(r) \| \leq c (c \geq m)$;
- (e2). $\phi_r(\| \|)$ is measurable and Doob decomposable;
- (e3) for any centrally symmetrical distribution function $F(Y)$ on $R^m$, $\int_0^m \phi_r(\|\|)dF(Y) = 0$.
There are quite a lot of Doob decomposable bounded functions. Using notation \( \Phi_{\tau, \varepsilon} \) to denote the set of Doob decomposable bounded functions, the next step is to seek a suitable function \( \phi_{\tau} \) in set \( \Phi_{\tau, \varepsilon} \) which will be used to design affinely equivariant outliers-tolerant estimator of covariance matrix \( \Sigma \).

### C.2 Determination of Minimax Optimal Function \( \phi_{\tau} \)

With a view to the formulae (14) and theorem 3 as well as the fact that \( d_{\varepsilon}(\phi_{\tau}, f) \) is in connection with function \( \phi_{\tau} \) and distribution \( F \), \( d_{\varepsilon}(\phi_{\tau}, f) \) is simply denoted as \( d_{\varepsilon} \) and can be simplified by the spherical coordinate transformation stated above,

\[
d_{\varepsilon}(\phi_{\tau}, f) = \frac{mk_{\lambda}^{\lambda}}{2} \int_{\tau} \phi_{\tau}^{j}(r) r^{\lambda-\beta-\gamma} f(r) dr \]

(30)

Using formula (30) and Cauchy-Schwitz inequality, the following inequality (28) can be deduced as

\[
d_{\varepsilon}(\phi_{\tau}, f) = \frac{m}{2} \cdot \frac{E_{\varepsilon} \phi_{\tau}^{2}(y_{\tau})}{\{E_{\varepsilon} \phi_{\tau}^{2}(y_{\tau}) \}^{\gamma}} \geq \frac{m}{2} \cdot \frac{1}{I_{\varepsilon}(f)} \]

(31)

where \( I_{\varepsilon}(f) = k_{\lambda} \int_{\tau} \omega_{\varepsilon}(r) r^{\lambda-\beta-\gamma} f(r) dr \) and

\[
\omega_{\varepsilon}(r) = \left\{ m + \frac{2r^{\gamma}(r)}{f(r)} \right\} .
\]

(32)

**Theorem 5.** If the principal part is \( N(0, I_{m}) \), then the least favorable density function of \( d_{\varepsilon}(\phi_{\tau}, f) \) in \( \Omega_{\varepsilon} \) is determined by the following function

\[
\tilde{f}(r) = \begin{cases} (1-\varepsilon) f_{0}^{\gamma}(r), & r \leq c_{\varepsilon} + m \\ (1-\varepsilon) f_{0}^{\gamma}(r), & r > c_{\varepsilon} + m \end{cases}
\]

(33)

Accordingly, the minimax optimal \( \phi_{\tau} \) in \( \Phi_{\tau, \varepsilon} \) is as follows

\[
\tilde{\phi}_{\tau}(r) = \begin{cases} r - m, & r \leq c_{\varepsilon} + m \\ c_{\varepsilon}, & r > c_{\varepsilon} + m \end{cases}
\]

(34)

And the equation (34) is satisfied

\[
\inf_{\phi_{\tau} \in \Phi_{\tau, \varepsilon}} \sup_{f \in \Omega_{\varepsilon}} d_{\varepsilon}(\phi_{\tau}, f) = \frac{1}{I_{\varepsilon}(\omega_{\varepsilon})} = \sup_{f \in \Omega_{\varepsilon}} \inf_{\phi_{\tau} \in \Phi_{\tau, \varepsilon}} d_{\varepsilon}(\phi_{\tau}, f)
\]

(35)

Where function \( \tilde{g}(r) \) is non-negative and defined on \([c - m, +\infty)\) as follows

\[
\tilde{g}(r) = c_{0} \int_{r,m} \frac{1}{2se} \left( \frac{s}{r} \right)^{c_{0}m} ds
\]

(36)

Where \( r \geq c_{\varepsilon} + m, c_{0} \) is the normalization factor.

