Online shaft encoder geometry compensation for arbitrary shaft speed

profiles using Bayesian regression

D H Diamond*, P S Heyns, A J Oberholster

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria 0002, South Africa

Abstract

The measurement of instantaneous angular speed is being increasingly investigated for its use in a wide

range of condition monitoring and prognostic applications. Central to many measurement techniques are

incremental shaft encoders recording the arrival times of shaft angular increments. The conventional ap-

proach to processing these signals assumes that the angular increments are equidistant. This assumption

is generally incorrect when working with toothed wheels and especially zebra tape encoders and has been

shown to introduce errors in the estimated shaft speed. There are some proposed methods in literature

that aim to compensate for this geometric irregularity. Some of the methods require the shaft speed to be

perfectly constant for calibration, something rarely achieved in practice. Other methods assume the shaft

speed to be nearly constant with minor deviations. Therefore existing methods cannot calibrate the entire

shaft encoder geometry for arbitrary shaft speeds.

The present article presents a method to calculate the shaft encoder geometry for arbitrary shaft speed

profiles. The method uses Bayesian linear regression to calculate the encoder increment distances. The

method is derived and then tested against simulated and laboratory experiments. The results indicate that

the proposed method is capable of accurately determining the shaft encoder geometry for any shaft speed

profile.

Keywords: Instantaneous Angular Speed, Shaft encoder, Zebra tape, Geometry compensation, Bayesian

regression

1. Introduction

Instantaneous angular speed (IAS) is a measure that can be widely applied to rotating machinery diag-

nostics and condition monitoring. As such, methods of determining IAS and its usages are being intensively

 $^{\circ}$ Corresponding author $Email\ address:$ dawie.diamond@yahoo.com (D H Diamond*)

Nomenclature		
a	Constant angular jerk [radians ³ /s]	Greek Symbols
b	Constant angular acceleration	β Precision of observed observations
	[radians ² /s]	θ Encoder increment distance [radians]
c	Constant angular velocity offset [radians/s]	$oldsymbol{\Sigma}_0$ Covariance matrix of prior distribution
E	RMS error of estimated geometry [radians]	Σ Covariance matrix of posterior distribution
M	Number of recorded revolutions	σ Standard deviation of prior probabilities [radians]
N	Number of encoder sections	
T	Time interval between two zero crossing times [s]	θ Instantaneous Angular Velocity [radians/s]
t	Time [s]	au Variable representing time between two
\mathbf{A}	Coefficient matrix	zero-crossing times [s]
В	Observation vector	Subscripts
\mathbf{m}_0	Prior parameter vector	m Revolution number
m	Posterior parameter vector	n Shaft encoder increment number
$\mathrm{d}t$	Sampling interval for data acquisition [s]	Superscripts
x	Parameter vector	prior Indicates a prior probability

researched by the engineering community. The applications of IAS measurements are vast; some examples include predicting of rotor bar failure in squirrel-cage motors [1], detecting gear defects [2], diagnosing gas pressure faults in diesel engine cylinders [3] and determining turbine blade damage [4]. It has also been shown that small fluctuations in IAS can indicate early spall defects in roller bearings [5].

IAS is often measured through an incremental shaft encoder. The operating principle behind different types of incremental encoders are qualitatively similar. Incremental encoders use a set of periodically alternating sections mounted onto a shaft, along with a stationary probe which measures the sections as they pass the probe. The alternating sections can take the form of, for example, gear teeth mounted onto a shaft, or a zebra tape stuck around the circumference of the shaft. The stationary probe can be optical, Hall effect,

capacitive or magnetic [6, 5]. Some encoders, such as high-precision optical encoders, are manufactured with the alternating sections and the probe encased together in a single unit [7].

The operating principle of an incremental encoder is demonstrated by Fig. 1. The signal from the stationary probe is directly dependent on the presence of an alternating section at the probe. When a section is at the stationary probe the resulting voltage may be high and then, as the section leaves the measuring range of the probe, it gradually decreases. This results in a signal that exhibits a sine wave-like nature. The signal is then amplified, filtered and processed by a comparator, resulting in a TTL square wave signal as shown in Fig. 1 [8]. Each rising edge of the square wave indicates the precise time an alternating section has arrived at the probe. In this article, this precise time is referred to as the zero-crossing time of a section¹.

