Optimal filter design for rotating machinery fault detection under time-varying speed conditions

Stephan Schmidt[1,*], Daniel N. Wilke[1], P. Stephan Heyns[1], Konstantinos C. Gryllias[2,3]

[1]Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
[2]Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300, 3001 Heverlee, Belgium
[3]Flanders Make @ KU Leuven, Belgium
[*]stephan.schmidt@up.ac.za

Abstract

In vibration-based condition monitoring, the damaged signal components are often masked by extraneous components and need to be extracted or enhanced to perform detection, identification, and trending. This can be performed using optimal filters. In this work, the optimal filtering problem is presented using different optimisation formulations, objective functions, and optimisation algorithms. The performance of the methods for fault detection, and trending are compared under time-varying speed conditions. The work demonstrates that the objective function, optimisation formulation and optimisation algorithm can have a significant impact on the performance of the methods and should be carefully selected for vibration-based condition monitoring applications under time-varying speed conditions.

1   Introduction

Vibration-based condition monitoring is one of the most popular condition monitoring methods for rotating machinery [1]. However, the fault signatures are often masked by extraneous signal components (e.g., dominant gear mesh interactions) and their detection is impeded by time-varying operating conditions. Many critical assets in the power generation and mining industries operate under time-varying operating conditions and therefore it is important to develop and investigate methods that can enhance the damage information under these conditions.

Digital filters can be used to extract the information of interest and to attenuate the extraneous components [2]. In blind deconvolution, this is achieved by designing an inverse filter that aims to attenuate the transmission path effects from the signal so that the original signal can be recovered [2]. Since we do not know the original source, this filter is obtained by maximising some measure of the filtered signal that is sensitive to the component of interest, e.g., kurtosis [3, 2], correlated kurtosis [4], and the indicator of second-order cyclostationarity (ICS2) [2, 5]. The L2/L1-norm [6], spectral negentropy [6] and the Hoyer index [6] of the envelope spectrum are some of the features that also has been used to design blind filters. Even though blind filtering follows the same process as blind deconvolution, blind filters do not deconvolve the signals, but enhance the components of interest [6]. Therefore, not all features (and resulting filters) can be used for blind deconvolution; some of the features only allow the damaged components to be enhanced. Since all the methods optimise filters that can improve the detectability of the damage, we refer to all the methods as optimal filtering methods in this work.

The performance of the optimal filtering methods depends on the feature that is maximised and the optimisation algorithm that is used. In previous work, the objective function method [3], the eigenvector algorithm to maximise the higher order statistics [3], the maximisation of the Generalised Rayleigh Quotient (GRQ) using a generalised eigenvalue decomposition [2], gradient-descent methods [7] and global optimisation methods [8] have been used to find the optimal filter coefficients. The GRQ method has been very popular in the past five years, e.g., it has been used to maximise the L2/L1-norm [6]; the Hoyer index [6]; the ICS2 [2, 5]; the spectral negentropy [6]; and the Gini index [9]. However, only limited papers evaluated and compared the performance of the different objective functions under varying speed conditions (e.g., [2, 5]).

It is, therefore, important to investigate the performance of different optimal filter design formulations for fault detection under time-varying speed conditions. More specifically, three optimisation formulations, namely, unconstrained gradient-descent, unconstrained gradient-descent with a normalised impulse response function and solving the optimality criterion problem in the form of a generalised Rayleigh quotient, are compared. The performance of the methods is quantified for fault detection under time-varying speed conditions.

Furthermore, two formulations of the ICS2 are compared (i.e., the formulation used in Refs. [2, 5] and a new formulation for optimal filtering); two formulations of the L2/L1 of the cyclic spectrum (i.e., the formulation used in [6] and a new formulation for optimal filtering); and the SES negentropy are compared. We evaluate the sensitivity of the methods to damage and the sensitivity of the methods to varying speeds. Subsequently, we use three performance measures to quantify the suitability of the methods for fault detection under varying speed conditions.

In Section 2, an overview of optimal filter design is presented, whereafter, three investigations are performed in Section 3 to compare the methods. Finally, the work is concluded in Section 4.

2 Optimal filter design

In this section, the optimal filter design approach is presented. Firstly, the signal processing procedure is discussed in Section 2.1, whereafter different optimisation formulations are presented in Section 2.2 and different objective functions are presented in Section 2.3. Lastly, the gradient-based optimisation methods considered in this work are discussed in Section 2.4.

2.1 Signal processing

Consider the measured vibration signal $\mathbf{x} \in \mathbb{R}^L$, with $\mathbf{x} = [x[1], x[2], \ldots, x[L]]$. The filtered signal $\mathbf{s} \in \mathbb{R}^D$ is obtained by convolving the signal $\mathbf{x}$ with a finite impulse response filter $\mathbf{h} \in \mathbb{R}^N$. More specifically, the signal $\mathbf{s}$ is calculated as follows [2]:

$$\mathbf{s} = \mathbf{x} \otimes \mathbf{h} = \mathbf{X}\mathbf{h} \tag{1}$$

where $\otimes$ denotes the convolution operator and $\mathbf{X} \in \mathbb{R}^{D \times N}$ is constructed from $\mathbf{x}$, with $D = L - N$ [2]. The Discrete Fourier Transform (DFT) of the filtered signal is defined by $\mathbf{d}_s = \mathbf{F}\mathbf{s}$ where $\mathbf{F} \in \mathbb{C}^{D \times D}$ is the DFT matrix of $\mathbf{s} \in \mathbb{R}^D$. The discrete Fourier transform of the squared signal is calculated as follows:

$$\mathbf{d}_{s^2} = \mathbf{F}(\mathbf{s} \odot \mathbf{s}) \tag{2}$$

where $\odot$ is the Hadamard multiplication operator, e.g., $\mathbf{x} \odot \mathbf{y} = [x[1] \cdot y[1], x[2] \cdot y[2], \ldots, x[L] \cdot y[D]]$ where $\mathbf{x} \in \mathbb{R}^D$, $\mathbf{y} \in \mathbb{R}^D$. The squared envelope spectrum is estimated from $\mathbf{b}_{s^2}^{SES} = \mathbf{d}_{s^2}^* \odot \mathbf{d}_{s^2}$ and the envelope spectrum is estimated from $\mathbf{b}_{s^2}^{ES} = |\mathbf{d}_{s^2}| = \sqrt{\mathbf{b}_{s^2}^{SES}}$, where $\mathbf{d}_{s^2}^* \in \mathbb{C}^D$ denotes that the element-wise complex conjugate of $\mathbf{d}_{s^2} \in \mathbb{C}^D$ is calculated.

We can replace the conventional DFT matrix $\mathbf{F}$ with the velocity synchronous DFT $\mathbf{V}$ [10, 2] in the spectrum calculations (e.g., equation (2)) to calculate the spectra in the order domain without performing computed order tracking in the time domain. The results obtained with the velocity synchronous DFT, i.e., $\mathbf{V}$, to calculate the spectra, are shown in this work. The filtered signal $\mathbf{s}$ and its envelope spectra, e.g., $\mathbf{b}_{s^2}^{SES}$ are functions of the filter $\mathbf{h}$, but this dependence is not explicitly shown to simplify the notation.

2.2 Optimisation formulations for filtering

The optimal filtering problem is formulated as follows:

$$\max_{\mathbf{h}} \psi(\mathbf{h}; \mathbf{x}), \text{ subjected to } \mathbf{h}^T\mathbf{h} = 1 \tag{3}$$

where the objective function is decomposed as follows: $\psi(\mathbf{h}; \mathbf{x}) = \frac{\eta(\mathbf{h}; \mathbf{x})}{\beta(\mathbf{h}; \mathbf{x})}$, with $\psi : \mathbb{R}^N \mapsto \mathbb{R}$, $\eta : \mathbb{R}^N \mapsto \mathbb{R}$ and $\beta : \mathbb{R}^N \mapsto \mathbb{R}$. Since the measurement is fixed during the optimisation process, we simplify the notation and use $\psi(\mathbf{h})$ instead of $\psi(\mathbf{h}; \mathbf{x})$. We will only solve objective functions of the following form: $\psi(\mathbf{h}) = \psi(c \cdot \mathbf{h})$, with $\{c \in \mathbb{R} | c \neq 0\}$ in this work, i.e., the function is invariant of the filter's impulse response magnitude and we can solve

the optimisation problem in equation (3) using unconstrained optimisation algorithms [11]. We will consider the following three optimisation formulations: Firstly, by writing $\frac{\eta(\mathbf{h})}{\beta(\mathbf{h})}$ as a non-linear Generalised Rayleigh Quotient (GRQ), i.e., $\frac{\mathbf{h}^T \mathbf{A}(\mathbf{h})\mathbf{h}}{\mathbf{h}^T \mathbf{B}(\mathbf{h})\mathbf{h}}$, and solving the GRQ using an optimality criterion method [11]. Secondly, by solving the unconstrained optimisation problem $\max_{\mathbf{h}} \psi(\mathbf{h})$ using gradient descent methods. Lastly, by solving the unconstrained optimisation problem with normalised impulse response functions, i.e., we reformulate the problem so that we solve $\max_{\mathbf{h}} \frac{\eta(\mathbf{g})}{\beta(\mathbf{g})}$, where $\mathbf{g} = \frac{\mathbf{h}}{\sqrt{\mathbf{h}^T \mathbf{h}}}$ using gradient descent methods. The latter method uses a normalised filter and can potentially improve the efficiency of the numerical optimisation methods and can be used to optimise functions that do not adhere to $\psi(\mathbf{h}) = \psi(c \cdot \mathbf{h})$ using unconstrained optimisation methods.

2.3 Objective functions

In this section, a brief overview is given of the objective functions considered in this work. All objective functions will be calculated using either the envelope spectrum $\mathbf{b}_{s^2}^{ES}$ or the squared envelope spectrum $\mathbf{b}_{s^2}^{SES}$.

2.3.1 L2/L1-norm

The L2/L1-norm of $\mathbf{b} \in \mathbb{R}^D$ is defined as follows:

$$\psi_{L2/L1}(\mathbf{h}) = \frac{||\mathbf{b}||_2}{||\mathbf{b}||_1} = \frac{\sqrt{\sum_{i=1}^{D} |b[i]|^2}}{\sum_{i=1}^{D} |b[i]|} \tag{4}$$

where $\mathbf{b}$ is dependent on $\mathbf{h}$, but this dependence is not shown to simplify the notation. In this work, we calculate the L2/L1-norm of the envelope spectrum $\psi_{L2/L1}\left(\mathbf{b}_{s^2}^{ES}\right)$ and the L2/L1 norm of the squared envelope spectrum $\psi_{L2/L1}\left(\mathbf{b}_{s^2}^{SES}\right)$ as two separate objective functions. The L2/L1-norm is a sparsity measure and is used to enhance the cyclostationary components in the signal without prior knowledge of the cyclic period. The GRQ for $\psi_{L2/L1}\left(\mathbf{b}_{s^2}^{ES}\right)$ proposed in Ref. [6] is used in this work. We did not implement the GRQ of $\psi_{L2/L1}\left(\mathbf{b}_{s^2}^{SES}\right)$.

