Advanced signal processing methods combined with automatic fault detection enable reliable condition monitoring for long periods of continuous operation. Rapid changes in acceleration become emphasized upon derivation of the signal $x^{(2)}$. Higher order derivatives, especially $x^{(a)}$, work very well in the whole range from slowly to very fast rotating rolling bearings. Real order derivatives $x^{(a)}$ provide additional possibilities. If $a < 0$, we can also talk about the fractional integration of displacement. It has interesting application potential for faults that occur in frequencies below the rotating frequency. The aim of the analysis is to introduce new features whose sensitivity is sufficient to detect faults from the absolute values of the dynamic part of the signals $x^{(a)}$ at an early stage, i.e. when the faults are still very small. Generalised moments $M^{(a)}_p$ and norms, $\| M^{(a)}_p \|_b = \left( \left( M^{(a)}_p \right)^{1/p} \right)^{1/b} = \left( \frac{1}{N} \sum_{i=1}^{N} |x^{(a)}|^{p} \right)^{1/p}$, can be defined by the order of derivation $(a)$, the order of the moment $(p)$ and sample time $(\tau)$, where $a$ and $p$ are real numbers. The number of signal values $N = \tau N_s$, where $N_s$ is the number of signal values taken in a second. The norm $\| M^{(a)}_p \|_b$ has the same dimensions as the corresponding signals $x^{(a)}$. Both the order of derivation and the order of the moment improve sensitivity to impacts. There are many alternative ways to normalise the moments: absolute mean deviation is an easy solution and the norm

$$\| M^{(a)}_p \|_b = \left( \left( M^{(a)}_p \right)^{1/p} \right)^{1/b} = \left( \frac{1}{N} \sum_{i=1}^{N} |x^{(a)}|^{p} \right)^{1/p}$$

has the same mathematical background as the corresponding signals $x^{(a)}$. The first time derivative of acceleration, known as jerk, has been used for assessing the comfort of travelling, for example in designing lifts, and for slowly rotating rolling bearings (1). Acceleration measurements have been performed more frequently with the introduction of accelerometers. The signals $x$ and $x^{(1)}$ can be detected successfully with $x$ and $x^{(1)}$. Acceleration measurements have been performed more frequently with the introduction of accelerometers. The signals $x$ and $x^{(1)}$ can be obtained from the $x^{(2)}$ signal through analogue or numerical integration (14, 15). It should be noted, however, that before accelerometers, the $x^{(2)}$ signal could have been generated by analogue derivation.

Higher order derivatives provide additional methods for vibration analysis (15). Lahdelma introduced the higher order derivatives in (16, 17). The first time derivative of acceleration, known as jerk, has been used for assessing the comfort of travelling, for example in designing lifts, and for slowly rotating rolling bearings (18). In (19) this was confirmed in practice and even better results were obtained with $x^{(1)}$. This is due to the fact that, although the acceleration pulses are weak and occur at long intervals, the changes in acceleration are rapid and become emphasised upon derivation of the signal $x^{(2)}$. Lahdelma has introduced the name $x^{(a)}$ for the fractional derivative.
’napse’ for the signal \( x^4 \). It is Finnish and means approximately ‘snap’ in English. Complex order derivatives introduced in \(^{[20]}\) offer additional possibilities for signal processing\(^{[18]}\). The integration of displacement has been introduced in \(^{[16]}\). Different approaches have been reviewed in \(^{[13]}\).

Statistical analysis provides various features for the signals: expectation and moments are used in developing features\(^{[21,22,23]}\). The central value alternatives are mean, median and mode. In vibration analysis, root mean square (rms) and peak values are the most commonly used features\(^{[24]}\). Dimensionless features are obtained by normalisation. Normalised moments skewness and kurtosis are widely used special cases corresponding to orders three and four, respectively. Probability distributions can be taken into account. If all the signal values are positive, the norms generalised to any real-valued order produce real-valued features. To obtain dimensionless features, these norms must be normalised. A generalised central moment introduced in \(^{[25]}\) works well, even with short sample times. The generalised norms introduced in \(^{[26]}\) have the same dimensions as the signal to be analysed.

