

# Differentiable Short-Time Fourier Transform Window Length Selection Driven by Cyclo-Stationarity

Douw Marx<sup>1,2</sup> and Konstantinos Gryllias<sup>1,2</sup>

<sup>1</sup> *Department of Mechanical Engineering, Division LMSD, KU Leuven, Celestijnenlaan 300, Box 2420, 3001 Leuven, Belgium*

*douw.marx@kuleuven.be*

*konstantinos.gryllias@kuleuven.be*

<sup>2</sup> *Flanders Make@KU Leuven, Belgium*

## ABSTRACT

The Short-Time Fourier transform is widely applied in the condition monitoring of rotating machinery. Even so, selecting the optimal window length for the Short-Time Fourier Transform remains a challenge. This work presents a procedure for adapting the Short Time Fourier Transform algorithm to be differentiable with respect to window length by using continuous window functions defined over the entire input signal duration. Thanks to this modification, a differentiable loss criterion can be defined to measure the Short-Time Fourier quality, and the gradient of the loss criterion with respect to window length can be computed. The optimal window length for a given loss criterion can then be efficiently solved for using a gradient-based optimization algorithm. Results from a simulated bearing dataset and three experimental bearing datasets are used to compare the optimal spectrograms obtained using different loss criteria. Specifically, a sparsity-based loss criterion is compared with two loss criteria inspired by the characteristic cyclo-stationarity nature of faults in rotating machinery. The results demonstrate the value of a continuous and differentiable window length selection method and highlight the importance of selecting appropriate loss criteria for defining STFT quality. Further, loss criteria that account for the cyclo-stationary nature of the signals are shown to be less likely to target single high-amplitude impulsive events compared to the sparsity-based loss criterion.

## 1. OPTIMAL WINDOW SELECTION FOR THE SHORT-TIME FOURIER TRANSFORM

The Short-Time Fourier transform (STFT) is a widely used method for analysing non-stationary signals as encountered in rotating machinery. It yields an informative time-frequency

representation by segmenting a signal with time-varying properties into consecutive windows and taking the Fourier transform of each window.

The Heisenberg Uncertainty principle dictates that the utility of the resulting time-frequency representation is dependent on a trade-off between time and frequency resolution, as governed by the selected STFT window length. Hence, optimal STFT window length selection is a crucial step in the analysis of non-stationary signals as encountered in rotating machinery.

As an alternative to searching for the optimal window length by trial and error or by evaluating all possible window lengths, we propose a STFT with associated loss criteria that can efficiently be optimised with respect to window length. The proposed STFT and loss criteria are a function of continuous window length, meaning that many conventional optimisation methods designed for continuous functions can be applied to conveniently solve for optimal window length. Furthermore, the proposed STFT is differentiable, allowing for a particularly efficient search of optimal window length using gradient-based optimizers.

A differentiable STFT can be useful for many signal processing tasks, but in this work, we focus on the application of the differentiable STFT to the condition monitoring of rolling element bearings. Specifically, we consider different unsupervised loss criteria for measuring spectrogram quality, including criteria that exploit knowledge of the periodic impulses expected in faulty rolling element bearings.

Previous work concerned with optimal STFT window length selection adjusted the window size in the frequency domain by defining the window size as a fixed number of cycles at each frequency (Mateo & Talavera, 2018), adjusted the window length based on the derivative of the instantaneous frequency (Xie, Lin, Lei, & Liao, 2012) or selected the window

Douw Marx et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

length by optimising the cone kernel distribution (Czerwinski & Jones, 1997).

More recent window selection methods exploited modern advances in automatic differentiation (Paszke et al., 2017) that make it possible to differentiate intricate loss criteria and optimise them using gradient-based optimisation methods.

Zhao et al. (2021), estimated optimal window parameters for the STFT by computing the gradients of the STFT with respect to window length and solving for the optimal window length by gradient-based optimisation. This differentiable STFT is further combined with neural networks and applied to a sound classification problem both for the case where the optimal window length is constant and for the case where it can vary in time.

Similarly, Lieber et al. (2022) proposed a differentiable STFT and applied it to STFT frequency tracking as well as spoken digit classification. In contrast with Zhao et al. (2021), a constant FFT size is used where a Hann window is zero-padded to match the full signal length, thereby making the STFT differentiable. The differentiability of the approach with respect to window length is proven, both for constant hop length and constant overlap. This work was then further extended by optimising for a time-dependent adaptive window length (Leiber, Marnissi, Barrau, & Badaoui, 2023).

In other related work, Fang et al. (Fang, Hu, Yang, Cao, & Jia, 2022) employed a differentiable framework for finding optimal filters for the deconvolution of bearing fault signals.

In this work, we propose a simplified variant of the differentiable STFT which neither requires a variable FFT size with changing window length (Zhao et al., 2021) nor zero padding (Leiber et al., 2022). We further introduce three unsupervised loss criteria that measure the STFT quality as function of window length. An unsupervised loss criteria implies that labeled samples are not used to solve a supervised problem as in earlier work (e.g. sound classification (Zhao et al., 2021), frequency tracking (Leiber et al., 2022)).

Since faulty machine data are generally not available to be used as target labels in the condition monitoring machines, an alternative loss criterion is required to drive the optimal window length optimisation problem. Although kurtosis-based metrics that encourage spectrogram sparsity have successfully been used to measure STFT quality (Zhao et al., 2021), such metrics are often ineffective for signals from rotating machinery that include impulsive signal components (Wodecki, Michalak, & Zimroz, 2021). Therefore, in addition to a differentiable STFT, we introduce three different loss criteria for optimizing the STFT window length. We demonstrate the benefit of accounting for the cyclo-stationarity nature of rolling element bearing signals in the design of these loss functions, thereby exploiting the knowledge that faults in rolling element bearings cause the resonate fre-

quency of the machine to be modulated by a characteristic fault frequency (McFadden & Smith, 1984). We finally show that the usefulness of a differentiable STFT depends on selecting an appropriate loss criterion that can appropriately measure the quality of the STFT for a given window length.