**Proof.** Because of the formula (36), formulae (37) can be deduced for \( r \leq c_{\varepsilon} + m \) as follows by using properties of Gaussian distribution density function

\[
\omega_{\varepsilon}(r) = - \left\{ \frac{2r^{\gamma}(r)}{f(r)} \right\} + m = r - m \leq c_{\varepsilon}
\]

(38)

And for \( r > c_{\varepsilon} + m \), the following expression (39) can be deduced

\[
- \frac{\tilde{g}(r) + (1-\varepsilon)f_{0}(r)}{\tilde{g}(r) + (1-\varepsilon)f_{0}(r) + m} = c_{\varepsilon}
\]

(39)

So we have the following expression for \( r \geq c_{\varepsilon} + m \)

\[
\omega_{\varepsilon}(r) = - \left\{ \frac{2r^{\gamma}(r)}{f(r)} \right\} + m = c_{\varepsilon}
\]

(40)

The expression (40) shows that \( \tilde{\phi}_{\tau}(r) = \omega_{\varepsilon}(r) \). So, for any distribution function \( F \in \Omega_{\varepsilon} \) the following inequation can be obtained

\[
d_{\varepsilon}(\omega_{\varepsilon}, f) = \frac{m}{2} \int_{\tau} k_{\lambda}^{\lambda} \int_{\tau} \phi_{\tau}^{j}(r) r^{\lambda+\beta-\gamma} f(r) dr
\]

\[
\leq \frac{m}{2} \int_{\tau} k_{\lambda}^{\lambda} \int_{\tau} \phi_{\tau}^{j}(r) r^{\lambda+\beta-\gamma} f(r) dr d_{\varepsilon}(\omega_{\varepsilon}, f)
\]

(41)
International Journal of Emerging Technology and Advanced Engineering

and

\[ \sup_{r \in \Omega_x} \inf_{\phi \in \Phi_r} d_r(\phi_r, F) \geq \sup_{r \in \Omega_x} \left\{ \frac{m}{2} \frac{1}{I_r(f)} \right\} \quad (42) \]

On the other hand, for any function \( F \in \Omega_x \) and \( \Phi_r \in \Phi_r \), the following inequation can be obtained

\[ \sup_{r \in \Omega_x} \inf_{\phi \in \Phi_r} d_r(\phi_r, f) \leq \inf_{\Phi_r} \sup_{r \in \Omega_x} d_r(\phi_r, f) \quad (43) \]

So, \((\bar{\phi}_r, \bar{f})\) is a saddle point of \( d_r(\phi_r, f) \) on the field \( \Phi_r \times \tilde{\Omega}_x \), thus

\[ \inf_{\Phi_r} \sup_{r \in \Omega_x} d_r(\phi_r, f) = d_r(\bar{\phi}_r, \bar{f}) = \sup_{r \in \Omega_x} d_r(\phi_r, f) \quad (44) \]

By comparison with definition of the least favorable distribution as well as the Minimax optimality, it is obvious that all the results given in theorem 6 are true.

C.3 Determination of the Optimal Functions \( \phi_\mu \) and \( \phi_\tau \) under Bounded HIS constraint

Let us construct two affinely invariable parameters which are defined by characteristic values of symmetrical matrix \([3,5]\), one of which is called logarithm scale parameter \( \tau \) and the other is called shape parameter \( \eta \):

\[
\begin{align*}
\tau &= \frac{1}{m} \ln|\Sigma| = \frac{1}{m} \sum_{i=1}^{m} \ln \lambda_i \\
\eta &= \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\ln \lambda_i - \tau)^2}
\end{align*}
\]  

(45)

Correspondingly, estimators of parameters \((\tau, \eta)\) are defined as follows

\[
\begin{align*}
\hat{\tau}(F) &= \frac{1}{m} \ln |\hat{\Sigma}(F)| = \frac{1}{m} \sum_{i=1}^{m} \ln \hat{\lambda}_i (F) \\
\hat{\eta}(F) &= \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\ln \hat{\lambda}_i (F) - \hat{\tau}(F))^2}
\end{align*}
\]  

(46)

Where \( \lambda \) and \( \hat{\lambda} \) are characteristic values of symmetrical matrix \( \Sigma \) and \( \hat{\Sigma}(F) \) respectively.