Vibration indices based on several higher derivatives in different frequency ranges were already introduced in 1992\(^{[18]}\). Operating conditions can be detected with a case-based reasoning (CBR) type application with linguistic equation (LE) models and fuzzy logic. The basic idea of the LE methodology, which was introduced in 1991, is the non-linear scaling that was developed to extract the meanings of variables from measurement signals\(^{[27,28,29]}\). The combined approach has been summarised in \(^{[33]}\). Features extracted from higher derivatives \( x^\alpha \) and \( x^m \) have been used, for example, in cavitation indicators. The index obtained from \( x^k \) is the best alternative in this case but the index obtained from \( x^m \) also provides good results throughout the power range. The cavitation indicator also provides warnings of possible risk during short periods of cavitation\(^{[30]}\). Several model-based indicators based on features of acceleration and higher derivatives were compared in four frequency ranges in \(^{[31]}\). Cavitation can be detected already with the features in a fairly low frequency range. Methodologies have been developed in real-world applications, for example a Kaplan water turbine\(^{[25,30,31]}\), the supporting rolls of a lime kiln\(^{[32]}\) and rolling bearings in a very fast rotating centrifuge\(^{[29,33]}\).

This article combines signal processing and feature extraction to form a unified analysis methodology. More details about experimental systems and applications, which have been used in methodology testing, are presented in \(^{[34]}\).

2. Signal processing

2.1 Order of derivation

The history of fractional derivatives is very long. Leibnitz invented the denotation \( \frac{d^ny}{dx^n} \) for the derivative. He dealt with the question of what if \( n \) is fractional, and especially the case \( n = \frac{1}{2} \), in his letter\(^{[40]}\) to L’Hospital on 30 September 1695. In 1738, Euler studied the derivation of the power function\(^{[35]}\). Laplace defined, in 1812, a fractional derivative with an integral\(^{[12,36]}\).

In 1822, Fourier made the following generalisation:

\[
\frac{d^nf}{dx^n} f(x) = \left( \sum_{i} \lambda_i \int_{-\infty}^{\infty} f(x) \cos \left( px - \lambda_i x + \frac{\pi}{2} \right) dx \right) \frac{d^nf}{dx^n} f(x) + \sum_{i} \lambda_i \int_{-\infty}^{\infty} f(x) \sin \left( px - \lambda_i x + \frac{\pi}{2} \right) dx \frac{d^nf}{dx^n} f(x)
\]

where the index \( i \) can have any positive or negative values, \( i \in \mathbb{Z}^{\pm} \). Functions \( f(x) \) and \( f(a) \) are now written as \( f(x) \) and \( f(a) \), respectively. The change of phase angle is a linear function of the order of derivative, \( i \in \mathbb{Z}^{\pm} \).

In 1823, Abel applied fractional derivatives in a practical problem, which is called the tautochrone problem\(^{[4,6]}\). Heaviside used fractional derivatives with order \( \frac{1}{2} \) in 1896 when he studied electric cables\(^{[37]}\). Gemant found a connection between the half order derivative and the behaviour of viscoelastic materials. His studies were strongly based on the theory presented by Heaviside\(^{[6,39]}\). Although Gemant considered this derivative a useful mathematical symbol, he was unsure about the meaning of this in relation to the body as a whole.

Liouville first studied the functions:

\[
y = A e^{\alpha x} + A e^{\beta x} + A e^{\gamma x} + \ldots\]

where \( m \) is a real or complex number. In 1832, he defined the fractional derivatives of Equation (2) by:

\[
\frac{d^m y}{dx^m} = \sum A_i e^{\mu_i x}
\]

where the order of derivation \( \mu \) is a real number\(^{[3]}\). In the same article, he presents another definition based on integration. He mentions in the article\(^{[39]}\) from 1835 that the order \( \mu \) can be a complex number as well. The present algorithms for fractional derivatives are usually based on the definitions of Riemann in 1847\(^{[40]}\) and Grünwald in 1867\(^{[41]}\).

In 1917, Weyl published an examination of periodic functions \( f(x) \) with zero mean value\(^{[4]}\). To obtain the fractional integral \( j^\mu f(x) \) he used the complex Fourier series, where the coefficients \( c_\mu \) have the form:

\[
c_\mu = 0; \quad c_\mu = c_\mu e^{i\frac{\pi \mu}{2}} (2\pi)^{-\mu}
\]

The integer \( \nu \geq 1 \) and the real number \( a > 0 \). For derivative \( D^\nu f(x) \), the corresponding coefficients are:

\[
c_\nu = c_\nu e^{\frac{i\pi \nu}{2}} (2\pi)^{\nu}
\]

Lahdelma studied the function:

\[
x(t) = X e^{\alpha t}
\]

where \( 0 < \omega \in R \) and \( X \) is a constant, \( t \) is a real variable and \( i = \sqrt{-1} \). In 1997\(^{[42]}\), he defined a real order derivative \( x^\alpha \) by:

\[
x^\alpha = \omega^\alpha X e^{\frac{a+ia t}{2}}
\]

which means that \( x^\alpha = (i \omega)^a x \). In 1999, Lahdelma introduced a definition for the complex derivative \( x^\alpha \) of (6) by:

\[
x^\alpha = x^{(a+bi)} = (i\omega)^a X e^{\alpha t}
\]

where \( z = a + bi \) and \( a, b \in R^{\pm} \).