2.3.2 Squared Envelope Spectrum (SES) negentropy

The SES negentropy is another sparsity measure and is defined as follows:

$$\psi_{SES-N}(\mathbf{h}) = \left\langle \frac{\mathbf{b}_{s^2}^{SES}}{\langle \mathbf{b}_{s^2}^{SES} \rangle} \log \left(\frac{\mathbf{b}_{s^2}^{SES}}{\langle \mathbf{b}_{s^2}^{SES} \rangle} \right) \right\rangle \tag{5}$$

with the mean of the vector $\mathbf{x} \in \mathbb{R}^L$ denoted $\langle \mathbf{x} \rangle$, i.e., $\langle \mathbf{x} \rangle = \frac{1}{L} \sum_{i=1}^{L} x[n]$. The GRQ form proposed in Ref. [6] is used for the GRQ formulation in this work.

2.3.3 Indicator of second-order Cyclostationarity (ICS2)

Two formulations of the ICS2 are considered. Formulation 1 of the ICS2 (referred to as ICS2 (f1)) was used in Ref. [2] and is defined as follows in this work:

$$ICS_{2,f1}(\mathbf{h}; \mathscr{A}_\alpha) = \frac{\psi_\alpha(\mathbf{h}; \mathscr{A}_\alpha)}{\left(\frac{1}{L} \sum_{n=1}^{L} s[n]^2 \right)^2} \tag{6}$$

where the numerator contains the sum of the targeted components is given by

$$\psi_\alpha(\mathbf{h}; \mathscr{A}_\alpha) = \left(\sum_{n \in \mathscr{A}_\alpha} b_{s^2}^{SES}[n] \right)^2 \tag{7}$$

and $\mathscr{A}_\alpha$ denotes the set of targeted indices and the amplitude of the squared envelope spectrum that corresponds to the nth index is denoted $b_{s^2}^{SES}[n]$. The dominator of equation (6) contains the squared mean energy of the filtered signal $\mathbf{s}$ in the time domain. The GRQ proposed in Ref. [2] is used for the ICS2 (f1)'s GRQ-based formulation.

The second ICS2 formulation, referred to as ICS2 (f2) in this work, is normalised by the DC component of the squared envelope spectrum, i.e.,

$$ICS_{2,f2}(\mathbf{h};\mathscr{A}_\alpha) = \frac{\psi_\alpha(\mathbf{h};\mathscr{A}_\alpha)}{\left(b_{s^2}^{SES}[0]\right)^2} \tag{8}$$

The denominator is the squared mean energy of the filtered signal in the angle domain. Formulations 1 and 2 are equivalent under constant speed conditions. However, the squared energy of the time domain signal and the squared energy of the angle domain signal are generally not equal under varying speed conditions, which can potentially impede its performance under varying speed conditions.

2.4 Gradient-based optimisation algorithms

In this work, we will use gradient descent and optimality criterion-based gradient-based optimisation algorithms. The gradient of the objective function in equation (3) is given by

$$\nabla_\mathbf{h}\psi(\mathbf{h}) = \frac{1}{\beta(\mathbf{h})} \cdot \nabla_\mathbf{h}\eta(\mathbf{h}) - \frac{\eta(\mathbf{h})}{\beta(\mathbf{h})^2} \cdot \nabla_\mathbf{h}\beta(\mathbf{h}) \tag{9}$$

with the vector differential operator denoted $\nabla_\mathbf{h} = [\frac{\partial}{\partial h_1}, \frac{\partial}{\partial h_2}, \ldots, \frac{\partial}{\partial h_N}]$ and $\nabla_\mathbf{h} \in \mathbb{R}^N$. The gradients of the numerator and denominator can be calculated using the chain rule. If the feature is written as a Generalised Rayleigh Quotient (GRQ), the numerator and denominator are decomposed as follows: $\eta(\mathbf{h}) = \mathbf{h}^T\mathbf{A}(\mathbf{h})\mathbf{h}$ and $\beta(\mathbf{h}) = \mathbf{h}^T\mathbf{B}(\mathbf{h})\mathbf{h}$, which means that the gradient can be simplified as follows:

$$\nabla_\mathbf{h}\psi(\mathbf{h}) = \frac{2}{\beta(\mathbf{h})} \cdot \left(\mathbf{A}(\mathbf{h})\mathbf{h} + \frac{1}{2}\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{A}(\mathbf{h})\right)\mathbf{h} - \psi \cdot \left(\mathbf{B}(\mathbf{h})\mathbf{h} + \frac{1}{2}\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{B}(\mathbf{h})\right)\mathbf{h}\right)\right) \tag{10}$$

If we use an optimality criterion approach and solve $\nabla_\mathbf{h}\psi(\mathbf{h}) = \mathbf{0}$, while assuming $\beta(\mathbf{h}) \neq 0$, we obtain the following equation:

$$\left(\mathbf{A}(\mathbf{h}) + \frac{1}{2}\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{A}(\mathbf{h})\right) - \psi \cdot \left(\mathbf{B}(\mathbf{h}) + \frac{1}{2}\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{B}(\mathbf{h})\right)\right)\right)\mathbf{h} = \mathbf{0} \tag{11}$$