The paper is structured as follows. In Section 2 we show how the STFT can be made differentiable with respect to window length by multiplying the signal with a set of windows having many values close to zero. Section 3 presents three loss functions for measuring spectrogram quality that can be optimised with respect to window length. Next, the effectiveness of the proposed gradient-based optimisation loss functions is demonstrated in Section 4 using a simulated bearing fault signal. Finally, the same method is applied to three experimental signals in Section 5 and conclusions are drawn in Section 6.

## 2. A DIFFERENTIABLE STFT

The STFT of signal  $x[i]$  of length  $L$  sampled at discrete time index  $i$  is defined as:

$$\mathcal{F}[m, k] = \sum_{n=0}^{N-1} x[m+n]w[n]e^{-j\frac{2\pi}{N}kn}, \quad (1)$$

where  $i, n, k$  and  $m$  are the discrete indexes of the signal time, window time, FFT frequency bin and resampled window time respectively, with  $w$  a window function of length  $N$ .

The number of samples in a window,  $N$ , will determine the time-frequency resolution of the resulting STFT. Longer windows will better capture the frequency content of the signal but will also lead to reduced time resolution. To solve for the optimal window length using gradient-based optimization, Eq. 1 needs to be differentiable with respect to the window length. This is done by making the window function  $w$  a function of signal sample time  $i$  rather than the conventional window time  $n$ . Consequently, the size of the FFT for each of the windows becomes equal to the signal length  $L$ :

$$\hat{\mathcal{F}}_{\theta}[m, k] = \sum_{i=0}^{L-1} x[i]w_{\theta}(i, m)e^{-j\frac{2\pi}{L}ki} \quad (2)$$

Note that this modification of the STFT is done exclusively for the sake of differentiability and that the corresponding increase in the frequency resolution does not necessarily contribute additional information to the differentiable STFT (Leiber et al., 2022).

Ultimately, this procedure corresponds to an element-wise multiplication of the signal  $x[i]$  with a window function  $w_{\theta}(i, m)$ , centred at  $m$  and taking values close to zero outside the interval  $[m - \frac{N}{2}, m + \frac{N}{2}]$ . This procedure is demonstrated graphically in the left hand side of Figure 1, where a

stack of signals are element-wise multiplied with a stack of time-offset windows.

In this work, the windowing function  $w$  is a Gaussian window, although it could in principle be parameterized by multiple free parameters as for the generalized Gaussian window.

$$w_\theta(i, m) = \exp\left(-\left(\frac{6(i-m)}{\theta}\right)^2\right) \quad (3)$$

The window function  $w$  (bottom left corner of Figure 1) is parameterised by continuous window length  $\theta$ , serving as a proxy for the window length  $N$ . In this work a Gaussian window function is used rather than a Hann window as in (Leiber et al., 2022) since the window function is close to zero for large values of  $n$ , meaning that the window does not require zero padding as the window length changes during optimisation. In Eq. 3, six standard deviations of the Gaussian window are considered as the full window length  $N$ , with the window function taking very small values outside of this region (Zhao et al., 2021).

In this work,  $M$  windows are equally spaced over the entire signal duration with window hop length or stride  $s = \lfloor \frac{L}{M} \rfloor$ , such that

$$m \in \{s, 2s, 3s, \dots, Ms\}. \quad (4)$$

The number of windows  $M$  can be as large as computational resources allow but should be sufficiently large to ensure that the new window sample rate can capture the fault frequency. In this work, we chose the number of windows  $M$  such that are at least 15 windows between two fault events.

Thanks to the modified STFT formulation in Eq. 2 the STFT is differentiable with the window length proxy  $\theta$ .

$$\frac{\partial \hat{\mathcal{F}}_\theta[m, k]}{\partial \theta} = \sum_{i=0}^{L-1} x[i] \frac{\partial w(i, m)}{\partial \theta} e^{-j \frac{2\pi}{L} ki} \quad (5)$$

A differentiable metric for STFT quality can now be computed from the STFT, backpropagating the computed loss function gradients and optimising the windowing parameters (Right hand side of Figure 1).

### 3. CYCLO-STATIONARY LOSS FUNCTIONS FOR DEFINING STFT QUALITY

We consider three loss criteria namely  $\mathcal{L}_1$ ,  $\mathcal{L}_2$  and  $\mathcal{L}_3$  for quantifying STFT quality for the purpose of bearing fault detection. In all cases, the squared magnitude of the STFT, or spectrogram, is used in the loss function and visualisation.

$$\mathcal{S}_\theta[m, k] = \left| \hat{\mathcal{F}}_\theta[m, k] \right|^2 \quad (6)$$

$\mathcal{L}_1$  aims to find the window length that leads to a maximally sparse STFT. This metric, called the concentration, is based on the kurtosis (Zhao et al., 2021) and is defined as:

$$\mathcal{L}_1(\theta) = \frac{c_4(\mathcal{S}_\theta)}{c_2(\mathcal{S}_\theta)} \quad \text{with} \quad c_p(v) = |v|^p \quad (7)$$

$\mathcal{L}_2$  is inspired by cyclostationary analysis in condition monitoring of rotating machinery (Antoni, 2009) and accounts for the periodic impulses of energy characteristic of bearing faults. This is done by maximizing the envelope spectrum over a range of expected fault frequencies. The envelope,  $E_\theta$  is computed by averaging the magnitude of the spectrogram over the frequency axis.

$$E_\theta[m] = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{S}_\theta[m, k] \quad (8)$$

The total loss then consists of summing the envelope spectrum Fourier coefficients for a range of cyclic frequency indices  $C = \{c_a, c_a + 1, \dots, c_b\}$  around the expected fault frequency. A range of frequencies is specified to ensure that the loss function does not fit the envelope to a non-fault-related signal component like the rotation or gear mesh frequency.