The definition \(^{[8]}\) of Lahdelma and the definition \(^{[3]}\) of Liouville are identical when \( \mu = a + bi, m = u + i \omega \) and we assume \( u \) to be zero. An interesting result of (8) is that the phase
angle changes with \( \varphi_a + \varphi_p = \frac{\alpha}{2} \omega t + \beta \ln \omega \), when the order of
derivation is a complex number \( z \). We find that \( \varphi_p = \frac{\alpha}{2} \pi \) does
not depend on \( \omega \), and \( \varphi_a = \beta \ln \omega \) increases with increasing \( \omega \),
see[14,26,43]. The amplitude is:
\[
X_z = e^{\frac{\alpha}{2} \omega t} \omega^z X \tag{9}
\]

The coefficients \((4)\) and \((5)\) introduced by Weyl are obtained as special cases of \((7)\), when \( \gamma \) has integer values in \( \omega = 2\pi f \) and \( \alpha \) is positive or negative. The real order derivative \( x^{(z)} \) is a special case of \((8)\), where \( \beta = 0 \).

The calculation of the time domain signal \( x^{(z)}(t) \), which is based on a rigorous mathematic theory\((10)\), is performed in three steps. The fast Fourier transform (FFT) is used for the
displacement signal \( x(t) \) to obtain the complex components \( \{X_k\} \), \( k = 0, 1, 2, \ldots, (N-1) \). The corresponding components of the
derivative \( x^{(z)}(t) \) are calculated as follows:
\[
X_{zk} = (i \omega t)^{z-k} X_k \tag{10}
\]

Finally, the resulting sequence is transformed with the inverse
Fourier transform \( FFT^{-1} \), which produces the signal \( x^{(z)}(t) \). Since
the vibration analysis is now based on the acceleration signals,
the components of the derivative are obtained with:
\[
X_{ak} = (i \omega t)^{\alpha-k} X_k \tag{11}
\]

where the complex components \( \{X_k\} \) are calculated from the
acceleration signal \( x^{(1)} \). The fast Fourier transform is explained
in\((40)\). To obtain complex derivatives we use \( z = \alpha + \beta i \) instead of \( \alpha \).
Alternatively, the derivatives can be calculated in the
three steps. The fast Fourier transform (FFT) is used for the
vibration analysis is now based on the acceleration signals,
with \( x(t) \) expressed as a complex number
angle changes with \( \gamma \), see\((14,20,43) \). The amplitude is:
\[
X_{z,0} = e^{\frac{\alpha}{2} \omega t} \omega^z X \tag{12}
\]

The real order derivative for sinusoidal signals
\( x = x(t) = X \sin \omega t \) is:
\[
\frac{d^x x}{dt^x} = X \omega^x \sin \left( \omega t + \frac{\alpha}{2} \pi \right) \tag{13}
\]

where \( \alpha \) is a real number, the amplitude \( X = \omega X \) and the change of phase angle \( \varphi_a = \frac{\alpha}{2} \pi \). The velocity and acceleration are
special cases of \((12)\):
\[
\frac{dx}{dt} = x^{(1)} = \omega X \sin \left( \omega t + \frac{\alpha}{2} \pi \right) \tag{14}
\]

Some faults, such as unbalance, misalignment, bent shaft and
mechanical looseness, can be detected by means of displacement
and velocity, ie signals \( x = x^{(0)} \) and \( x = x^{(1)} \). On the other hand,
rolling bearing faults or faults in gears, as well as cavitation, can be
detected more efficiently with the acceleration signal. For these
faults, sensitivity of the features increases to some limit with the
increasing order of derivative. More examples are presented in\((34)\).
Higher order derivatives provide more sensitive solutions, ie the ratios of features between the faulty and non-faulty cases become
higher. Additional flexibility can be achieved with the real order of
derivation\((17,42,45) \). The derivatives \( x^{(n)} \) and \( x^{(1)} \) have been used in
condition monitoring\((14,42,45) \), and \( x^{(0)} \) also in active damping\((37) \).