Furthermore, if $\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{A}(\mathbf{h}) - \psi\nabla_\mathbf{h}\mathbf{B}(\mathbf{h})\right)\mathbf{h} = \mathbf{0}$, we can solve the generalised eigenvalue problem in the following form:

$$(\mathbf{A}(\mathbf{h}) - \psi \cdot \mathbf{B}(\mathbf{h}))\mathbf{h} = \mathbf{0} \tag{12}$$

to find the filter coefficients $\mathbf{h}$. Equation (12) has been iteratively solved to find the filters $\mathbf{h}$ in blind deconvolution and blind filtering problems and is referred to as the iterative Generalised Eigenvalue Decomposition (i-GED) approach in this work. Therefore, by solving equation (12), we assume $\mathbf{h}^T\left(\nabla_\mathbf{h}\mathbf{A}(\mathbf{h}) - \psi\nabla_\mathbf{h}\mathbf{B}(\mathbf{h})\right)\mathbf{h} \approx \mathbf{0}$. The i-GED starts with an initial guess for $\mathbf{h}$ to calculate $\mathbf{A}(\mathbf{h})$ and $\mathbf{B}(\mathbf{h})$, whereafter the eigenvalue problem in equation (12) is solved to obtain a new set of eigenvectors. The filter's impulse response function $\mathbf{h}$ is set equal to the eigenvector that corresponds to the largest eigenvalue and the process is repeated until convergence is reached.

With the gradient descent methods, the Conjugate Gradient (CG) algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (a quasi-Newton method) are used with analytical gradients. The analytical gradients of equation (9) were verified against finite differences. To utilise the gradient descent algorithms, we need to supply the negative of the objective functions in Section 2.3, i.e., we transform the problem to: $\min_\mathbf{h} -\psi(\mathbf{h})$. The tolerance for termination was set to 10^{-6} and the maximum number of iterations were set to 250. In this work, the length of the filter's impulse response function is 64.

3 Results

The signal considered in this work is used to model outer race bearing damage and is generated with

$$x(t) = M(t) \cdot (\kappa \cdot r(t) \otimes x_b(t) + \varepsilon(t)) \tag{13}$$

where $r(t)$ is a single degree-of-freedom's impulse response function, $x_b(t)$ is a train of dirac functions with a periodicity of 3.62 shaft orders and $\varepsilon(t)$ is standardised Gaussian noise. The function $M(t) = \left(\frac{\omega}{2 \cdot \pi \cdot 10}\right)^2$ is used to model the amplitude modulation due to the varying speed. The impulse magnification factor κ is used to change the signal-to-noise ratio of the damage in this work. Only linear speed signals are considered in this work, i.e., $\omega'(t)$ is constant and therefore the acceleration of the reference shaft is denoted ω'.

The signal $x(t)$ and the train of impulses $M(t) \cdot (\kappa \cdot r(t) \otimes x_b(t))$ are shown in figure 1 for a linear speed-varying case with $\kappa = 0.72$. The speed is shown in figure 1(a). The train of impulses is superimposed on the signal $x(t)$ in figures 1(b) and 1(c). The Power Spectral Density (PSD) of the time domain signal, the PSD of the train of impulses and the centre frequency of the impulse response function of the train of impulses, which is 3750 Hz, is shown in figure 1(d). The frequency-domain SES, calculated using the DFT matrix $\mathbf{D}$, is shown in figure 1(e) and the order-domain SES, calculated using the velocity synchronous DFT matrix $\mathbf{V}$, is shown in figure 1(f). The components smear in the frequency domain in figure 1(f) since they lose periodicity under the time-varying speed conditions. The cyclic order of the impulses (3.62 events/revolution and its harmonics) are much sparser and more prominent in figure 1(f), which enables early damage detection.
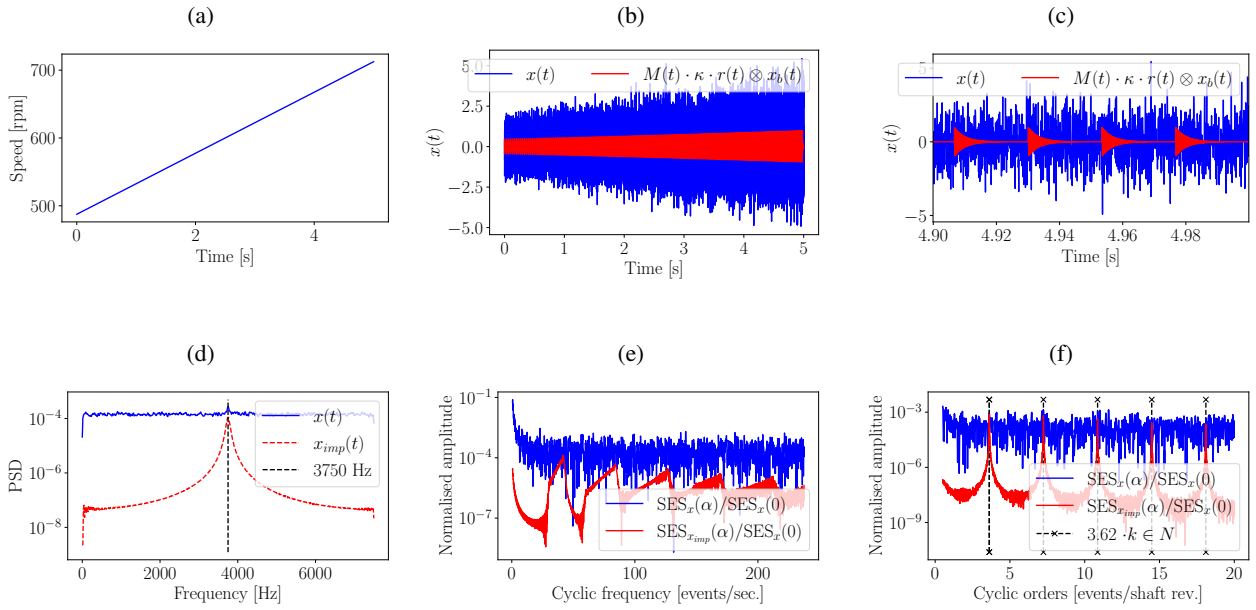

Figure 1: Different representations of the signal $x(t)$ and the train of impulses $x_{imp} = M(t) \cdot \kappa \cdot r(t) \otimes x_b(t)$ are shown for $\kappa = 0.72$ and the speed profile in figure 1(a). The time domain signal and zoomed time domain signals are shown in figures 1(b) and 1(c), the PSD of the time domain signals are shown in figure 1(d). The SES, using the temporal signal, and the SES, using the angular domain signal, are shown in figures 1(e) and 1(f) respectively. The SES is normalised using the DC component of $x(t)$'s SES to make the results easier to visualise (without including the DC component in the plot), while still interpreting the magnitude of the components relative to the DC component.