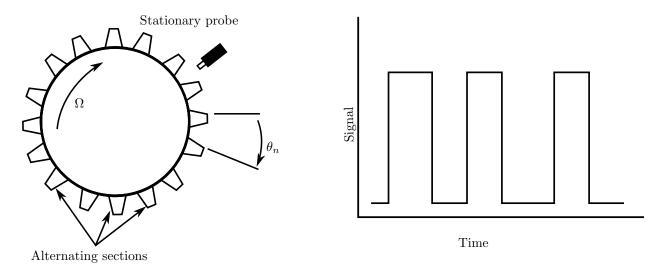


Figure 1: Operating principle behind incremental shaft encoders.

Different methods can be used to determine encoder zero-crossing times. For instance, the shaft encoder signal can be sampled using a signal analyzer and then processed afterwards to determine the zero-crossing times. Another approach, commonly referred to as the Elapse Time method, uses a high speed counter to determine the number of clock pulses between zero-crossing times. The clock speeds of these systems, typically between 80 - 100 MHz [6, 5, 9], are much higher than the acquisition rates of general purpose signal analyzers. It is therefore beneficial to use high speed counters where possible, as one would be able to resolve the zero-crossing times more accurately. Regardless of the method used to determine the zero-crossing times,

¹The term 'zero-crossing time' is sometimes used exclusively to indicate the exact time a signal changes sign, like a sine wave passing through 0 V. For the purposes of this article and related literature [6], the term 'zero-crossing time' is used to indicate the exact time a signal has passed a prescribed threshold value (2.5 V in the case of TTL signals).

the IAS must be calculated from the zero-crossing times.

The conventional method of determining IAS from two consecutive zero-crossing times assumes that the shaft speed is constant in that particular encoder section [8, 10, 11]. Mathematically, this is expressed by Eq. (1).

$$\dot{\theta}_n(t) = \frac{\theta_n}{t_{n+1} - t_n} \qquad t_n \le t \le t_{n+1} \tag{1}$$

In Eq. (1), $\dot{\theta}_n(t)$ denotes the IAS during the $n^{\rm th}$ section of the shaft encoder. θ_n denotes the angular distance of then the encoder section. The time, t, is restricted to being between then the and n+1 the zero-crossing times. If Eq. (1) is applied to all encoder increments for all measured revolutions, the IAS is determined. This approach is used in almost all IAS determination strategies. There are however many factors that can affect the accuracy of the resulting IAS measurement. Three factors that significantly affect the accuracy of the method are discussed below. All three of these factors result from invalid assumptions.

The first assumption is that one can resolve the zero-crossing times of the sections to an infinitesimal precision. The problem is that the raw shaft encoder signal is processed digitally. This enforces a discrete nature into the sampling of the signal. For instance, suppose the signal analyzer has a data acquisition rate of 100 kHz. The sampling interval of the system is thus calculated by Eq. (2) below.

$$dt = \frac{1}{100\ 000} = 1 \times 10^{-5} \text{ seconds}$$
 (2)

If the zero-crossing time occurs, for instance, 1×10^{-6} seconds after a sampled value, the acquisition system will miss this and record the zero-crossing time as having happened exactly at the next sample. This is one source of error and is called the quantisation error or quantification error [8, 12, 13]. The simplest way to mitigate the quantification error is to use the Elapsed Time method with a very high clock speed to determine the period between zero crossing times [6, 5, 9]. Another possible solution to this problem is the use of interpolation on the shaft encoder signal between two consecutive samples. However, this can only be performed if the sampled signal is not processed by the comparator. Some authors have also proposed low-pass filtering on the square-wave encoder signal. This smoothes the abrupt jumps between 'high' and 'low' and allows one to use interpolation to find the real zero-crossing time [13].