$$\mathcal{L}_2(\theta) = \sum_{c \in C} \left( \sum_{m=0}^{M-1} E_\theta[m] e^{-j \frac{2\pi}{M} cm} \right) \quad (9)$$

Similar to  $\mathcal{L}_2$ ,  $\mathcal{L}_3$  also accounts for the periodic impulses of energy typical of bearing faults but does not assume that the envelope signal will be sinusoidal as in  $\mathcal{L}_2$ . Instead,  $\mathcal{L}_3$  maximizes the sum of the autocorrelation of the envelope for a range of lag values  $P = \{p_a, p_a + 1, \dots, p_b\}$  around the expected lag between fault impulses.

$$\mathcal{L}_3(\theta) = \sum_{p \in P} \left( \sum_{m=0}^{M-1} E_\theta[m] E_\theta[m+p] \right) \quad (10)$$

In this work, the range of cyclic frequencies and lag values used in  $\mathcal{L}_2$  and  $\mathcal{L}_3$  are chosen to span the range  $[0.9f_c, 1.1f_c]$  where  $f_c$  is the expected fault frequency for given rotation speed. Consequently, the range of cyclic frequencies and lag values are defined as:

$$\begin{aligned} c_{a,b} &= \left\lfloor \frac{1 \pm 0.1f_c M}{f_s} \right\rfloor \\ p_{a,b} &= \left\lfloor \frac{1 \pm 0.1f_s}{f_c} \right\rfloor. \end{aligned} \quad (11)$$

All operations required for computing any of the loss func-

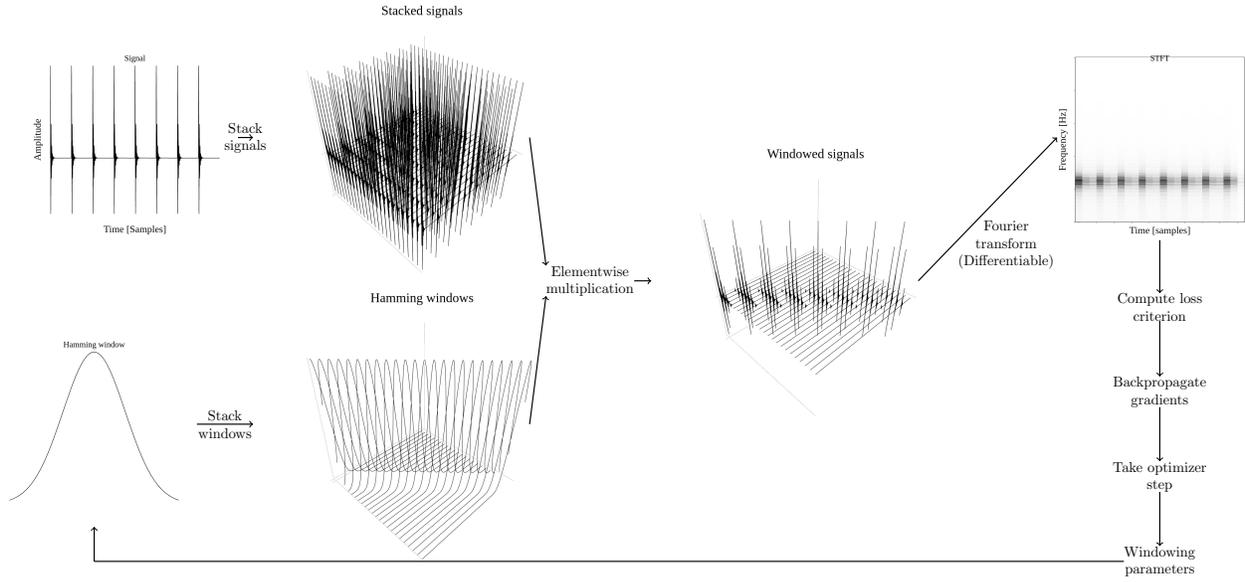


Figure 1. Procedure: Multiplication of the signal with a window function that has the same length as the input signal leads to a differentiable STFT. After computing a loss criterion measuring the spectrogram quality, the optimal window length can be solved for using a gradient based optimiser.

tions can be conveniently differentiated by automatic differentiation. Since the gradient of the STFT with respect to the window length is known from Eq. 5, the gradient of the loss with respect to the window length can be computed using the chain rule.

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \hat{\mathcal{F}}_\theta} \frac{\partial \hat{\mathcal{F}}_\theta}{\partial \theta} \quad (12)$$

Finally, the gradient is used in a gradient-based optimiser to update the optimal window until convergence.

#### 4. OPTIMAL STFT WINDOW LENGTH SELECTION FOR A SIMULATED BEARING SIGNAL

In this section, the proposed optimisation algorithm is applied to a simulated bearing fault signal. The signal, shown in Figure 2, is based on a phenomenological bearing model (McFadden & Smith, 1984). Periodic excitations of the natural frequency of the machine when a rolling element passes through a fault region is modelled as a first-order time response convolved with a Dirac comb. Although no slip of rolling elements are present in the simulation, the amplitude  $a$  of the bearing impulses are normally distributed,  $a \sim \mathcal{N}(\mu = 1, \sigma = 0.1)$ , and the signal is contaminated with Gaussian noise,  $\nu \sim \mathcal{N}(\mu = 0, \sigma = 0.1)$ . Further properties of the simulated signal are listed in Table 1.

