Abstract: The problem of robust fault detection in command inputs in the presence of modelling errors in linear time-invariant systems is addressed. The problem leads to a hypothesis test with a test statistic taking form of a weighted quadratic sum of independent normally distributed variables. It is rather difficult to compute the threshold for such a test and the corresponding error probabilities in an exact manner. Instead, the approximative results are derived by assuming asymptotic distribution of the test statistic towards normal distribution. Convergence conditions in case of growing number of summands are presented. Performance of the resulting detector is demonstrated on a simulated DC motor-generator rig. Copyright © 2002 IFAC

Keywords: Fault detection, likelihood-ratio test, central limit theorem.

1. INTRODUCTION

The issue of robust fault detection has been treated in many papers, c.f. Patton (1994). Uncertainties, as inevitable ingredient of any model, can be formally characterised in various ways, e.g. by means of "unknown inputs", disturbance distribution matrix or parameter uncertainties thus resulting in various algorithms for generating detection alarms. A well elaborated option is also to formulate the detection problem as the problem of statistical decision making (Basseville and Nikiforov, 1993). Unfortunately, only rarely does the distribution of the test statistic conform to some of the well known distributions for which it is easy to calculate error probabilities. Therefore, approaches based on asymptotic results should be adopted as e.g. asymptotic local approach of Basseville and Nikiforov (1993).

This paper addresses the problem of detecting faults in instrumentation related to command inputs (e.g. position of the actuator). The paper provides a more detailed elaboration of the tentative results reported by Juričić and Žele (2001). The underlying system is considered to be linear and time-invariant with single input and single output. In order to account for possible (bad) influence of modelling errors the stochastic embedding technique of Goodwin et al. (1992) is applied.

The paper is organised as follows. In the second section a robust detection problem is formulated as a problem of hypothesis testing. The third section provides a solution of the test based on approximating the distribution of the test statistic with a normal distribution. Illustrative simulation results are presented in the fourth section. A summary of weak convergence theorems, used to calculate the asymptotic distribution, is given in appendix.

2. PROBLEM STATEMENT

2.1. Description of the modelling errors

As no model is perfect, the idea is to assume that the system response consists of three parts: the nominal response, unmodelled effects and statistically independent noise term. For a linear system with input \( u(t) \) and output \( y(t) \) this reads as follows

\[
y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + G_3(q^{-1}) u(t) + n(t) \tag{1}
\]

with \( u(t) \) and \( y(t) \) denoting process input and output, \( q^{-1} \) being the shift operator, \( n(t) \) normally distributed white noise \( n \sim \text{N}(0, \sigma_n^2) \), \( G_3 \) finite impulse response
(FIR) model of the unmodelled dynamics. Furthermore,
\[ A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_m q^{-m} \]
\[ B(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_d q^{-d} \]
are polynomials of the nominal transfer function. The modelling error is described with the stochastic term (Goodwin et al., 1992)
\[ G_d(q^{-n}) = \sum_{l=1}^{L} \eta_l q^{-n_l} \]
Expression (1) can be rewritten in the regression form as follows
\[ y_F(t) = \varphi^T(t) \theta_0 + \psi^T(t) \eta + n(t). \]  
(2)

The applied symbols mean
\[ \varphi(t) = \begin{b*} y_F(t-1), \ldots, y_F(t-n_a), u_F(t-1), \ldots, u_F(t-n_b) \end{b*} \]
\[ y_F(k) = \begin{b*} [y_F(q^{-1})]_k, u_F(k) = [y_A(q^{-1})]_k \end{b*} \]
\[ A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_m q^{-m} \]
with \( A \) assumed to be stable polynomial. The second term is the finite impulse response (FIR) model with
\[ \psi(t) = [u(t-1), \ldots, u(t-L)]^T, \quad \eta = [\eta_1, \ldots, \eta_L]^T. \]

The idea of the stochastic embedding approach is that undermodelling can be represented by a zero-mean normally distributed random variable
\[ \eta \sim N(0, \Sigma_\eta) \]
\[ P_{\eta\eta} = \text{diag} \\alpha^k, \quad \alpha > 0, \quad 1 > \lambda \geq 0. \]