Fractional integration of the displacement has an interesting
application potential. The negative order of derivation, ie integration, amplifies the amplitudes at a low frequency. For example, derivation with order \(-1.1\) results in:
\[
\frac{d^{(-1.1)} x}{dt^{(-1.1)}} = x^{(-1.1)} = \omega^{(-1.1)} \sin \left( \omega t - 1.1 \frac{\pi}{2} \right) \tag{15}
\]

2.2 Frequency range
For sinusoidal signals with constant amplitude \( X \), the derivation results show an increase in the amplitude \( X \), with \( \omega > 1 \) if \( a > 0 \), see \((7)\). The increase is linear in a log-log scale. Correspondingly,
decreasing lines result if \( a < 0 \). The derivation does not change the amplitude if \( a = 0 \). If the allowed vibration levels on the logarithmic scale form ascending or descending lines over certain frequency ranges, there is a risk of major errors when interpreting root mean square (rms) values in assessing fault severity. The errors can be
eliminated with \( x^{(0)} \) signals. This problem is discussed in more
detail in\((17,42) \).

3. Feature extraction
Feature extraction is based on velocity \( x^{(3)} \), acceleration \( x^{(2)} \)
and higher derivatives \( x^{(5)} \) and \( x^{(4)} \). The other signals have been
obtained from acceleration through analogue\((30) \) or numerical
integration and derivation\((31) \).

3.1 Statistical features
The mathematical expectation, expected value, or briefly the
expectation, of a random variable is a very important concept
in probability and statistics. For a discrete random variable \( X \), having possible values \( x_i, i = 1, \ldots, N \), the probability of each value is defined by a probability function \( P(X = x_i) \). A vibration signal
\( x(t) \) is a sample with a length of \( T \) from a continuous variable
\( X \). Values are selected from the signal with an equal-sized time
interval \( T/(N-1) \). The probability of each value is defined by a
density function \( f(x) \), ie \( P(X = x) = f(x) \), and the expectation of
\( X \) is defined by:
\[
E(X) = \sum_{i=1}^{N} x_i P(X = x_i) = \sum_{i=1}^{N} x_i f(x_i) \tag{16}
\]

Additional random variables can be constructed by means of
the function \( Y = g(X) \), which is integrable between \((\infty, \infty) \). The expectation of \( Y \) is:
\[
E(Y) = E[g(X)] = \sum_{i=1}^{N} g(x_i) P(X = x_i) = \sum_{i=1}^{N} g(x_i) f(x) \tag{17}
\]

For a continuous random variable \( Y \), the expectation is
defined by:
\[
E(Y) = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \tag{18}
\]

Note that this does not involve the probability density function of
\( Y = g(X) \).

The mean is the first moment about the origin. Higher
moments \( M^k \) about the origin can be defined as expectations
\( E(X^k) \). The means can also be defined about some central
value \( c \in R \), for example the moments about the mean defined by:
\[
M^k = E[(X - E(X))^k] \tag{19}
\]
where \( k \) is a positive integer. Other central values are medians and modes, for example. The second moment about the mean, known as the variance \( \text{Var}(X) = \sigma_X^2 \), is a measure of dispersion, or scatter, of the values of the random variable near the mean. The variance can be represented by:

\[
\sigma_X^2 = M_2 = E\left[ (X - E(X))^2 \right] = E\left( X^2 \right) - \left[ E\left( X \right) \right]^2 \tag{20}
\]

The positive square root of the variance, \( \sigma_X \), or briefly \( \sigma \), which is called the standard deviation, is used more often since it has the same dimension as the variable. Another measure of dispersion is the mean absolute deviation (MAD), or briefly mean deviation (MD), which is defined as the expectation of \( |X - E(X)| \).

Dimensionless features can be obtained by normalising the moments (19), for example by standard deviation \( \sigma_X \):

\[
\gamma_i = \frac{E\left[ (X - E(X))^i \right]}{\sigma_X^i} \tag{21}
\]

The feature \( \gamma_i \) is called the coefficient of skewness, or briefly skewness, and the feature \( \gamma_i \) is the coefficient of kurtosis. The skewness is a measure of asymmetry: \( \gamma_1 = 0 \) for a symmetric distribution. If \( \gamma_i > 0 \), the skewness is called positive skewness and the distribution has a long tail to the right, and \textit{vice versa} if \( \gamma_i < 0 \). The kurtosis is a measure of the concentration of the distribution near its mean. For a Gaussian signal \( \gamma_i = 3 \). Flatter, also described as long-tailed or heavy-tailed, distributions have \( \gamma_i < 3 \) and for distributions with high peakedness \( \gamma_i > 3 \). For a sinusoidal signal \( \gamma_i = 1.5 \). An alternative definition of kurtosis reduces the ratio by 3 to give the value zero for the normal distribution.