3.1 Datasets

Two datasets were generated using equation (13) to compare the overall performance of the methods:

1. In the first dataset, the impulse magnification factor κ is increased from 0 to 1 under the speed conditions shown in figure 1(b). This dataset is used to compare the performance of the methods to detect damage in Section 3.2.

2. In the second dataset, the angular acceleration of the signal is varied and the impulse magnification factor κ is constant. A linear speed profile is parameterised as follows:

$$\omega(t) = \left(20 \cdot \pi + \omega' \cdot (t - 2.5)\right) \tag{14}$$

where $\omega' \in \mathbb{R}$ is a parameter used to control the acceleration, e.g., $\omega' = 0$, simulates constant speed conditions. A dataset is generated using $\kappa = 0.72$ for $0 \leq \omega' \leq 23.56$ rad/s^2 to quantify the sensitivity of the methods to changes in speed in the presence of damage. The sensitivity of the methods to varying speeds is evaluated in Section 3.3.

The overall performance of the methods, using the results in Sections 3.2 and 3.3, is summarised in Section 3.4.

3.2 Sensitivity to changes in impulse magnification factors κ

In this section, the sensitivity of the filtered signals, obtained using different optimal filter design methods, are compared using the first dataset in Section 3.1. The optimal filter design methods are separately applied to each signal with the results shown in figure 2 as a function of impulse magnification factor κ. The results of the normalised filter formulation (i.e., when supplying $\mathbf{g} = \mathbf{h}/\sqrt{\mathbf{h}^T \mathbf{h}}$ to the objective function) are not included in figure 2. This is because the normalised filter's results are similar to the non-normalised filter's results and excluding it simplified the figure. However, the performances of the normalised and non-normalised filters are quantified and discussed in Section 3.4.

The results of each objective function (refer to Section 2.3) are presented in a separate row in figure 2. In the left column, the final function value of the objective function is presented. In the right column, the ICS2 (f2) of the filtered signal $\mathbf{s} = \mathbf{Xh}$, with $\mathbf{h}$ obtained by optimising the specific objective function, is presented to make it possible to compare the different methods.

The ICS2 features in figures 2(a) and 2(c) show the earliest indications of damage between $\kappa = 0.38$ and $\kappa = 0.44$, earlier than the L2/L1 (ES) in figure 2(e), the L2/L1 (SES) in figure 2(g) and the SES negentropy in figure 2(i). This is expected because the ICS2 features target specific cyclic component of 3.62 shaft orders and its harmonics, whereas the L2/L1 (ES), L2/L1 (SES) and SES negentropy features are blind (i.e., specific cyclic components are not targeted).

In figures 2(b) and 2(d), the variance between the different algorithms used to optimise the ICS2 (f1) feature is much smaller than the ICS2 (f2) feature. The ICS2 (f2) feature optimised using the i-GED algorithm did not converge on the dataset and performed very poorly. The ICS2 (f2) feature using the BFGS algorithm had slightly smaller function values, which indicates that it was not able to optimise the objective function as well as the CG for the cases considered in this work.

For the blind features, the SES negentropy using the CG and BFGS algorithms in figure 2(i) and the L2/L1 (SES) with the CG and BFGS algorithms in 2(g) display the best performance; the trends are near monotonic, and the damage is clearly detected after $\kappa = 0.52$. The L2/L1 (ES) feature in figure 2(e) can detect the damage with the CG and i-GED algorithms, but its trend displays strong non-monotonic behaviour, which is undesirable for diagnostics. We found the L2/L1 (SES) is much easier to optimise than the L2/L1 (ES) when using gradient descent methods. The SES is estimated from the squared value of the spectrum of the squared signal $\mathbf{d}_{s2}$, with the ES calculated as the square root of the SES. The ES is the square root of the SES and the absolute value of $\mathbf{d}_{s2}$. The absolute value is non-smooth, which impedes the performance of the CG and BFGS algorithms. In figures 2(f), 2(h) and 2(j), the SES negentropy using the BFGS and CG and the L2/L1 (ES) using the CG and the i-GED achieve slightly larger function values than the L2/L1 (SES) feature for $\kappa \geq 0.75$. This indicates that the cyclostationarity of the damage is more enhanced. However, the L2/L1 SES and SES negentropy is more monotonic, which makes them desirable for fault diagnostics. Ultimately, quantitative comparisons are needed to rank the methods. This will be performed in Section 3.4 using the speed sensitivity measure defined in the next section.

3.3 Sensitivity to changes in speed

In condition monitoring, it is important that the features are insensitive to changes in speed to reduce the uncertainty in the components' condition under time-varying conditions. The second dataset discussed in Section 3.1 is used to evaluate the sensitivity of the methods to speed and will be used to quantify the overall performance of the methods in Section 3.4. Only the results of the ICS2 (f1) feature are considered in this section to motivate the sensitivity ratio performance metric used in this work.