To demonstrate the gradient-based optimization of a STFT loss criteria, Figure 3a shows the loss landscape of the objec-

Table 1. Phenomenological model parameters

Simulation parameter	Value
Sampling Rate	20.48 kHz
Signal Duration	0.7 s
Transient amplitude mean	1 $m/s^2$
Transient amplitude standard deviation	0.1 $m/s^2$
Measurement Noise standard deviation	0.33 $m/s^2$
Fault mode	Outer race
Ball diameter	8.4 mm
Pitch circle diameter	71.5mm
Number of rolling elements	16
Carrier frequency	4230 Hz
Transient damping ratio	0.05
Contact angle	15.17 Degrees
Operating speed	500 RPM

tive function  $\mathcal{L}_2$  in Eq. 9. The optimisation trajectory from the optimisation start point, traced by the optimiser, is superimposed on the response surface, demonstrating that the optimiser can find the optimal window length (indicated with \*) for the prescribed loss function.

Further, to demonstrate the time-frequency resolution trade-off, spectrograms for non-optimal candidate window lengths A and C, annotated in Figure 3a, are shown in Figures 3b and 3c respectively. Candidate A has a low frequency resolution, while candidate C has a low time resolution. All optimal window lengths obtained in this paper for a given loss criterion will necessarily make a compromise between time and frequency resolution, with certain loss criteria favoring higher

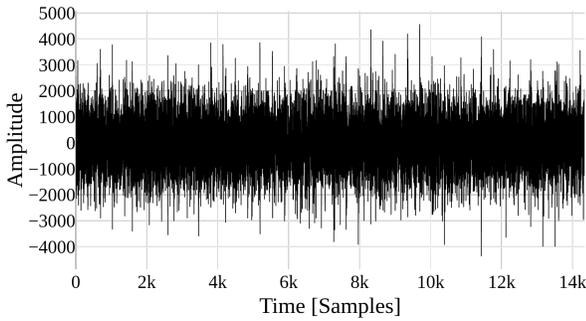


Figure 2. Synthetic outer race fault bearing signal contaminated with noise.

time resolution, and others favoring higher frequency resolution.

The choice of loss criterion and the resulting optimal window length can ultimately influence machine diagnosis, either directly, or in downstream signal processing tasks that employ the STFT. For instance, for the bearing diagnostics problem, a window length resulting in a very high frequency resolution but low time resolution could fail to capture all periodic impulses in time. As a result, the periodic fault signature characteristic of a bearing fault might not be detected in the spectrogram. Conversely, a loss criteria that tends to select window lengths with very high time resolution will have a coarse frequency resolution meaning that the specific resonant frequency bands that convey the bearing fault information might not be present in the STFT or downstream signal processing methods that make use of the STFT.

After convergence of an optimizer minimizing the example loss function  $\mathcal{L}_2$  in Figure 3a, the optimal window length is obtained ( $B^*$  in Figure 3a) and the spectrogram with optimal window length can be visualized. Spectrograms obtained after solving for the optimal window length according to each of the three loss criteria  $\mathcal{L}_1$ ,  $\mathcal{L}_2$  and  $\mathcal{L}_3$  are shown in Figures 4a, 4b and 4c respectively.

In all cases the fault information is more clearly visible in the spectrogram as compared to the time domain signal shown in Figure 2. However, even though all of the aforementioned time-frequency representations are for exactly the same signal, it is clear that different loss criteria can lead to different optimal spectrograms. For example, the sparsity-based metric  $\mathcal{L}_1$  (Figure 4a) tends to magnify single fault impacts in the spectrogram, thereby disregarding weaker impulses present in the signal that could have confirmed the presence of regular repeating impulses characteristic of a fault. Conversely, metrics that account for the expected cyclo-stationarity in a faulty bearing, namely  $\mathcal{L}_2$  and  $\mathcal{L}_3$  (Figures 4b and 4c), have impulses of all transients at a similar intensity, but tend to smear energy in the spectrogram over the time axis. Ultimately, the optimal window length solution

is dependent on the selected loss criteria, with the loss criteria inspired by cyclo-stationarity analysis tending to choose optimal window lengths that capture most periodic impulses present in the signal.

Finally, the same simulated signal is used to demonstrate the benefits of having access to gradient information from the differentiable STFT during window length optimisation. For a random starting window length in the continuous range of (0 samples, 200 samples), the median number of function evaluations required to reach a window length optimisation convergence of one signal sample is recorded in Table 2. Four different optimisation methods are applied to each of the proposed loss criteria in Eqs. 7, 9 and 10, with each problem solved 30 times to obtain the median recorded in the table.

Table 2. Median (30 trials) number of function evaluations required to converge to a window length tolerance of one sample. Optimisation algorithms that exploit the gradients provided by the differentiable STFT tend to require fewer loss function evaluations to reach convergence.

Loss	Gradient-based		Gradient-free	
	Adam	BFGS	Nelder-Mead	Powell
$\mathcal{L}_1$	9.0	20.0	60.5	66.5
$\mathcal{L}_2$	23.0	11.5	63.0	63.5
$\mathcal{L}_3$	16.0	9.5	67.0	67.0

Two of the optimisation algorithms (Adam, BFGS) make use of the gradient of the loss criterion with respect to the window length. Conversely, the other two algorithms (Nelder-Mead, Powell) do not have access to the gradient information made available through the differentiable framework proposed in this work, although they still rely on the continuous nature of Eq. 2.

The first notable result is that using a continuous window function allows for the use for standard optimisation algorithms that require significantly fewer function evaluations as compared to the 200 function evaluations that would have been required if the full grid of window lengths in the (0, 200) range (tolerance of one sample) were computed. The second result is that the gradient-based optimisation algorithms requires significantly fewer function evaluations to reach the desired tolerance as compared to the gradient-free methods. In all of the cases tested here, even if the computational cost of computing the gradient equally expensive as computing the loss, the gradient based methods would be more computationally efficient on the whole. Therefore, the formulation presented here does not only yield smooth and continuous optimisation problems that can be solved using standard optimisation algorithms, but additionally, the differentiability of the approach ensures that the window length can efficiently be optimized for using gradient based optimisation.