For any Gaussian distribution \( F \), it can be deduced that the infinitesimal influences\([3,5]\) of estimators \((\hat{\tau}(F), \hat{\eta}(F))\) are equal to

\[
\begin{align*}
\hat{\tau}(F) &= \frac{2\phi(\tau, F)}{E_{\phi}(d^2(Y, \mu, \Sigma))} \\
\hat{\eta}(F) &= \frac{m\phi(\eta, F)}{E_{\phi}(d^2(Y, \mu, \Sigma))}
\end{align*}
\]  

\[ (47) \]

Without lose of generality, an assumption of \((\mu, \Sigma) = (0, I)\) is used in this section. It is easy to deduce two bias change sensitivity functionals:

\[
\begin{align*}
r_{\mu}(\phi) &= \sup_{r \in \Omega_x} \frac{\phi_r(r)}{\tau} \left( r \right) dr \\
r_{\eta}(\phi) &= \sup_{r \in \Omega_x} \frac{\phi_r(r)}{\tau} \left( r \right) dr (48)
\end{align*}
\]

Denoting the median of \( \chi^2(m) \) distribution as \( m_{\mu} \), two bounded functions are designed as follows

\[
\check{\phi}_r(a) = \frac{1}{\max\{r,a\}}, \quad \check{\phi}_\eta(b) = \frac{r-m_{\mu}}{\max\{r-m_{\mu}, b\}}
\]  

(49)

Where \( a \geq 0, b \geq 0 \), and \( \max(\cdots) \) is maximum function.

Theorem 6. For any two nonnegative dual constants \( c_r \geq r_{\tau}(\phi_{\tau,0}) \) and \( c_\eta \geq r_{\eta}(\phi_{\eta,0}) \), there exists two dual constants \( a \) and \( b \) which satisfy the following equations:

\[
r_{\tau}(\phi_{\tau,a}) = c_1, \quad r_{\eta}(\phi_{\eta,b}) = c_2
\] 

(50)

Where \( \phi_{\tau,a} \) is the function which arrives at the minimum of \( d_r(\phi_r, f_0) \) in the set \( \{\phi_r \in \Phi_r : \|f(Y, \tau, F_0)\| \leq c_1\} \), and \( \phi_{\eta,b} \) is the function which arrives at the minimum of \( d_\eta(\phi_\eta, f_0) \) in \( \{\phi_\eta \in \Phi_\eta : \|f(Y, \eta, F_0)\| \leq c_2\} \).

III. ITERATIVE ALGORITHMS FOR OUTLIERS-TOLERANT ESTIMATORS OF \((\Sigma, M)\)

Theoretically, the optimal functions \( \{\phi_{\mu}, \phi_{\tau}, \phi_{\eta}\} \) given above and equation (8) can be adopted to determine the optimal outliers-tolerant estimators of process parameters \((\mu, \Sigma)\). But, because of nonlinearity of equation (8), the estimators are difficult to be calculated in practice. Now, a set of iterative algorithms are constructed to find out the optimal outliers-tolerant estimators.
Assume that the stationary process run and a series of time series data are collected. The finite-sample format of formula (8), which are

\[
\begin{align*}
\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) & = 0 \\
\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) & = 0
\end{align*}
\]

\[\begin{align*}
\sum_{i=1}^{n} \phi_i & \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) Y(t_i) - \hat{\mu} Y(t_i) = 0 \\
\sum_{i=1}^{n} \phi_i & \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) Y(t_i) - \hat{\mu} Y(t_i) = 0
\end{align*}\]

\[
\begin{align*}
\hat{\mu} & = \frac{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) Y(t_i)}{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)} \\
\hat{\Sigma} & = \frac{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right) (Y(t_i) - \hat{\mu})(Y(t_i) - \hat{\mu})'}{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)}
\end{align*}
\]

So, using some notation symbols as follows

\[
\begin{align*}
\tilde{\phi}_\mu(Y(t_i), \mu, \Sigma) & = \frac{\phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)}{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)} \\
\tilde{\phi}_\eta(Y(t_i), \mu, \Sigma) & = \frac{\phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)}{\sum_{i=1}^{n} \phi_i \left( d^2 \left( Y(t_i), \hat{\mu}, \hat{\Sigma} \right) \right)}
\end{align*}
\]

Then a series of iterative algorithms for \( k(=0,1,2,\cdots) \) can be constructed as follows

\[
\begin{align*}
\hat{\mu}_{k+1} & = \frac{\sum_{i=1}^{n} \tilde{\phi}_\mu(Y(t_i), \hat{\mu}_k, \hat{\Sigma}_k) Y(t_i)}{\sum_{i=1}^{n} \tilde{\phi}_\mu(Y(t_i), \hat{\mu}_k, \hat{\Sigma}_k)} \\
\hat{\Sigma}_{k+1} & = \frac{\sum_{i=1}^{n} \tilde{\phi}_\eta(Y(t_i), \hat{\mu}_k, \hat{\Sigma}_k) (Y(t_i) - \hat{\mu}_k)(Y(t_i) - \hat{\mu}_k)'}{\sum_{i=1}^{n} \tilde{\phi}_\eta(Y(t_i), \hat{\mu}_k, \hat{\Sigma}_k)}
\end{align*}
\]