2.2. Prediction error
Let us now see what happens if a fault affects the actuator. Such a case will be modelled by a bias term \( u^*(t) = u(t) + f_u \) with \( u(t) \) being command signal and \( u^*(t) \) being actual command input. The term \( f_u \) is assumed to take values from the discrete set \( f_u \in \mathcal{F} = \{f_u^1, \ldots, f_u^k\} \). This is a simplified version of the problem discussed by Juričić and Žele (2001) in order to avoid certain statistical dependence problems. Expression (2) now reads
\[ y_F(t) = (\varphi^T(t) + A \varphi^T(t)) \theta_0 + (\psi^T(t) + A \psi^T(t)) \eta + n(t) \]  
(3)

where
\[ A \varphi(t) = f_u [ \begin{b*} 0, \ldots, 0, 1, \ldots, 1 \end{b*} \}_{n_a} \]
\[ f_u = A(q^{-1}) f_u \]
and
\[ A \psi(t) = f_u [ \begin{b*} 1, \ldots, 1 \end{b*} \]}

If a window of \( N \) data samples is taken, equation (3) reads
\[ y_F(t) = (G(t) + \Delta G(f_u)) \theta_0 + (F(t) + \Delta F(f_u)) \eta + n(t) \]  
(4)

with
\[ y_F(t) = [y_F(t-N+1), \ldots, y_F(t)]^T, \]
\[ F(t) = [\psi(t-N+1), \ldots, \psi(t)]^T, \]
\[ G(t) = [\varphi(t-N+1), \ldots, \varphi(t)]^T, \]
\[ \Delta G(f_u) = f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} I_{N_{nu}}, \]
\[ \Delta F(f_u) = f_u I_{N_{ld}} \]
and
\[ n(t) = [n(t-N+1), \ldots, n(t)]^T. \]

From expression (4) it is easy to calculate the vector of \( N \) prediction errors
\[ \epsilon(t) = [\epsilon(t-N+1), \ldots, \epsilon(t-N+2), \ldots, \epsilon(t)]^T \]

simply as follows
\[ \epsilon(t) = y_F(t) - G(t) \hat{\theta}_0 = G(t) \Delta \theta + F(t) \eta + n(t) + \]
\[ f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} I_{N_{nu}} \hat{\theta}_0 + f_u I_{N_{ld}} \eta \]  
(5)

where \( \Delta \theta = \theta_0 - \hat{\theta}_0 \) and \( I_{N_{xu}} \) is \( r \times s \) matrix with all elements being 1.

Remark 1
Matrices \( E[\eta \eta^T] = P_{\eta\eta} \), \( E[\Delta \theta \Delta \theta^T] = P_{\theta\theta} \), \( E[\Delta \theta \eta^T] = P_{\theta\eta} \)
noise variance \( \Sigma_n^2 \) as well as the parameters \( \hat{\alpha}, \hat{\lambda} \)
used in (5) are computed off-line from normal operating records by means of identification. \( \checkmark \)

2.3. Composite hypothesis testing
We are now ready to state the main detection problem. Given the vector of prediction errors \( \epsilon(t) = [\epsilon(t-N+1), \epsilon(t-N+2), \ldots, \epsilon(t)]^T \), one has to decide whether there is bias present in the actual command signal.

Since \( \Delta \theta, \eta \) and \( n \) are assumed to be normally distributed, \( \epsilon \) is normally distributed as well, i.e. \( \epsilon(t) \sim N(\mu(f_u), \Sigma(f_u, f_u)) \) with mean
\[ E[\epsilon(t)] = \mu(f_u) = f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} I_{N_{nu}} \hat{\theta}_0 \]
covariance
\[ E[\epsilon(t) \epsilon(t)^T] = \Sigma(f_u, f_u) = G(t) P_{\theta\theta} G(t)^T + F(t) P_{\eta\eta} F(t)^T + \]
\[ + G(t) P_{\eta\eta} G(t)^T + F(t) P_{\eta\eta} F(t)^T + \]
\[ + f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} I_{N_{nu}} P_{\eta\eta} I_{N_{nu}} + \]
\[ + f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} P_{\eta\eta} I_{N_{nu}} + G(t) P_{\eta\eta} G(t)^T + \]
\[ + f_u [ \begin{b*} 0, \ldots, 0 \end{b*} ]_{N_{nu}} P_{\eta\eta} I_{N_{nu}} + \]
\[ + I_{N_{nu}} P_{\eta\eta} G(t)^T \]  
(6)