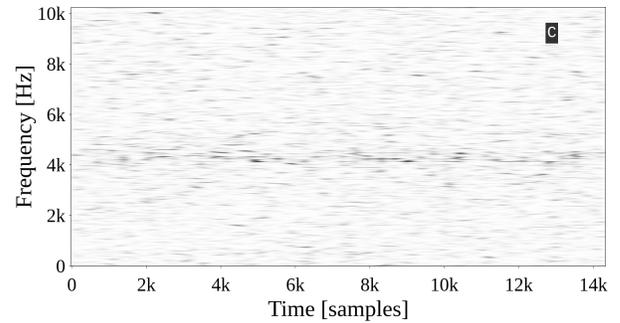
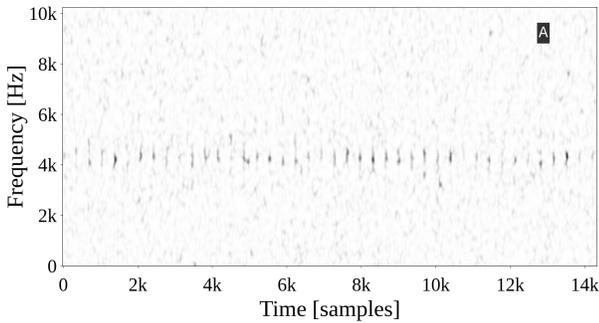
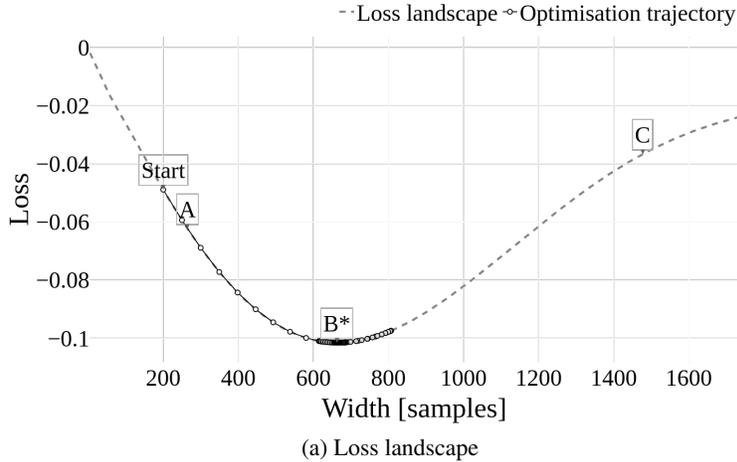


Figure 3. Optimisation landscape and candidate solutions for  $\mathcal{L}_2$ . The window lengths for non-optimal candidate solutions A and C are indicated on the loss landscape.

## 5. APPLICATION ON THREE EXPERIMENTAL DATASETS

The proposed window length selection method is evaluated on three experimental datasets, including two public datasets (Qiu, Lee, Lin, & Yu, 2006; *Case Western Reserve University Bearing Dataset*, n.d.) and an in-house dataset measured at KU Leuven, Belgium. The dataset specifications are listed in Table 3.

Table 3. Experimental dataset specifications.

Dataset	CWRU	IMS	KUL
Signal name	210	Test 2, Bearing 1	
Fault mode	Inner Race	Outer Race	Inner Race
Fault frequency [Hz]	159.92	236.43	193.51
Sampling frequency [Hz]	12000	20480	51200
Duration [s]	0.58	0.2	0.2

Models in the rest of this paper are optimised using the Adam optimiser with 100 optimisation steps and a learning rate of 50 to ensure full convergence. The learning rate is comparatively large since the optimisation problem is directly solved for continuous window length in number of samples. Signals

are further standardized to have a standard deviation of 1000 to avoid machine precision issues when computing the loss.

The optimal spectrograms for a given dataset and loss criterion are shown in Figures 5 to 7. Each set of figures show spectrograms for exactly the same signal from a given data set, but use different loss criteria for defining the window length optimally. Interestingly, the choice of loss criterion has a greater influence on some data sets as compared to others. For example, different loss criteria lead to very similar spectrograms for the IMS dataset (Figure 5), with none of the loss criteria leading to spectrograms that are different from the others. However, for the CWRU (Figure 6) and KUL dataset (Figure 7) the choice of loss criteria can have a significant influence in the resulting optimal spectrogram with some loss criteria defining significantly different optimal time-frequency trade-offs than others.

Similar to the results obtained for the simulated data set, the sparsity loss function  $\mathcal{L}_1$  tends to favour shorter optimal window lengths that magnify single impulsive events that are not necessarily related to the bearing fault frequency. Of the tested experimental signals, this behaviour is most apparent