In feature extraction, we assume that all the values are equally probable, \( ie \) the probability distribution function is constant:

\[
P(x = x_i) = \frac{1}{N} \text{ for all } x_i, \ i = 1, \ldots, N.
\]

The sample size \( N \) has an important effect on the calculations. The features can also be obtained from sampling distributions. The unbiased population variance is estimated by the sample variance:

\[
s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{22}
\]

which is based on the pairwise analysis of the differences of the values \( x_i, \ i = 1, \ldots, N \). Unbiased variance is needed for small samples. For large samples, for example \( N > 100 \), the denominator value \( N \) can be used.

In vibration analysis, the variable is displacement \( x_i \), or more often its derivatives \( x^{(i)} \) and \( x^{(2)} \). We can also use \( x^{(i)} \), where \( a \in R \) is a real number. Standard deviation can be obtained from signals \( x^{(i)} \):

\[
\sigma_a = \left[ \frac{1}{N} \sum_{i=1}^{N} (x^{(i)} - \bar{x}^{(i)})^2 \right]^{1/2} \tag{23}
\]

and features normalised by the standard deviation, such as kurtosis:

\[
\gamma_a = \frac{1}{N \sigma_a} \sum_{i=1}^{N} \left( x^{(i)} - \bar{x}^{(i)} \right)^4 \tag{24}
\]

where \( \bar{x}^{(i)} \) is the arithmetic mean of the signal values \( \{x^{(i)}\} \).

The aim of condition monitoring is to detect faults at an early stage and vibration analysis usually deals with the dynamic part of the signal. In a sufficiently long signal, for example, the mean value of signals \( x^{(1)} \) and \( x^{(2)} \) is zero. Otherwise, the measurement point, for example the bearing housing, would move away from the machine. The assumption \( \bar{x}^{(0)} = 0 \) means that the root mean square of \( x^{(i)} \) is equal to the standard deviation: \( x^{(i)}_{\text{rms}} = \sigma_x \).

### 3.2 Generalised moments

Features can be calculated by means of a generalised absolute moment about the origin:

\[
\left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \tag{25}
\]

where the real number \( a \) is the order of derivation, the real number \( p \) is the order of the moment and \( \tau \) is the sample time, \( ie \) the moment is obtained from the absolute values of signals \( x^{(i)} \). The number of signal values \( N = \tau N_s \), where \( N_s \) is the number of signal values which are taken in a second. Alternatively, the signal values \( x^{(i)} \) can be compared to the mean \( \bar{x}^{(i)} \):

\[
\left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \tag{26}
\]

The sample time \( \tau \) connects the moments to the control applications. The mean of the signal does not fluctuate considerably from sample to sample if the sample time is long enough.

The generalised central absolute moment about \( c = \bar{x}^{(a)} \) can be normalised by means of the standard deviation \( \sigma_a \) of the signal \( x^{(a)} \):

\[
\left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \tag{27}
\]

which was presented in(23). The peaks of the signal have a strong effect on the moments (25), (26) and (27). The moment (27) can be used in the same way as kurtosis(23). The moments calculated for higher order derivatives \( x^{(3)} \) and \( x^{(4)} \) are more sensitive to impacts than the ones calculated for velocity. The sensitivity of the moment improves when the order \( p \) of the moment increases.

Features (27) can be understood as moments of standardised random variables. The moment \( \left. \frac{1}{N} \right) \left. \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \) is the central moment of order \( p \) of the signal. The standard deviation \( \sigma_a \) can be obtained from \( \left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \) by taking the square root:

\[
\sigma_a = \left( \left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right|^p \right) \right)^{1/2} \tag{28}
\]

There are many alternative ways of normalisation, for example we can use the absolute mean deviation or briefly mean deviation:

\[
\left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} - \bar{x}^{(i)} \right| \right) \tag{29}
\]

### 3.3 Generalised norms

A norm defined by:

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} \right|^p \right\| = \left( \left. \frac{1}{N} \sum_{i=1}^{N} \left| x^{(i)} \right|^p \right) \right)^{1/2} \tag{30}
\]

where \( p \neq 0 \) was introduced in(28). This is the \( l_p \) norm of \( x^{(a)} \) and we can write it in an alternative way:
It has the same dimensions as the corresponding signals $x^{(a)}$. The $l_p$ norms are defined in such a way that $1 \leq p < \infty$. In this study, the order $p$ is allowed to be less than one. If $p < 1$ then (30) is a quasinorm because it violates the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$. The norm (30) combines two trends: a strong increase caused by the power $1/p$. For the order $p = 1$, there is no amplification. The significance of the highest peaks will decrease if $p < 1$.