The actuator fault detection problem refers to choosing between two hypotheses:
The decision is drawn by checking the log-likelihood ratio test in which the hypothesis $H_0$ is rejected if the following condition holds (Rohatgi, 1976)

$$\lambda(\mathbf{e}) = \log \frac{p_{H_0}(\mathbf{e}(t))}{\sup_{\mathbf{f}_u} p_{H_1}(\mathbf{e}(t))} < k$$

(7)

Since $\mathbf{e}(t) \sim N(0, \Sigma(t,0))$ under $H_0$, $\mathbf{f}_u=0$ and $\mathbf{e}(t) \sim N(\mu(t), \Sigma(t))$ under $H_1$, $\mathbf{f}_u \neq 0$, the statistical test (10) leads to

$$\lambda(\mathbf{e}) = \mathbf{e}^T \Sigma^{-1} \mathbf{e} - (\mathbf{e} - \mu_1)^T \Sigma^{-1} (\mathbf{e} - \mu_1) > k$$

(8)

where $\Sigma = \Sigma(t,0)$ is the residuals covariance under $H_0$, $\Sigma = \Sigma(t, f_u)$ is the residuals covariance under $H_1$, $\mu = \mu(t, f_u)$ is the mean of the residuals under $H_1$ and $\hat{f}_u = \arg \sup_{f_u} p_{H_1}(\mathbf{e}(t))$, $f_u = A(q^{-1}) \hat{f}_u$.

The threshold $k$ in (8) should be determined so that probability of rejecting $H_0$ when it actually true is $\beta$. To calculate the threshold $k$, the probability density function of the test statistic (8) under $H_0$, i.e. $\mathbf{e}(t) \sim N(0, \Sigma(t,0))$, should first be determined.

3. LIMIT DISTRIBUTION OF THE TEST STATISTIC

Let us start with the following observation. Covariance matrix $\Sigma$ is a quadratic function of $f_u$. Since $\Sigma = \Sigma + \Delta \Sigma(f_u)$, where $\Delta \Sigma(f_u)$ may be negative definite, the difference $\Sigma^{-1}_0 - \Sigma^{-1}$ is not necessarily non-zero the Lindeberg condition can be violated.

Note that the first term in (8) has $\chi^2$-distribution

$$\mathbf{e}^T \Sigma^{-1} \mathbf{e} \sim \chi^2(N)$$

(9)

For large enough $N$, it can be approximated by the normal distribution $N(N, 2N)$, which can be easily deduced from the theorem 3 (c.f. appendix). The convergence rate (see theorem 5) is of order $1/n^{1/2}$.

To treat the second term in (8) let us first introduce the transformed vector

$$\mathbf{w} = \Sigma^{1/2}_0 (\mathbf{e} - \mu_1)$$

(10)

which is normally distributed $\mathbf{w} \sim N(\Sigma^{-1/2}_0 \mu_1, I)$.

The second term of the test statistics (8) can now be expressed via $\mathbf{w}$ as follows

$$(\mathbf{e} - \mu_1)^T \Sigma^{-1}_1 (\mathbf{e} - \mu_1) = \mathbf{w}^T \Sigma^{1/2}_0 \Sigma^{-1}_1 \Sigma^{1/2}_0 \mathbf{w}$$

(11)

After UD decomposition $\Sigma^{1/2}_0 \Sigma^{-1}_1 \Sigma^{1/2}_0 = U^T DU$ the expression (11) reads

$$w^T \Sigma^{1/2}_0 \Sigma^{-1}_1 \Sigma^{1/2}_0 w = w^T U^T D U w = \xi^T D \xi$$

(12)

where $\xi = [\xi_1, \xi_1, \ldots, \xi_N] = U \mathbf{w}$ is a vector of independent normally distributed random variables $\xi \sim N(U \Sigma^{-1/2}_0 \mu_1, I)$.