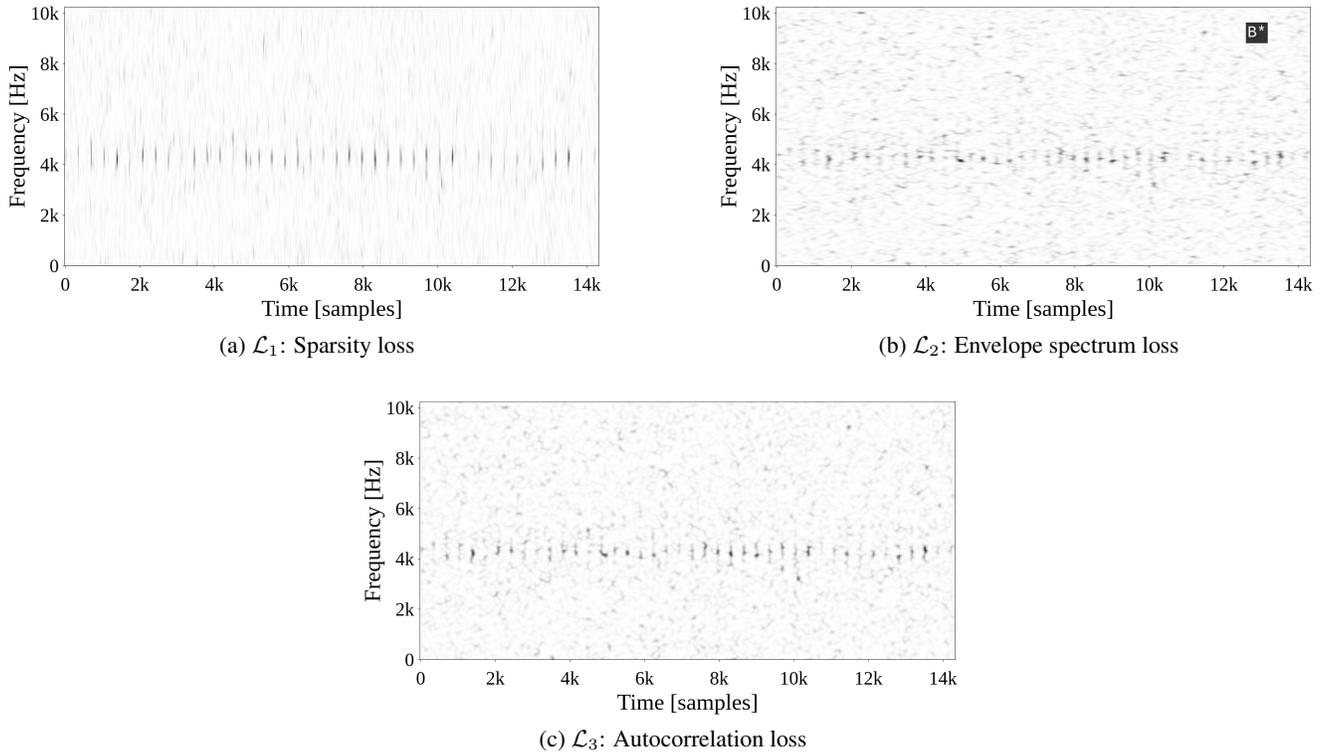


Figure 4. Optimal spectrograms for different loss criteria for simulated bearing fault signal. The sparsity-based loss  $\mathcal{L}_1$  tends to favour time resolution, magnifying single fault impacts in the spectrogram while the cyclo-stationarity based loss functions  $\mathcal{L}_2$  and  $\mathcal{L}_3$  smear energy in the spectrogram over the time axis.

for the KUL dataset (Figure 7a). Here, the  $\mathcal{L}_1$  sparsity loss (Figure 7a) chooses a window length that highlights a small subset of impulses and suppresses other impulses to the extent that they are barely perceptible. In contrast, the loss criteria  $\mathcal{L}_2$  and  $\mathcal{L}_3$  (Figures 7b and 7c), that account for the cyclo-stationary nature of the signal, tend to show all individual impulses at a similar intensity, albeit at the cost of having the impulses more smeared over the time axis. Interestingly, for the CWRU data set (Figure 6), the auto correlation loss  $\mathcal{L}_3$  (Figure 6c) also suffers from a similar problem as the sparsity loss  $\mathcal{L}_1$  with the optimal window length choosing to highlight a selection of high amplitude samples.

Ultimately, the results demonstrate that different loss criteria for defining spectrogram quality can lead to very different selected optimal window lengths, highlighting the importance of selecting the correct loss function for a given problem. The choice is however application specific. For instance, one can argue that for the KUL data (Figure 7) the sparse spectrogram of loss criterion  $\mathcal{L}_1$  (Figure 7a) will be most useful for diagnosing a fault visually from the spectrogram. On the other hand, for cases where the spectrogram is used in a subsequent signal processing step focused on the periodicity of the impulses in the spectrogram (e.g. cyclic modulation spectrum), loss criterion  $\mathcal{L}_2$  and  $\mathcal{L}_3$  would be more appropriate.

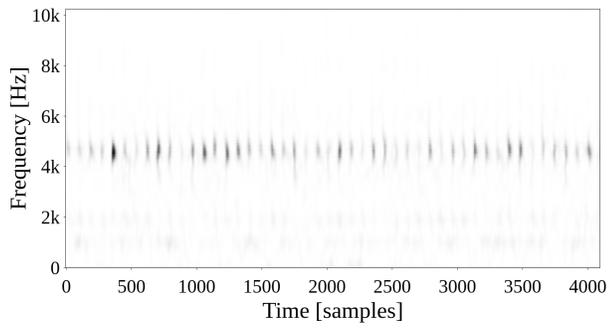
Depending on the desired result in the application, and the data at hand, the most appropriate loss criterion must be selected.

## 6. CONCLUSIONS AND FUTURE WORK

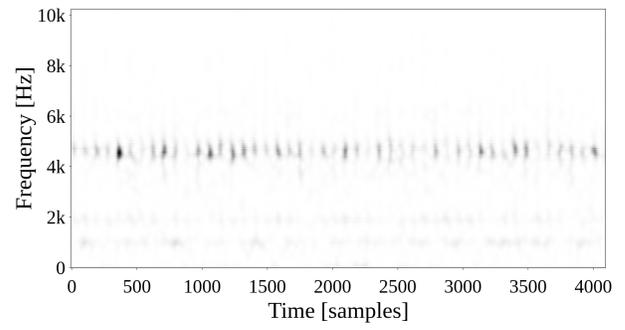
This paper presented a framework for making the STFT differentiable with respect to window length by using a continuous window over the entire signal duration. Three different loss functions for quantifying the spectrogram quality are applied to a simulated dataset and three experimental datasets.

The results demonstrate the value of a continuously variable window length that allows for efficiently solving for the optimal window length using gradient-free or gradient-based optimisers. Further, the importance of the appropriate unsupervised loss criterion is made apparent, with a sparsity-based loss criteria often highlighting isolated, high-amplitude transients, while loss criteria that respect the periodic nature of the bearing fault signals yield impulses with a more uniform power at each fault impact in the spectrogram.