The next source of error emanates from the assumption that the shaft speed is constant between two zero-crossing times. From a mathematical point of view this is obviously a fallacy as it leads to discontinuities in the IAS, which posits an infinitely large acceleration at the zero-crossing times. Fig. 2 illustrates the effect of assuming a constant shaft speed between two encoder sections. When using IAS computationally, the fact that there are two different speed values at each zero-crossing time is often ignored. During

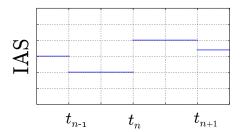


Figure 2: Discontinuities in IAS as a result of assuming constant angular speed between zero-crossing times.

post-processing only one of the shaft speeds is ascribed to each zero-crossing time. The typical assumption would be to ascribe $\dot{\theta}(t_n) = \dot{\theta}_n$, $\dot{\theta}(t_{n+1}) = \dot{\theta}_{n+1}$ etc. This results in a discrete representation of the IAS, meaning that the IAS is only known at the zero-crossing times.

This can lead to possible aliasing in the measured IAS. In assuming constant shaft speed between zero-crossing times one effectively misses higher frequency content that occurs between two zero-crossing times. There is no way to low-pass filter the signal in a way that will prevent aliasing. The only possible solution is to add a mechanical filter to the rotational system, such as a fly wheel [8, 12], or to have an incremental encoder with a sufficient number of sections. Fyfe and Munck [14] investigated the effect of using polynomial interpolation functions to determine the shaft instantaneous position to be used in resampling a vibration signal during computed order tracking. It was observed that using higher order interpolation functions reduced the noise floor of the order spectrum in the measured vibration signal, as opposed to assuming the IAS to be constant. To the best knowledge of the authors, the effect on the IAS by using polynomial interpolation functions between zero-crossing times of the encoder has not been investigated.

The third incorrect assumption often made is to assume all sections of a shaft encoder have equal angular distances. Equal section distances can be approached in very high-precision optical encoders [6] but is generally not realized for other types of shaft encoders. In one study conducted by Rémond [9], the coherence between the signal from two different optical encoders at a constant shaft speed as well as from two revolutions of the same optical encoder was determined. It was found that correcting for the section distances resulted in an insignificant improvement in measurement precision. This shows that optical encoders, when used in conjunction with a high speed counting clock (100 MHz in this article's case), can be used to measure IAS very accurately. One manufacturer of high-quality optical encoders, Heidenhain, report a variation of up to 5% in incremental distance [15]. This appears to be a conservative estimation from the manufacturer. Rivola and Troncossi [10] found a peak-to-peak variation in incremental distance of 20% in a zebra tape shaft encoder. For gears and optical encoders, manufacturing tolerances cause the sections to have different widths. Geometric inconsistencies in zebra tape systems are caused by inaccuracies during the printing of

the zebra tape as well as misalignment during the fastening process. The zebra tape but joint, where the start and end meet, is another source of geometric inequality. These geometric differences in a shaft encoder cause high order content in the IAS if left uncompensated for.

The focus of this article is the compensation of geometric inconsistencies between shaft encoder sections. The existing techniques for performing geometry compensation can be classified as either offline or online calibration methods.

The offline calibration methods require the use of a calibration run or a calibration measurement to determine the shaft encoder section distances. The shaft encoder signal is then used with each encoder section distance to calculate the IAS with Eq. (1).

Leclere et al. [13] performed calibration tests on two shaft encoders mounted onto a work piece driven by a Numerically Controlled (NC) lathe. For calibration of the geometric differences, the authors used a method that averaged out the geometric errors. They synchronously averaged the duration between zerocrossing times for each encoder section over many revolutions. This method requires that the rotation speed of the shaft be constant, or that there be no integer multiple periodic deviations of shaft speed. The authors used the calibration method for different shaft speeds and found good correspondence among the calibration results from different speeds. They noted, however, that the condition of constant shaft speed was very difficult to achieve, even on a NC lathe.

Rivola and Troncossi [10] performed IAS measurements on a camshaft of a racing motorbike engine. The purpose of the experiment was to determine the camshaft angular position at the opening and closing of each valve. They used a zebra tape with an optical probe as the shaft encoder. They noted that the zebra tape had unequal geometric increments, especially at the butt joint, and devised a simple method to compensate for the geometry of the encoder. Before installing the cam-valve mechanism to the camshaft, a calibration test was conducted. The camshaft was disconnected from the motor that would be used in the experiment and actuated by a pneumatic torque wrench. When the camshaft reached approximately 1000 RPM, the torque wrench was disconnected. Shaft encoder measurements were then taken for about two seconds. The reasoning behind this approach is that, since the camshaft was not being driven by any mechanism during the acquisition time, the shaft speed should decrease linearly without any torsional vibrations.