Since $\mathbf{D}$ is a diagonal matrix with elements $d_j, j=1..N$, the expression (12) can be easily rewritten as a sum of quadratic terms

$$\xi^T \mathbf{D} \xi = d_1 \xi_1^2 + d_2 \xi_2^2 + \ldots + d_N \xi_N^2$$

(13)

with $\xi$ being normally distributed with unit variance and mean $\mu_i$ ($\mu_i$ being the $i$-th element of the vector $U \Sigma^{-1/2}_0 \mu_1$).

Expression (13) can be rearranged in the following way

$$\xi^T \mathbf{D} \xi = \sum_{j=1}^N d_j (\xi_j - \mu_j)^2$$

(14)

The second term in the sum is obviously normally distributed.

The distribution of the first term is the same as that of the re-scaled expression:

$$Z_N = \sum_{j=1}^N d_j (\xi_j - \mu_j)^2$$

(15)

Since the covariance matrix (6) is full rank with finite eigenvalues and (in most practical cases) not ill-conditioned, it can be assumed that the weights in (13) obey the second condition of theorem 3. Then $Z_N$ converges in distribution to the standard normal distribution. In other words, "fair" distribution of the weights $d_j$ is required to assure convergence and appropriate rate of convergence (see theorem 5). Indeed, if, for example, $\nu_{\text{max}} \in \{1, \ldots, N\}$ is such that $d_{\text{max}} = \max\{d_j, j=1, \ldots, N\} > d_{\nu_{\text{max}}}, j=1, \ldots, N$ and $j \neq \nu_{\text{max}}$, then the distribution of (16) degenerates to the $\chi^2$-distribution.

We can now summarise the results above and state that under conditions in theorem 3 the distribution of the sum (14) can be approximated by normal distribution

$$d_1 \xi_1^2 + d_2 \xi_2^2 + \ldots + d_N \xi_N^2 \sim -N \left( \sum_{i=1}^N d_i (\xi_i - \mu_i)^2, \sum_{i=1}^N d_i^2 (2 + 2 \mu_i^2) \right)$$

(16)

Since both terms of the test statistics (8) are asymptotically normally distributed, it can be shown that the difference of both is asymptotically normal as well. However, there is some correlation between both terms which is not easy to calculate. The most conservative option, i.e. maximal negative correlation (-1) will therefore be used. The result then reads as follows
\[ e^{T} \Sigma_{0}^{-1} e - (e - \mu_{1})^{T} \Sigma_{1}^{-1} (e - \mu_{1}) \sim N(\mu_{1}, \sigma_{1}^{2}) \]  

where

\[ \mu_{1} = m_{0} \frac{N}{m_{0}}^{2} \]

\[ \sigma_{1}^{2} = \left( \sqrt{2N} + \frac{1}{m_{0}} \sum_{i=m_{0}}^{1} \left[ 2 + 4d_{i}^{2} \right] \right)^{2} \]

Thus the statistical test (8) amounts to checking the following condition

\[ \frac{e^{T} \Sigma_{0}^{-1} e - (e - \mu_{1})^{T} \Sigma_{1}^{-1} (e - \mu_{1})}{\sigma_{1}} \sim c_{1} - \beta \tag{19} \]

If (19) holds true, presence of fault in command input is inferred. The threshold \( c_{1} - \beta \) can be determined as the confidence level of normal distribution at the degree of confidence 1-\( \beta \).

4. SIMULATED EXAMPLE

The model of the DC motor-generator process is used to demonstrate the proposed statistical test. The original model of the process is the 4th order transfer function (Juričić and Žele, 2002)

\[ G_{p}(q^{-1}) = \frac{0.1341q^{-1} + 0.015q^{-2} - 0.0316q^{-3} - 0.1153q^{-4}}{1 + 1.4414q^{-1} + 0.1563q^{-2} + 0.0666q^{-3} + 0.2187q^{-4}} \]

\[ T_{S}=0.05s \]

The process operates in a closed loop controlled by the PI controller

\[ C(q^{-1}) = \frac{1.23 - 1.2q^{-1}}{1 - q^{-1}} T_{S}=0.05s \tag{21} \]

To show the effectiveness of the statistical test (19) in the presence of modelling errors, the nominal process model is taken to be of the first order

\[ G_{m}(q^{-1}) = \frac{0.2253}{1 - 0.9748q^{-1}} T_{S}=0.05s . \tag{22} \]

The parameters of noise (\( \sigma_{e}^{2}=3.5 \times 10^{-7} \)) and undermodelling (\( \alpha=10^{-6}, \lambda=0.7 \)) were estimated offline during fault-free operation when the process was sufficiently excited by filtered white noise.