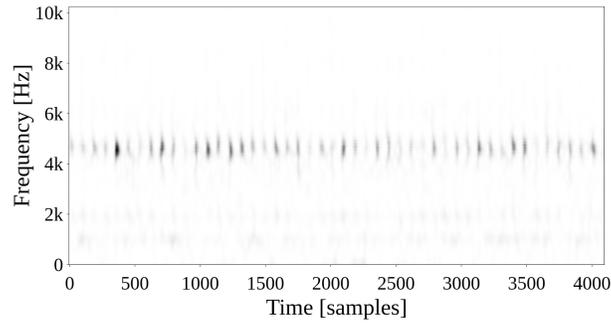
In future work, the proposed framework can be generalised to infer multiple window parameters and can be applied for window length selection in popular cyclo-stationary indicators



(a) IMS,  $\mathcal{L}_1$ : Sparsity loss



(b) IMS,  $\mathcal{L}_2$ : Envelope spectrum loss



(c) IMS,  $\mathcal{L}_3$ : Autocorrelation loss

Figure 5. Optimal spectrograms for different loss criteria: IMS dataset.

like the cyclic spectral correlation (Antoni, Xin, & Hamzaoui, 2017).

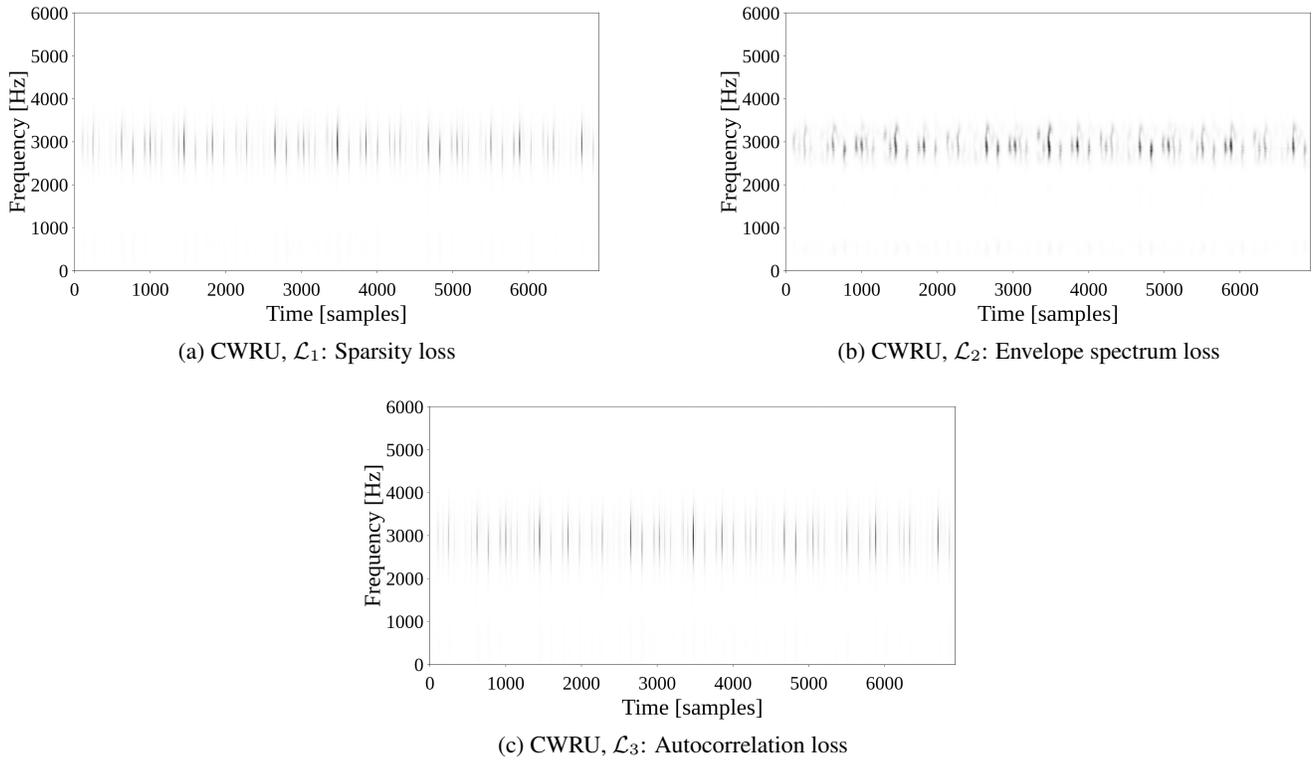


Figure 6. Optimal spectrograms for different loss criteria: CWRU dataset.