They then proceeded to assume a constant shaft speed for each revolution, and a linear speed profile was fitted to the set of shaft speeds over time using a least squares approach. The result was a linearly varying IAS which they then assumed to be accurate, and they could thus determine the encoder section distances.

The other group of methods are called online methods here. These are methods that do not use a separate calibration measurement to determine the encoder section distances. The calibration is performed using the signal that is contaminated by the torsional vibrations of interest.

Resor et al. [6] proposed a method for calibration that can be applied to shafts with nearly constant speeds. They synchronously averaged the time between section passages as a percentage of the rotation period for each different shaft revolution. This allowed them to take into account small variations in shaft speed and therefore perform more accurate geometry compensation. Although the method allows for slight variations of shaft speed between revolutions, the speed within a revolution is not permitted to vary. The method also cannot be used where the shaft speed is transient, such as during a run-up. Braut et al. [16] used a method very similar to Resor et al. [6] to compensate for the geometry of a zebra tape on a test bench while experimenting with the effect of gears on IAS.

Janssens et al. [17] also developed an online method for shaft geometry compensation of a zebra type. The method they proposed can be used for shaft steady or transient conditions. However, it assumes the use of a high-quality zebra tape with no misalignment during fastening, so that the only source of error is the butt joint increment. This method can therefore determine the position of the butt joint from the signal and calculate the section distance of the butt joint, assuming all other sections are equal.

In reviewing the relevant literature, the current limitations for geometric compensation methods can be summarised as follows:

- 1) Many methods make use of a calibration run. The shaft speed is required to be either constant or linearly varying and void of torsional vibration. These requirements can be met approximately under laboratory conditions but cannot be guaranteed in real working conditions.
- 2) Current online methods assume either a constant shaft speed or constant increment distances, except at the butt joint. Both these conditions are limiting, because they do not allow one to measure transient conditions and because geometric effects are invariably present throughout all sections in most shaft encoders.

The need is thus identified for an online calibration method that allows the shaft speed the flexibility of transient states and that calculates every encoder section distance. Such a method can be useful in a variety of applications, for example:

•Online calibration when constant speed calibration data is not available.

- Online calibration under temperature fluctuations. Thermal expansion of the shaft or encoder segments can alter the geometric pattern of the encoder [6].
- Online calibration when the stationary probe changes position. This can be because the probe was
 accidentally bumped or had to be moved.

The focus of this article is thus on deriving a novel algorithm that can perform the identified calibration action. The method is derived and then validated by means of simulated and laboratory experiments.

2. Shaft encoder geometry compensation

2.1. Assumptions

Suppose a shaft encoder with N sections is used to record data over M revolutions of a shaft. Four assumptions pertaining to the shaft encoder and the IAS are made:

- 1) Each shaft section distance remains constant for all M revolutions.
- 2) The sum of all N sections add up to one whole revolution, or 2π radians.
- 3) The IAS between two zero-crossing times can be expressed as a second order polynomial. This is in contrast to the traditional notion that the IAS is constant (zero order polynomial) between two zerocrossing times.
- 4) The IAS is a continuous function throughout time. This practically means that no abrubt jumps in the IAS at zero-crossing times are allowed.

2.2. Derivation

First, a variable representing time is introduced in Eq. (3).

$$\tau_{n,m}(t) = t - t_{n,m} \qquad t_{n,m} \le t \le t_{n+1,m}$$
(3)

$$\tau_{n,m}(t_{n+1}) = T_{n,m} \tag{4}$$

In Eq. (3), $t_{n,m}$ represents the zero-crossing time of encoder section n in revolution m. The time, t, represents any time instant between $t_{n,m}$ and the next zero-crossing time, $t_{n+1,m}$. The time period between two consecutive zero-crossing times is represented by $T_{n,m}$. For increment n in revolution m, the IAS can be expressed as in Eq. (5).