The results of the ICS2 (f1) feature is shown in figure 3 in a similar two-column format as figure 2. The ICS2 (f1) feature in figure 3(a) is much dependent on speed. This dependence is because the ICS2 (f1) feature is normalised by the time domain energy as opposed to the angle domain energy. This is corroborated by the
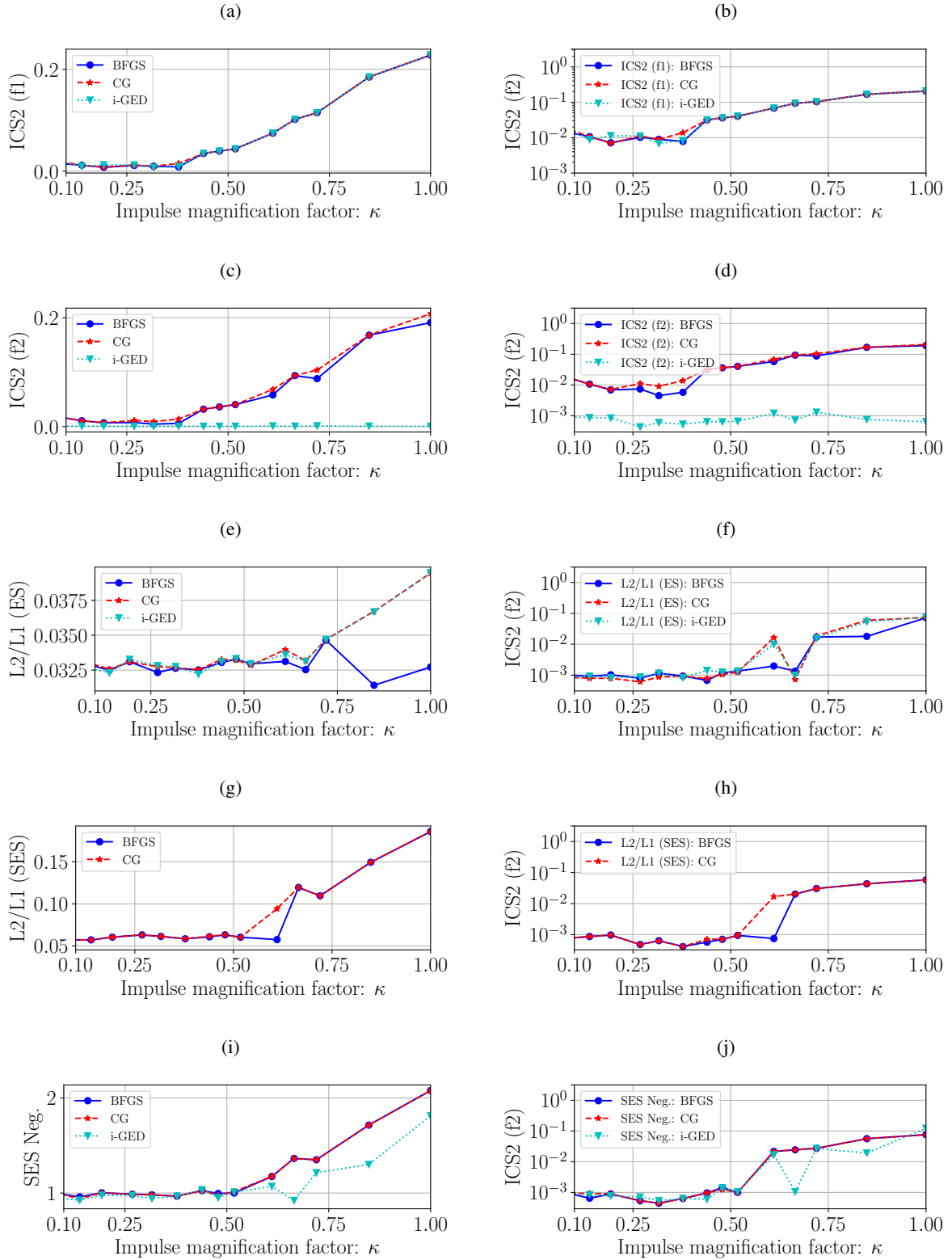
Figure 2: The fault trending results obtained using dataset 1 in Section 3.1. For a specific row, the left figure, which is presented on a linear scale, contains the final objective function value of a specific feature. The right figure, which is presented on a semi-log scale, contains the ICS2 of the optimally filtered signal **s**, with the filter's impulse response function obtained by maximising the objective function described in the legend of the figure in the right figure. None of the impulse response functions were normalised.
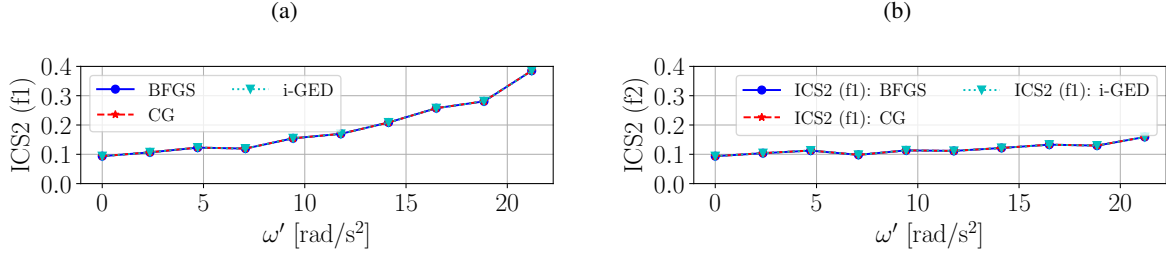
Figure 3: The results obtained using dataset 2 in Section 3.1 is displayed in a similar format to figure 2. The impact magnification factor was $\kappa = 0.72$ for all signals.

results obtained when calculating the ICS2 (f2) of the filtered signals, shown in figure 3(b). The ICS2 (f2) of the filtered signals in figure 3(b) is much more robust to speed fluctuations than the ICS2 (f1) of the filtered signal in figure 3(b) since it uses the angle domain signal's energy for normalisation.