**Example 1**

Bias in command signal was introduced at \( t=36s \) by adding the offset \( f_{u}=0.01 \) (5% of the input signal value) to the controller output. The process output in this case is shown in Figure 1. Due to closed-loop operation, the effect of the actuator bias cannot be seen in the output signal. Figure 2 shows the residuals of the first-order nominal model. After the reference changes at \( t=17.5s \), a significant increase in the residuals can be detected. This is due to the fact that the first-order nominal model does not provide an accurate description of the process at high frequencies. Another rise in the residuals can be observed when the actuator bias comes into play (at \( t=36s \)). By using the proposed statistical test (19), false alarms, which can arise due to the effect of modelling errors, are avoided. Figure 3 illustrates the way the proposed test (19) works at the level of significance \( \beta=2\% \).

The statistical test can be interpreted as the sum of the residuals weighted by \( \Sigma_{0}^{-1} \) and \( \Sigma_{1}^{-1} \) and normalised by \( \mu_{1}, \sigma_{1} \). The test (19) does not trigger any alarm during the transient response since the modelling error is taken into consideration through the terms \( \Sigma_{0}^{-1}, \Sigma_{1}^{-1}, \mu_{1}, \sigma_{1} \), which depend on the modelling error as described by equation (6). However, in the time period from 21 to 36 seconds, a slight deviation from zero on the part of the test statistics can be observed. There are two reasons for this: poorly estimated process gain; and an over-optimistic estimate of modelling error. The statistical test increases significantly as soon as the fault appears and consequently exceeds the threshold, thus leading to an alarm (c.f. expression 19). The simulated example confirms that the statistical test is able to distinguish between the effect of the modelling error and the actuator fault.

**Example 2**

An interesting phenomenon can be observed for the actuator bias \( f_{u}=0.0025 \) introduced at \( t=36s \). Again, there is a step change in the reference at \( t=17.5s \). It is apparent that the effect of modelling errors and the effect of actuator fault tend to compensate each other, thus leaving an illusion of improved certainty of the predicted process behaviour. Consequently, the variance of the prediction error under a faulty regime might decrease compared to the variance of the prediction error under fault-free operation. The actuator fault in this case cannot be detected by using the statistical test (19) (see Figure 4).

**Example 3**

The performance of the test statistics in the case of a small negative bias \( f_{u}=-0.0025 \) (1.25% of the input signal value) has also been examined. Figure 5 depicts the test statistics for the actuator bias which appeared at \( t=32s \). Although the statistical test increases after fault occurrence, it does not exceed the threshold. This is a consequence of an over-conservative estimate of noise and undermodelling delivered by the stochastic embedding method. The sensitivity of the test statistics could be improved by using a less conservative description of the modelling error.

5. CONCLUSIONS

A statistical test is proposed for the robust detection of faults in actuator command inputs in the presence of modelling errors. For this purpose the stochastic embedding description of the modelling error is applied. The test amounts to checking whether a weighted and normalised quadratic sum of the residuals exceed a threshold. The weights, the mean value and the variance of the weighted quadratic sum
of residuals depend on the extent of undermodelling. The paper addresses concerns convergence conditions under which the test statistic can be approximately described by normal distribution. The experiments performed on the simulated DC motor-generator set-up show that the proposed statistical test is useful for detecting faults in command inputs in SISO systems while significantly reducing false alarms that arise due to modelling errors. Several issues remain for further research, one of which is accounting for the effects of unmodelled nonlineairity.