Using notation \textit{med} to express the median value of a series, the iterative original values are

\[
\begin{align*}
\hat{\mu}_0 & = \text{med} \{ Y(t_i) \} \\
\hat{\mu}_{k+1} & = \text{med} \{ Y(t_i) \} \\
\hat{\Sigma}_{k+1} & = \text{med} \{ \hat{\Sigma}(t_i) \}
\end{align*}
\]

where, \( \bar{Y}_j(t_i) = y_j(t_i) - \text{med} \{ y_j(t_i) \} \ (j = 1, \cdots, m) \).

IV. DETECTION AND DIAGNOSIS OF PULSE-TYPE CHANGES

When a multi-dimensional stationary process is monitored, substantial attention is paid to abrupt abnormal changes such as pulse-type changes, step-type jumps and compound of many kinds of abrupt changes. Because of the facts that step-type changes can be turn into pulse-type changes in the corresponding difference process, this section discusses how to detect as well as to identify pulse-type changes.

Assume that a multi-dimensional time series \( Y(t_i) \in \mathbb{R}^n, t_i \in T \) is ergodic and measurable and its statistical mean and covariance matrix are \( \mu \) and \( \Sigma \) separately. If all of the slices \( Y(t_i) \ (t_i \in T) \) are Gaussian, it can be proved that the distribution of \( \bar{Y}(t_i) = \Sigma^{-\frac{1}{2}}(Y(t_i) - \mu) \) is \( N(0, I) \).

Relying on this fact, the following T-detection statistics \( T_{\mu,\Sigma}(Y) \sim \chi^2(m) \) is widely used to detect abrupt abnormal changes in m-dimensional process

\[
T_{\mu,\Sigma}(Y) = \frac{(Y - \mu)\Sigma^{-1}(Y - \mu)}{m}
\]

In order to use this T-detection statistics \( T_{\mu,\Sigma}(Y) \sim \chi^2(m) \) in practical processes monitoring, parameters \( \mu \) and \( \Sigma \) must be given or determined at first step. In most cases, parameters \((\mu, \Sigma)\) are unknown and must be estimated. A lot of theoretical analysis and practical applications show that, because of the bad influence resulting from outliers in sampling time series, the ratio of erroneous alarm as well as false alarm may increase if the ordinary least squared estimators are substituted for the parameters \((\mu, \Sigma)\) in formula (60). In other words, using fallible estimators of the parameters \((\mu, \Sigma)\) may result in wrong judgment in monitoring practical processes.

In order to avoid wrong judgment as much as possible, it is necessary to use these outliers-tolerant estimators of \((\mu, \Sigma)\) given in sections 2 and section 3 when the T-detection statistics \( T_{\mu,\Sigma}(Y) \) is selected to detect abrupt changes of a multi-dimensional stationary process. Now, some practical steps to monitor a stationary process are described as follows:
(a) Selecting suitable function $\phi_\mu$ in formula (24) and functions $(\phi_r, \phi_\eta)$ in (34) or (49), constructing function $\phi_r(r) = (r \phi_r(r) - \phi_\eta(r)) / m$ , and using the algorithms (52)–(55) to get outliers-tolerant estimators $(\hat{\mu}_a, \hat{\Sigma}_a)$ ;

(b) Substituting $(\hat{\mu}_a, \hat{\Sigma}_a)$ for parameters $(\mu, \Sigma)$ and to calculate Mahalanobis distance $T_{(\mu, \Sigma)}(Y(t_i))$ between measurement data $Y(t_i)$ and “center” $(\mu_a)$ ;

(c) Selecting a positive constant $C_a$ , which makes sure that $\alpha \times 100\%$ sampling points fall into the set $S_2(\alpha) = \{Y(t_i) : T_{(\mu, \Sigma)}(Y(t_i)) > C_a\}$ , and all of the samples in set $\{Y(t_i) \in \mathbb{R}^m, t_i \in T\}$ are divided into two parts

$$
\begin{align*}
S_1(\alpha) &= \{Y(t_i) : T_{(\mu, \Sigma)}(Y(t_i)) \leq C_a\} \\
S_2(\alpha) &= \{Y(t_i) : T_{(\mu, \Sigma)}(Y(t_i)) > C_a\}
\end{align*}
$$

(57)

where the parameter $\alpha$ is selected as 0.05 or 0.1 which reflects the ratio of outliers as well as the probability of pulse-type change happening in the process.