We can use the standard deviation of the features in figure 3 to quantify the sensitivity of the methods to speed when κ is fixed. Let $\Psi_i(\kappa, \omega'_m)$ denote feature i's function value, obtained with an optimally filtered synthetic signal with an impulse magnification factor of κ and angular acceleration of ω'_m. We can use $\Psi_i(\kappa, \omega'_m)$ to calculate the standard deviation of the features in figure 3 using

$$\sigma_{\Psi_i} = \sqrt{\frac{1}{N} \sum_{m=1}^{N} \left(\Psi_i(0.72, \omega'_m) - \bar{\Psi}_i(0.72, \omega'_m) \right)^2} \tag{15}$$

where ω'_m denotes the mth angular acceleration point in figure 3 and $\bar{\Psi}_i(0.72, \omega'_m) = \left\langle \Psi_i(0.72, \omega'_m) \right\rangle_m$. The standard deviation of the results in figures 3(a) and 3(b), calculated using equation (15), is 0.087 and 0.018, respectively. This highlights that the ICS2 (f1) feature is more sensitive to varying speed conditions than the ICS2 (f2) feature. This will be used to calculate the sensitivity ratio performance metric in the next section.

3.4 Summary of performance

Three performance measures are used to compare the optimal design formulations in this work. The first method is the Sensitivity Ratio (SR), which is defined as follows:

$$\text{SR} = \frac{\Psi_i(1, \omega') - \Psi_i(0, \omega')}{\sigma_{\Psi_i}} \tag{16}$$

where $\Psi_i(\kappa, \omega')$ is defined in the previous section. The numerator of equation (16) quantifies the change in the feature value from the healthy $\kappa = 0$ to damage $\kappa = 1$ case using the results in figure 2. The denominator of equation (16) is defined in equation (15) and quantifies the sensitivity of the feature to speed using the second dataset. The second performance measure is the monotonicity of the features with respect to κ as defined in Ref. [12]. The third performance metric is the trendability of the features with respect to κ as defined in Ref. [12]. The monotonicity and trendability are important performance metrics for prognostics [12] and is calculated using the results in figure 2.

The results are summarised in table 1 for the optimal filtering formulations. The metrics of two features of the filtered signals are included in table 1; the final objective function value and the ICS2 (f2) of the filtered signal **s**. The number of function evaluations is also summarised in the last column to compare the efficiency of the methods.

The results in table 1 indicate that the ICS2 (f1) using the CG algorithm performed the best to enhance the damage; it had the highest SR, monotonicity and trendability values for the ICS2 (f2) of **s** feature and required relatively few iterations to converge (second least after the L2/L1 (ES) i-GED feature). However, if the final objective function is used, the method's sensitivity ratio is much lower than the ICS2 (f2). This indicates that the feature value of the ICS2 (f1) is sensitive to speed, which is undesirable for time-varying speed conditions. The ICS2 (f2) using the CG algorithm performed second best.

When comparing the ICS2 (f2) of **s** results of the blind features, it is concluded that the SES negentropy using the CG algorithm performed the best to enhance the damage; it has the highest SR, monotonicity and

Feature	Alg.	Filter normed	Final obj. fun. SR	Mon.	Trend.	ICS2 (f2) of **s** SR	Mon.	Trend.	# fun eval. Med. $\pm$ MAD
ICS2 (f1)	BFGS	N	2.517	0.273	0.879	10.973	0.273	0.880	88 ± 24
ICS2 (f1)	BFGS	Y	2.517	0.273	0.879	10.972	0.273	0.880	88 ± 20
ICS2 (f1)	CG	N	2.524	0.455	0.882	11.006	0.455	0.883	82 ± 22
ICS2 (f1)	CG	Y	2.524	0.455	0.882	11.006	0.455	0.883	82 ± 22
ICS2 (f1)	i-GED	N	2.515	0.455	0.879	10.968	0.455	0.880	250 ± 0
ICS2 (f2)	BFGS	N	8.048	0.273	0.875	8.048	0.273	0.875	136 ± 70
ICS2 (f2)	BFGS	Y	8.048	0.273	0.875	8.048	0.273	0.875	136 ± 70
ICS2 (f2)	CG	N	11.005	0.364	0.883	11.005	0.364	0.883	124 ± 32
ICS2 (f2)	CG	Y	11.005	0.364	0.883	11.005	0.364	0.883	124 ± 32
ICS2 (f2)	i-GED	N	-1.008	-0.091	0.208	-1.008	-0.091	0.208	250 ± 0
L2/L1 (ES)	BFGS	N	0.222	0.182	0.203	2.724	0.000	0.638	258 ± 2
L2/L1 (ES)	BFGS	Y	0.224	0.182	0.205	2.726	0.000	0.638	258 ± 2
L2/L1 (ES)	CG	N	3.246	0.000	0.764	4.065	0.091	0.731	672 ± 79
L2/L1 (ES)	CG	Y	3.246	0.000	0.762	4.074	0.091	0.732	678 ± 87
L2/L1 (ES)	i-GED	N	3.370	0.000	0.736	4.418	0.182	0.714	56 ± 21
L2/L1 (SES)	BFGS	N	6.209	0.000	0.797	4.267	0.182	0.775	254 ± 4
L2/L1 (SES)	BFGS	Y	6.209	0.000	0.797	4.267	0.182	0.775	254 ± 4
L2/L1 (SES)	CG	N	6.256	0.000	0.840	4.343	0.273	0.820	298 ± 79
L2/L1 (SES)	CG	Y	6.256	0.000	0.840	4.343	0.273	0.820	298 ± 79
SES Neg.	BFGS	N	6.062	0.091	0.806	4.088	0.182	0.811	161 ± 40
SES Neg.	BFGS	Y	6.062	0.091	0.806	4.088	0.182	0.811	161 ± 40
SES Neg.	CG	N	6.279	0.182	0.810	5.100	0.000	0.810	300 ± 86
SES Neg.	CG	Y	6.279	0.182	0.810	5.100	0.000	0.810	289 ± 75
SES Neg.	i-GED	N	3.962	0.182	0.695	3.567	-0.182	0.633	250 ± 0