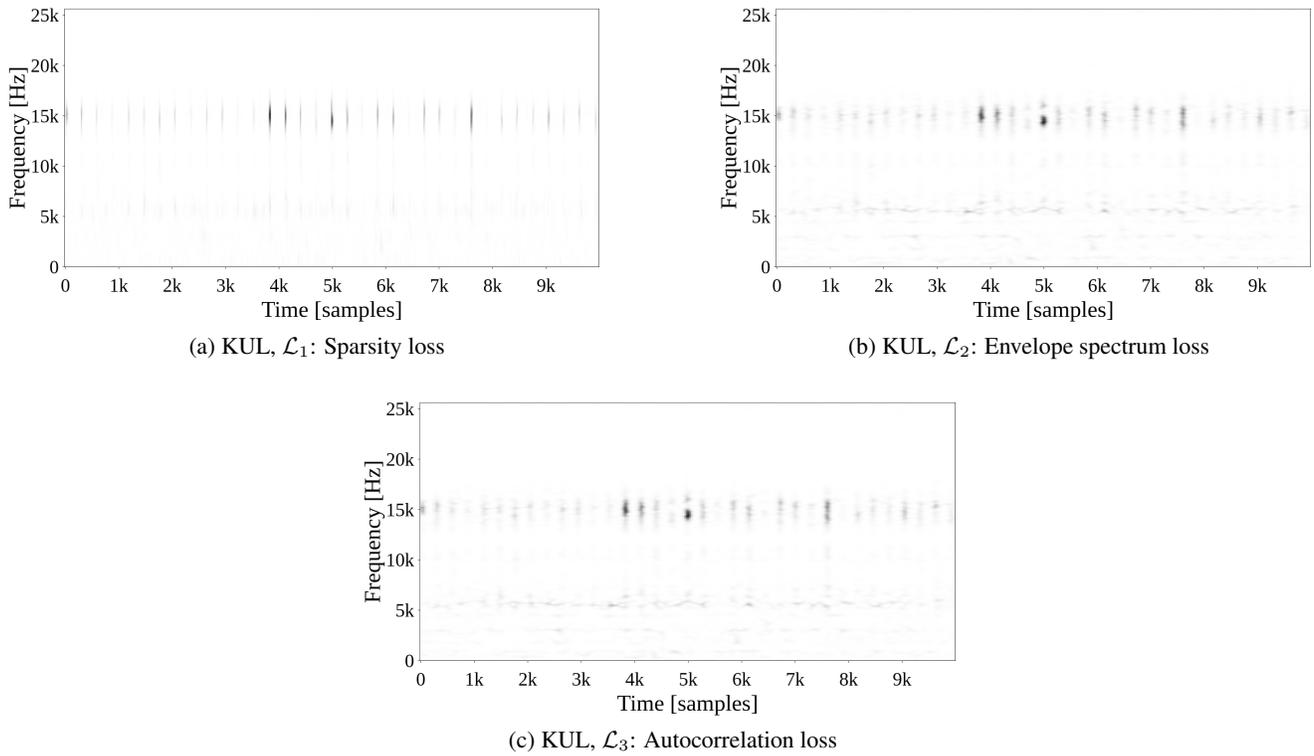


Figure 7. Optimal spectrograms for different loss criteria: KUL dataset.

## ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the European Commission under the Marie Skłodowska-Curie program through the ETN MOIRA project (GA 955681).

## REFERENCES

- Antoni, J. (2009, May). Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4), 987–1036. doi: 10.1016/j.ymssp.2008.10.010
- Antoni, J., Xin, G., & Hamzaoui, N. (2017, August). Fast computation of the spectral correlation. *Mechanical Systems and Signal Processing*, 92, 248–277. doi: 10.1016/j.ymssp.2017.01.011
- Case Western Reserve University Bearing dataset. (n.d.). <https://engineering.case.edu/bearingdatacenter>.
- Czerwinski, R., & Jones, D. (1997, February). Adaptive short-time Fourier analysis. *IEEE Signal Processing Letters*, 4(2), 42–45. doi: 10.1109/97.554468
- Fang, B., Hu, J., Yang, C., Cao, Y., & Jia, M. (2022, February). A blind deconvolution algorithm based on backward automatic differentiation and its application to rolling bearing fault diagnosis. *Measurement Science and Technology*, 33(2), 025009. doi: 10.1088/1361-6501/ac3fc7
- Leiber, M., Barrau, A., Marnissi, Y., & Abboud, D. (2022, August). A differentiable short-time Fourier transform with respect to the window length. In *2022 30th European Signal Processing Conference (EUSIPCO)* (pp. 1392–1396). Belgrade, Serbia: IEEE. doi: 10.23919/EUSIPCO55093.2022.9909963
- Leiber, M., Marnissi, Y., Barrau, A., & Badaoui, M. E. (2023, June). Differentiable Adaptive Short-Time Fourier Transform with Respect to the Window Length. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1–5). doi: 10.1109/ICASSP49357.2023.10095245
- Mateo, C., & Talavera, J. A. (2018, June). Short-time Fourier transform with the window size fixed in the frequency domain. *Digital Signal Processing*, 77, 13–21. doi: 10.1016/j.dsp.2017.11.003
- McFadden, P., & Smith, J. (1984, September). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, 96(1), 69–82. doi: 10.1016/0022-460X(84)90595-9
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... Lerer, A. (2017). Automatic differentiation in PyTorch. In *NIPS-W*.
- Qiu, H., Lee, J., Lin, J., & Yu, G. (2006, February). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, 289(4-5), 1066–1090. doi: 10.1016/j.jsv.2005.03.007
- Wodecki, J., Michalak, A., & Zimroz, R. (2021, February). Local damage detection based on vibration data analysis in the presence of Gaussian and heavy-tailed impulsive noise. *Measurement*, 169, 108400. doi: 10.1016/j.measurement.2020.108400
- Xie, H., Lin, J., Lei, Y., & Liao, Y. (2012, July). Fast-varying AM-FM components extraction based on an adaptive STFT. *Digital Signal Processing*, 22(4), 664–670. doi: 10.1016/j.dsp.2012.02.007
- Zhao, A., Subramani, K., & Smaragdis, P. (2021, June). Optimizing Short-Time Fourier Transform Parameters via Gradient Descent. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 736–740). doi: 10.1109/ICASSP39728.2021.9413704

## BIOGRAPHIES



**Douw Marx** received his B.S. and M.Sc degree in mechanical engineering from the University of Pretoria, South-Africa. He joined the Noise and Vibration Research Group in the Department of Mechanical Engineering, KU Leuven, Belgium as a PhD researcher in 2021. His research interests include fault detection using signal processing and unsupervised deep learning.



**Konstantinos Gryllias** holds a 5 years engineering diploma degree and a PhD degree in Mechanical Engineering from National Technical University of Athens, Greece. He holds an associate professor position on vibroacoustics of machines and transportation systems at the Department of Mechanical Engineering of KU Leuven, Belgium. He is also the manager of the University Core Lab Flanders Make @ KU Leuven MPro of Flanders Make, Belgium. His research interests lie in the fields of condition monitoring, signal processing, prognostics and health management of mech. & mechatronic systems.