$$\dot{\theta}_{n,m}(\tau_{n,m}) = a_{n,m}\tau_{n,m}^2 + b_{n,m}\tau_{n,m} + c_{n,m} \qquad 0 \le \tau_{n,m} \le T_{n,m}$$
 (5)

In Eq. (5), $\dot{\theta}_{n,m}(\tau_{n,m})$ is the IAS, $a_{n,m}$ is the constant angular jerk, $b_{n,m}$ is the constant angular acceleration and $c_{n,m}$ the speed offset during section n in revolution m. To calculate each section distance, Eq. (5) is

integrated, resulting in Eq. (6). Recall that $\theta_{n,m}$ is constant for all m. The subscript m is therefore dropped for the incremental distance.

$$\theta_n = \int_0^{T_{n,m}} \dot{\theta}_{n,m}(\tau_{n,m}) d\tau_{n,m} = \frac{1}{3} a_{n,m} T_{n,m}^3 + \frac{1}{2} b_{n,m} T_{n,m}^2 + c_{n,m} T_{n,m}$$
 (6)

Enforcing continuity of speed between sections results in Eq. (7).

$$\dot{\theta}_{n,m}(T_{n,m}) = \dot{\theta}_{n+1,m}(0)
a_{n,m}T_{n,m}^2 + b_{n,m}T_{n,m} + c_{n,m} - c_{n+1,m} = 0$$
(7)

Enforcing continuity of acceleration between sections results in Eq. (8).

$$2a_{n,m}T_{n,m} + b_{n,m} - b_{n+1,m} = 0 (8)$$

Substituting Eq. (6) into Equations (7) and (8) by cancelling out $a_{n,m}$ results in Equations (9) and (10) respectively.

$$3\theta_n - \frac{1}{2}b_{n,m}T_{n,m}^2 - 2c_{n,m}T_{n,m} - c_{n+1,m}T_{n,m} = 0$$
(9)

$$6\theta_n - 2b_{n,m}T_{n,m}^2 - b_{n+1,m}T_{n,m}^2 - 6c_{n,m}T_{n,m} = 0$$
(10)

Writing out Equations (9) and (10) for all values of n and m results in a set of equations. There are N+2MN unknowns throughout all equations and there are 2MN-2 equations (the last section in the last revolution is not included in the analysis as it would add two unknowns from the next revolution.). The final equation is added to the system to ensure that the sum of all N sections adds up to one revolution, as illustrated in Eq. (11).

$$\sum_{n=1}^{N} \theta_n = 2\pi \tag{11}$$

Equations (9), (10) and (11) are now assembled into a system of equations as in Eq. (12) shown along with the corresponding matrix dimensions.

$$\mathbf{A} \underset{(2MN-1\times N+2MN)}{\mathbf{x}} \underset{(N+2MN\times 1)}{\mathbf{B}} = \mathbf{B}$$
(12)

The parameter vector, \mathbf{x} , contains all the unknown values. The first N values of \mathbf{x} are $\{\theta_1, \theta_2, \ldots, \theta_N\}^\intercal$. The remaining 2MN values of \mathbf{x} are $\{b_1, b_2, \ldots, b_{1}, c_{1}, c_{2}, \ldots, b_{N,M}, c_{N,M}\}^\intercal$. All the elements of the observation vector, \mathbf{B} , are zero (from Eq. (9) and Eq. (10)) except the first value, which is 2π (from Eq. 11). The coefficient matrix, \mathbf{A} , consists of all the parameter coefficients in Equations (9)-(11). The system of equations described by Eq. (12) has N + 2MN unknowns and only 2MN - 1 equations. The system is therefore underdetermined and has an infinite number of solutions.

The equation can be solved for \mathbf{x} by noting that prior information about the encoder sections are available. Although the encoder sections do not have exactly equal angular distances, they are at least manufactured with that intent. This thus serves as a point of reference and allows one to solve Eq. (12) with Bayesian linear regression.