(d) Using all of the samples in $S_1(\alpha)$ , the updated LS estimators of parameters $(\mu, \Sigma)$ are obtained by

$$
\begin{align*}
\hat{\mu}(\alpha) &= \frac{1}{n_1} \sum_{S_1(\alpha)} Y(t_i) \\
\hat{\Sigma}(\alpha) &= \frac{1}{n_1} \sum_{S_1(\alpha)} (Y(t_i) - \hat{\mu}(\alpha))(Y(t_i) - \hat{\mu}(\alpha))^T
\end{align*}
$$

(58)

where $n_1$ is the number of samples in the set $S_1(\alpha)$.

(e) Substituting $(\hat{\mu}(\alpha), \hat{\Sigma}(\alpha))$ for the parameters $(\mu, \Sigma)$ in formulae (53) and computing Mahalanobis distance $T_{(\hat{\mu}(\alpha), \hat{\Sigma}(\alpha))}(Y(t_i))$ of samples $Y(t_i) \in S_2(\alpha)$ , the detection and diagnosis can be used as follows: if the Mahalanobis distance $T_{(\hat{\mu}(\alpha), \hat{\Sigma}(\alpha))}(Y(t_i)) \geq c$ of $Y(t_i)$ , then $Y(t_i)$ will be judged as abnormal, and the magnitude of outlying center of process will equal to $Y(t_i) - \hat{\mu}(\alpha)$ ; otherwise, $Y(t_i)$ will be redeployed from $S_1(\alpha)$ into $S_1(\alpha)$ , and the step will return to (d) until all the samples in $S_2(\alpha)$ are diagnosed.

(f) To go with the multi-dimensional process running along, the updating set $S_1(\alpha)$ and the estimators of $(\mu, \Sigma)$ corresponding with $S_1(\alpha)$ can be used to implement online monitoring of process.

V. SIMULATION

With the Monte Carlo simulation$^{[10]}$ of a 3-dimensional stochastic stationary process, distribution of which is $N(\mu, \Sigma)$ where $\mu = (0, 10, -10)$ and $\Sigma = diag(1, 2, 3)$ , $n = 100$ series data are generated and plotted in Fig 1(a). There are ten outlier patches which are inserted into this series at ten divergent times. Then, there are ten peaks which emerge in Fig 1(b).

Using all of the simulation series data shown in Fig 1, two kinds of estimator values of parameters $(\mu, \Sigma)$ are calculated with the least squared (LS) estimator algorithm and Rd-type estimator algorithm separately. Identification results are listed in table 1. Results of T-detection statistics are shown in Fig 2.
Results given in Table 1 evidently validate that the outliers-tolerant estimators are not only inoculated with the LS-estimators in the case that there are not outliers in stationary process but also are more reassuring than the LS-estimators in the case that the stationary process are contaminated. Comparison between Fig 2(a) and Fig 2(b) indicates similar facts stated above, which is, in terms of monitoring abnormal changes in multi-dimensional stationary processes, the Rd-estimators are more suitable and more reliable than the LS-estimators.

In order to test the online detection ability of the T-detection statistics, the 3-dimensional stationary series data given in Fig 1(a) and Fig 1(b) are appended with 10 sampling data and one of them (No. 105) is set to be much out of the center location, which are shown in Fig 3(a) and Fig 3(b) separately.

Table 2 shows that the outliers-tolerant T-detection statistics is effective to directly identify abrupt changes in multi-dimensional stationary series. The outliers-tolerant T-detection can be used to monitor abrupt changes in stationary process, validity of which is reliable.

### VI. CONCLUSION

In this paper, oriented at these open problems stated above, some detailed researches are accomplished for multi-dimensional stationary process and a series of outliers-tolerant estimators are built for covariance matrix $\Sigma$ as well as location vector $\mu$ of the process. The optimality and the outliers-tolerant properties of the proposed estimators are proved. Using these optimal outliers-tolerant estimators of ($\Sigma$, $\mu$), a new detection index, named T-detection statistics, is established to detect as well as to monitor abrupt changes in process. Simulation results validate that new detection approaches given in this paper are effective and practicable to monitor states and to detect abnormal changes in multi-dimensional stationary processes in different engineering fields.
And, approaches given in this paper are helpful to reduce the ratio of false alarm as well as the ratio of missing alarm when a process is under monitoring and to enhance the safety when a complicated system runs.

REFERENCES