ACKNOWLEDGEMENT

The work is supported by the Slovenian Ministry of Science and Technology and partly by DAAD, Bonn (through the first author). Useful comments received from Martin Raič and anonymous referees are gratefully acknowledged.

REFERENCES


APPENDIX: BASIC CONVERGENCE RESULTS

**Theorem 1:** Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with $E(Y_k)=0$. If

$$\sum_{k=1}^n \text{var}(Y_k) < \infty$$

then the random sum $W_n = Y_1 + Y_2 + \ldots + Y_n$ converges with probability 1 as $n \to \infty$.

**Proof:** see Durrett (1995).

This theorem says that variables $W_n$ are well defined random variables. As the next step it is necessary to investigate when they converge.

**Definition 1:** A sequence of random variables $Y_1, Y_2, \ldots, Y_n$ is said to converge weakly (or converge in distribution) if the distribution functions $F_n(y)=P(Y_n \leq y) \to F(y)=P(Y \leq y)$ for all $y$ that are continuity points of $F$.

**Theorem 2 (Lindeberg, Feller):** For each $n$ let $Y_{nk}, 1 \leq k \leq n$ be i.i.d. random variables with $E(Y_{nk})=0$ and if the following conditions hold:

(i) $\lim_{n \to \infty} \sum_{k=1}^n \text{var}(Y_{nk}) = 1$

(ii) for all $\varepsilon > 0$

$$\lim_{n \to \infty} \sum_{k=1}^n E[Y_{nk}^2 \mid Y_{nk}] > \varepsilon = 0$$

then the sequence $W_n = \sum_{k=1}^n Y_{nk}$ weakly converges to the normal distribution, i.e.

$$\lim_{n \to \infty} P(W_n \leq w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^w e^{-t^2/2} dt$$

**Proof:** can be found in Durrett (1996), see also Rohatgi (1976).

The next result is a direct consequence of the Lindeberg-Feller theorem:

**Theorem 3:** Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with $E(X_i)=0$ and $\text{var}(X_i)=\sigma^2<\infty$. Let $\lambda_n \in R, 1 \leq n \leq n$ such that for all $n$

(i) $\sum_{k=1}^n \lambda_{nk} = \frac{1}{\text{var}(X_i)}$

(ii) $\lim_{n \to \infty} \max_{k \leq n} |\lambda_{nk}| = 0$

then the sequence $W_n = \sum_{k=1}^n \lambda_{nk} X_{nk}$ (A.1) weakly converges towards normal distribution $N(0,1)$.

**Proof:** Let us define $Y_{nk} = \lambda_{nk} X_{nk}$. Lindeberg condition applies for all $\varepsilon > 0$

$$\lim_{n \to \infty} \sum_{k=1}^n E(X_{nk}^2) / \text{var}(X_{nk}) \cdot \max_{k \leq n} |\lambda_{nk}| = 0$$

The last equality relies on the Lebesgue theorem on monotone convergence (see Durrett, 1996).

Having shown the convergence, it remains to answer the question on how fast does the sequence (A1) converge. First, one has to recall the fundamental result of Berry and Essen (Petrov, 1975):

**Theorem 4:** Let $Y_1, Y_2, \ldots, Y_n$ be a sequence of i.i.d. random variables with $E(Y_k)=0$, $\text{var}(Y_k)=\sigma^2<\infty$ and $E(|Y_k|^p) = p<\infty$. Let $F_n(W_n)$ denote the distribution function of the sum $W_n = Y_1 + Y_2 + \ldots + Y_n$. Then it follows that
An immediate consequence of this result can be formulated as follows

**Theorem 5:** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $E(X_1)=0$, $E(X_1^2)=\sigma^2<\infty$ and $E(|X_1|^3)=\rho<\infty$. Let $F_n(W_n)$ denote the distribution function of the sum $W_n=\lambda_1 X_1+\lambda_2 X_2+\ldots+\lambda_n X_n$. The assumption (ii) in theorem 3 is assumed to hold. Then it follows that

$$
F_n\left(W_n < w\right) - \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{w^2}{2\sigma^2}} \leq \frac{3\rho}{\sigma^3 \sqrt{n} \left| \sum_{i=1}^{n} \lambda_i \right|}
$$

Proof is omitted.