The absolute mean:

$$\|x^{(a)}\|_a = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i^{(a)}|^p\right)^{1/p}$$

and the rms value:

$$\|x^{(a)}\|_{\text{rms}} = \left(\frac{1}{N} \sum_{i=1}^{N} |x_i^{(a)}|^p\right)^{1/2}$$

are special cases of (31). We also obtain from (31) the absolute harmonic mean:

$$\|x^{(a)}\|_h = \frac{N}{\sum_{i=1}^{N} \frac{1}{|x_i^{(a)}|^p}}$$

if the order $p = -1$ and $|x_i^{(a)}| \neq 0$.

The norm (30) is a Hölder mean, also known as the power mean. The norm values increase with increasing order, i.e. for the $l_1$ and $l_p$ norms holds:

$$\|M_{a}^p\|_p \leq \|M_{a}^q\|_q$$

if $p < q$. The increase is monotonous if all the signals are not equal.

Different weights can be introduced to the calculations by the density function, for example:

$$f(|x_i^{(a)}|) = \frac{w_i}{\sum w_i}$$

where $w_i$ is the weight of the value $|x_i^{(a)}|$.

The rms values of displacement and velocity can be obtained from (33) by using the values zero and one for the order $a$, respectively:

$$\|x\|_r = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^2\right)^{1/2}$$

and:

$$\|v\|_r = \left(\frac{1}{N} \sum_{i=1}^{N} v_i^2\right)^{1/2}$$

These norms can be used for detecting unbalance, misalignment, bent shaft and mechanical looseness. For rolling bearing faults, displacement and velocity should be replaced by acceleration or higher derivatives. For a very low rotation speed, the rms values are not sensitive for bearing faults because the effect of a few weak impacts is small in the sum (33), where $N$ is a large number. For high rotation speeds, frequent strong impacts affect rms values significantly. The rms values can then be used to detect rolling bearing faults, especially if $a \geq 2$.

The computation of the norms can be divided into the computation of equal-sized sub-blocks, i.e. the norm for several samples can be obtained as the norm for the norms of individual samples. The same result is obtained using the moments:

$$\|M_{a}^p\|_p = \left(\frac{1}{K} \sum_{i=1}^{K} \|M_{a}^p\|_i\right)^{1/p}$$

where $K$ is the number of samples $\{x^{(a)}_{i}\}_{i=1}^{K}$. Each sample has $N$ signal values. Weights can be introduced by means of densities functions. It is useful to calculate the norms from short samples since the number of signal values per second is quite high. The sample time $\tau$ is an essential parameter in the calculation of moments and norms.

When the order $p \to 0$, we obtain from (30) the absolute geometric mean, i.e.:

$$\lim_{p \to 0} \|x^{(a)}\|_p = \|x^{(a)}\| = \left(\prod_{i=1}^{N} x_i^{(a)}\right)^{1/N}$$

If the $f$-mean $M_{f}$ of the absolute values $|x^{(a)}|$ is calculated with a logarithmic function, $f(x) = \ln(x)$, the result is the same:

$$M_{f}(|x^{(a)}|) = f^{-1}\left(\frac{1}{N} \sum_{i=1}^{N} f(|x_i^{(a)}|)\right) = \exp\left(\ln\left(\prod_{i=1}^{N} x_i^{(a)}\right)^{1/N}\right) = \left(\prod_{i=1}^{N} x_i^{(a)}\right)^{1/N}$$

The exponential function is the inverse function of $\ln(x)$. The norm (31) represents the norms from the minimum to the maximum, which correspond to the orders $p = -\infty$ and $p = \infty$, respectively. When $p < 0$, all the signal values should be non-zero, i.e. $|x_i^{(a)}| \neq 0$. Therefore, the norms with $p < 0$ are reasonable only if the signal values near the zero are removed.

Rolling bearing faults in slowly rotating machinery can be detected with the maximum norm presented by:

$$\lim_{p \to \infty} \|x^{(a)}\|_p = \|x^{(a)}\| = \max_{i=1,...,K} x_i^{(a)}$$

which in diagnostics is called peak value. An efficient solution is to use peak values $x_{\text{peak}}^{(a)}$, $x_{\text{peak}}^{(1)}$ and $x_{\text{peak}}^{(a)}$. To avoid the domination of a distinct peak, the peak value can be calculated, for example, as an average of the highest three peaks.

A feature can also be defined as a maximum of the norms $\|M_{a}^p\|_p$ calculated from different samples $i = 1,...,K$, i.e.:

$$\max_{i=1,...,K} \|M_{a}^p\|_p = \max_{i=1,...,K} \left(\|M_{a}^p\|_i\right)$$

The number of signal values in each sample is equal and defined by the sample time and the number of signal values in a second.

A weighted sum of the norms obtained for different orders of derivatives represented by:

$$\|k_{a} = \sum_{a=p}^{N} w_{a} \|x^{(a)}\|_p$$

is a norm as well. In a special case, integer orders can be used: $a_{1} = 0, a_{2} = 1, ..., a_{n} = n$.