Table 1: A summary of the results that were obtained using the methods. The objective function (or feature), the optimisation algorithm and whether the filter's impulse response function was normalised before the function evaluation (Y) or not (N) are shown in the first three columns of the table. Thereafter, the Sensitivity Ratio (SR), the monotonicity and the trendability of the final objective function is shown. The same measures are also calculated and documented for the ICS2 (f2) of the filtered signal **s**. Lastly, the Median $\pm$ the Median Absolute Deviation (MAD) number of the function evaluations is also summarised for each case.

second to highest trendability. The L2/L1 (ES) using the i-GED performed second best when using the SR and number of function evaluations as guidance. The L2/L1 (SES) using the CG algorithm had a very good SR, with the highest monotonicity and trendability of the blind features. The L2/L1 (ES) performs slightly worse than the L2/L1 (SES) when using the gradient descent methods; the performance measures are lower, and it required more function evaluations.

According to the SR, monotonicity and trendability performance metrics, the CG generally performed much better than the BFGS and i-GED algorithms. Even though the GRQ with the i-GED published in literature had good convergence characteristics, it does not use the consistent gradient, which can impede the performance of new objective functions (e.g., ICS2 (f2)). The filter normalisation process did not make a significant impact, since all the objective functions considered in this work are filter magnitude invariant. However, some of the filter normalisation cases required slightly fewer function evaluations to achieve the same performance as the non-normalised filters (e.g., ICS2 (f1) (BFGS), SES Negentropy (CG)) or resulted in better ICS2 (f2) of **s** scores (e.g., L2/L1 (SES) (CG)). The main benefit of the filter normalisation process is that unconstrained optimisation algorithms can be applied to objective functions that are sensitive to the magnitude of the filter's impulse response function.

Lastly, we also investigated the performance of the methods when using the DFT matrix, i.e., **F**, instead of the velocity synchronous DFT matrix, i.e., **V**. We found that the blind and targeted methods performed poorly when using **F** instead of **V** under large angular accelerations. Therefore, it is important to calculate the features in the cyclic order domain as opposed to the cyclic frequency domain.

4 Conclusion

In this work, optimal filters were obtained using different objective function formulations, different optimisation formulations and different optimisation algorithms. The methods were applied on a synthetic dataset and the performance of the methods was quantified. The ICS2 feature performed the best, however, it is a targeted method, and its performance depends on the formulation that is used. The SES negentropy performed the best of the blind features. The results indicate that the optimisation algorithm influences the performance and efficiency of the optimisation process. The conjugate gradient algorithm is proposed as a starting point for new objective functions. However, this should be further investigated on different datasets. Furthermore, the sufficient tolerance for termination and the maximum number of iterations to achieve acceptable results for condition monitoring also need to be investigated on different datasets.

Acknowledgements

References

[1] M. Tiboni, C. Remino, R. Bussola, and C. Amici. A review on vibration-based condition monitoring of rotating machinery. *Applied Sciences*, 12(3):972, 2022.

[2] M. Buzzoni, J. Antoni, and G. d'Elia. Blind deconvolution based on cyclostationarity maximization and its application to fault identification. *Journal of Sound and Vibration*, 432:569–601, 2018.

[3] J.Y. Lee and A.K. Nandi. Extraction of impacting signals using blind deconvolution. *Journal of Sound and Vibration*, 232(5):945–962, 2000.

[4] G.L. McDonald, Z.Q., and M.J. Zuo. Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection. *Mechanical Systems and Signal Processing*, 33:237–255, 2012.

[5] E. Soave, G. D'Elia, M. Cocconcelli, and M. Battarra. Blind deconvolution criterion based on Fourier–Bessel series expansion for rolling element bearing diagnostics. *Mechanical Systems and Signal Processing*, 169:108588, 2022.

[6] C. Peeters, J. Antoni, and J. Helsen. Blind filters based on envelope spectrum sparsity indicators for bearing and gear vibration-based condition monitoring. *Mechanical Systems and Signal Processing*, 138:106556, 2020.

[7] L. He, D. Wang, C. Yi, Q. Zhou, and J. Lin. Extracting cyclo-stationarity of repetitive transients from envelope spectrum based on prior-unknown blind deconvolution technique. *Signal Processing*, 183:107997, 2021.

[8] X. Gu, S. Yang, Y. Liu, R. Hao, and Z. Liu. Multi-sparsity-based blind deconvolution and its application to wheelset bearing fault detection. *Measurement*, 199:111449, 2022.

[9] Y. Miao, J. Wang, B. Zhang, and H. Li. Practical framework of Gini index in the application of machinery fault feature extraction. *Mechanical Systems and Signal Processing*, 165:108333, 2022.

[10] P. Borghesani, P. Pennacchi, S. Chatterton, and R. Ricci. The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery. *Mechanical Systems and Signal Processing*, 44(1-2):118–133, 2014.

[11] J.A. Snyman and D.N. Wilke. *Practical mathematical optimization*. Springer, 2018.

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