Each norm has a weight factor $w_{a}$ to make the sum dimensionless. Several individual faults or a combined fault can be detected. Also, the importance of different terms can be dealt with using the weight factors. Dimensionless vibration indices introduced in 1992 belong to norms of this type, see $\|x^{(a)}\|$.
inverse of the norm $\|x\|_{\alpha}$ provides an indication of the health of the machine.

### 3.4 Signal distribution

The bins $F_{k}^{(a)}$ of the histograms can be based on the standard deviation $\sigma$ of the corresponding signal $x^{(a)}$, for example, in the following way: $(k=1) |x^{(a)}| < 2\sigma$, $(k=2) 2\sigma \leq |x^{(a)}| < 3\sigma$, $(k=3) 3\sigma \leq |x^{(a)}| < 4\sigma$, $(k=4) 4\sigma \leq |x^{(a)}| < 5\sigma$ and $(k=5) |x^{(a)}| \geq 5\sigma$, where $a$ is the order of derivation. Large values for the features $\sigma$, and the fractions $F_{k}^{(a)}$, $k=4$ and $5$, are related to faulty situations, and large values for the fractions $F_{k}^{(a)}$, $k=1 \ldots 3$, are obtained in normal conditions. Similar results can be obtained with bins defined by the absolute average of the signals, and the resulting easier calculation is useful for developing intelligent sensors.

The bins can be equally well based on the mean deviation, which requires fewer calculations than the standard deviation. The fractions related to the highest bins have been found to be useful features for detecting strong impacts. Any norm $\|M_{\alpha}^{x}\|$, with a higher order $p$ could provide a suitable definition for the highest bins in these cases.

### 4. Condition indices

#### 4.1 Measurement indices

Vibration signals can be utilised in process or machine operation by combining features obtained from derivatives. Dimensionless vibration indices can be combined in a measurement index:

$$\text{MIT}^{\alpha}_{a_{1}a_{2} \ldots a_{n}} = \frac{1}{n} \sum_{i=1}^{n} b_{a_{i}} \frac{\|x^{(a_{i})}\|_{\alpha}}{\|x^{(a_{i})}\|_{\alpha,0}} \tag{45}$$

where the norms $\|x^{(a)}\|_{\alpha}$ are obtained from the signals $x^{(a)}$, $i = 1, \ldots, n$. Each norm is divided by its reference value, denoted by index zero, and multiplied by a weight factor $b_{a_{i}}$. The sum $\sum_{i=1}^{n} b_{a_{i}} = n$. The reference values correspond to the good conditions. In (45) all the features were based on the rms values of $x^{(a)}$, ie norms $\|x^{(a)}\|_{\alpha} = \|M_{\alpha}^{x}\|$. Alternatively, peak values or kurtosis can be used for some norms corresponding to indices $\text{MIT}^{\alpha}_{a_{1}a_{2} \ldots a_{n}}$ and $\text{MIT}^{\alpha}_{a_{1}a_{2} \ldots a_{n}}$, respectively. Measurement indices obtained by using different orders $a_{i}$ and $p$, for example:

$$\text{MIT}^{\alpha}_{a_{1}a_{2} \ldots a_{n}} = \frac{1}{2} \left[ \text{MIT}^{\alpha}_{a_{1}} + \text{MIT}^{\alpha}_{a_{2}} \right] = \frac{1}{2} \left[ \frac{\|x^{(a_{1})}\|}{\|x^{(a_{1})}\|_{\alpha,0}} + \frac{\|x^{(a_{2})}\|}{\|x^{(a_{2})\|_{\alpha,0}} \right] \tag{46}$$

provide good results by combining two specialised features: the rms value reacts to changes in the overall signal levels and the peak value to high impacts. Indices (46) are special cases of (45): we use the norms $\|M_{\alpha}^{x}\|$ and $\|M_{\alpha}^{x}\|$. The importance of the features is defined by the weight factors $b_{a_{i}}$, which can be adjusted on the basis of accumulated knowledge in each application case.

The inverse of the index MIT, denoted as SOL, provides a direct indication of the condition of the machines: small values indicate poor condition and high values good condition. The machine is in good condition if $\|x^{(a)}\|_{\alpha} = \left(\|x^{(a)}\|_{\alpha,0}\right)^{-1}$ for all the terms in (45). In this case SOL = MIT = 1, and weakening health is seen when the SOL index decreases from one. These indices are not restricted to vibration signals. The generalised health index (Figure 1) can combine several measurements:

$$\text{SOL} = \text{SOL} \text{ (vibration, pressure, temperature, electric current, rotation speed, ...)}.$$  

#### 4.2 Intelligent condition and stress indices

The analysis can be further improved by taking into account non-linear effects. The scaling function scales the real values of variables to the range of $[-2,2]$ with two monotonously increasing functions: one for the values between $-2$ and $0$, and one for the values between $0$ and $2$. The scaling function $f$ consists of second-order polynomials and the scaled values, which are called linguistic levels, are obtained by means of the inverse function $f^{-1}$, see (30). The polynomials should be monotonous increasing functions in order to produce realisable systems: constraints are presented in (30). All the variables brought to the same scale are analysed with linear methods. For process measurements, the scaling function combines normal operation $[-1,+1]$ with the handling of warnings and alarms. In condition monitoring, non-linear scaling has been used for statistical features and features based on the signal distribution.

For each feature, the level 0 can be obtained as a median of the values in the training set and the levels $-1$ and $1$ as medians of the lower and higher halves of the values, respectively. In (30), the cavitation index is a stress index obtained by scaling the relative value of the norm (43):

$$I_{\alpha}^{(a)} = f_{\alpha}^{(a)} \left( \text{relative max} \left( \|M_{\alpha}^{x}\| \right) \right) \tag{47}$$

In this case, only one feature is needed if the orders $p$ and $a$ are chosen properly. The same results can be obtained by scaling the norm max $\|M_{\alpha}^{x}\|$ defined by (43), since only the parameters of
the scaling function will change. Low stress corresponds to value –2 and not allowable stress to value 2. Operating conditions with high stress should be avoided.

A condition index can be based on several features, which are all scaled to the range [–2,2]:

$$IC_k = \sum_{i=1}^{n} w_k f_i^{-1} \left[ \max \left( \| M_i \| \right) + \sum_{j=1}^{m} w_j f_j^{-1} (F_j^{(a)}) \right]$$

where \( w_k \) is the coefficient and \( f_i^{-1} \) the scaling function of the feature \( k \). Features include maximum norms \( \| M_i \| \), and other features \( F_j^{(a)} \), for example bins of the histograms (Section 3.4). Features can have specific frequency ranges. The number of features was six in a lime kiln application(30) and three for very fast rotating rolling bearings(33). Index (48) is obtained from the features of the signal \( x(t) \), but an index can also combine the features of different signals as in (45) and (46). Good conditions correspond to value 2 and not allowable conditions to value –2. Condition indices (47) and (48) can also combine several measurements (Figure 1). The condition is impaired if the equipment is operated in a high stress area for a long time.

### 5. Applications

Methodologies have been developed in experimental systems(45,47), the gearbox of a seawater pump(26), a water turbine(23,26,28,30) and a roller bearing of a centrifuge(33), and several applications in pulp mills, including the bottom scraper of a continuous digester(47), the supporting rolls of a lime kiln(32,46) and a roller bearing in a washer(33). Short sample times and relatively small requirements for frequency ranges make the generalised norms and condition indices feasible for online applications to detect impacts. Even low frequency ranges 10-1000 Hz are useful for \( x(t) \) in many cases where the fault causes impacts. A good frequency range is well adapted to process phenomena and the signal that is analysed. Examples of paper machine applications were presented in(15). Features from different frequency ranges can be adapted to process phenomena and the signal that is analysed.

Higher frequency ranges are needed for very fast rotating rolling bearings(29,33). Features can have specific frequency ranges. The number of features was six in a lime kiln application(30) and three for very fast rotating rolling bearings(33). Index (48) is obtained from the features of the signal \( x(t) \), but an index can also combine the features of different signals as in (45) and (46). Good conditions correspond to value 2 and not allowable conditions to value –2. Condition indices (47) and (48) can also combine several measurements (Figure 1). The condition is impaired if the equipment is operated in a high stress area for a long time.

### 6. Conclusions

Some faults can be detected from displacement and velocity signals by means of rms and peak values. For faults causing impacts, the sensitivity increases with the order of derivation to some limit, and the signal \( x(t) \) is a good alternative in many applications. Real order derivation and integration introduce additional possibilities. High sensitivities to small faults and anomalies make generalised moments and norms informative features for early fault diagnosis. Moments need to be normalised in order to obtain dimensionless features. The \( f_i^{-1} \) norms have the same dimensions as the corresponding signals. Non-linear effects can be taken into account by scaling functions, and several features are combined in condition, measurement and health indices. In some cases only one norm is needed. The generalised norm can be defined by the order of derivation, the order of the moment and sample time. The low order of derivatives can be compensated for by using higher order moments.

### References

The International Journal of Condition Monitoring | Volume 1 | Issue 2 | November 2011