2.3. Bayesian linear regression

Bayesian linear regression is a probabilistic regression technique that finds its origin in Bayes's theorem. The derivation of the regression technique is well established and can be found in [18]. The crux of the method lies in the fact that one can choose prior probabilities for system parameters, which leads to accurate inferences. The fact that the prior probabilities are incorrect is tolerable. As mentioned earlier, the prior belief with regards to the encoder sections is that they have equal distances. This is expressed in Eq. (13).

$$\theta_n^{\text{prior}} = \frac{2\pi}{N} \tag{13}$$

The prior probabilities of the other unknown quantities must now be determined. A reasonable assumption to make is that the IAS is constant between two zero-crossing times, i.e. $a_{n,m}^{\text{prior}} = 0$ and $b_{n,m}^{\text{prior}} = 0$. This is the assumption made in most published literature concerning IAS, as was described in the Introduction. This leads one to the prior belief as expressed in Eq. (14).

$$c_{n,m}^{\text{prior}} = \frac{\theta_n^{\text{prior}}}{T_{n,m}} \tag{14}$$

All prior quantities are expressed as a multivariate Gaussian distribution with a mean vector denoted by \mathbf{m}_0 and a covariance matrix denoted by Σ_0 . Both these quantities are expressed in Eq. (15).

In Eq. (15), Σ_0 is an identity matrix multiplied by a variance parameter σ^2 . The variance describes how accurate the mean vector is thought to be. A large variance describes low accuracy and a small variance describes high accuracy. In Bayesian linear regression a large variance is often chosen so as to not let the

prior probability dominate the inference. In Bayesian theory the variance is called a hyperparameter.

The goal of an inference is to determine the posterior probability from a likelihood function and a prior probability. The prior probability is described by a Gaussian multivariate distribution with a mean and covariance given in Eq. (15). The likelihood function indicates how likely the parameter vector is to have generated the observed data. The observed data in this case is each element in the vector \mathbf{B} from Eq. (12). The probability of having observed these values is again expressed by a Gaussian distribution with mean value centred on each value in \mathbf{B} and a precision parameter β . The precision parameter is the inverse of probability distribution's variance and is also chosen at the beginning of the inference. If the precision parameter takes on a small value, it implies that there is a lot of noise in the observed values. If the precision parameter takes on a large value, it implies that the observed values contain less noise.

All the values are now in place to perform Bayesian linear regression. The posterior probability is described by a multivariate Gaussian with a mean \mathbf{m} and a covariance matrix Σ . The covariance matrix can be calculated directly with Eq. (16) and the mean can be calculated with Eq. (17).

$$\Sigma = \Sigma_0 + \beta \mathbf{A}^\mathsf{T} \mathbf{A} \tag{16}$$

$$\mathbf{m} = \mathbf{\Sigma}^{-1} \left(\mathbf{\Sigma}_0 \mathbf{m}_0 + \beta \mathbf{A}^{\mathsf{T}} \mathbf{B} \right) \tag{17}$$

The result is therefore a probability distribution over each of the section distances, acceleration terms and speed offset values. The probabilities ascribed to each value can be informative, but at this stage it is assumed that the mean of the posterior distribution describe the solution to the initial problem, shown in Eq. (18).

$$\mathbf{x} = \mathbf{m} \tag{18}$$

The IAS can easily be calculated by substituting the relevant values into Equations (7) and (8), solving the jerk values, and then ultimately substituting the values into Eq. (5).

3. Experimental validation

3.1. Numerical validation

3.1.1. Numerical setup

Four simulated experiments were conducted to validate the proposed method. The purpose of the method is to recover the encoder section distances from knowledge of only the number of sections and the zero-crossing times. The Root of Mean Square (RMS) error is defined as in Eq. (19).

$$E = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(\theta_n^{\text{real}} - \theta_n\right)^2}$$
 (19)

In Eq. (19), θ_n defines the estimate for section n. Four different transient shaft speeds are simulated. The four shaft speeds are given in Equations (20-23), in units of radians/second. A description of each angular speed time history is presented next to each equation.

$$\dot{\theta}_1(t) = 100 + 10t \text{ (Linear ramp up)} \tag{20}$$

$$\dot{\theta}_2(t) = 100 + 5\sin((2\pi)30t)$$
 (Sinusoidal speed with offset) (21)

$$\dot{\theta}_3(t) = 100 + 10t + 5\sin((2\pi)30t)$$
 (Linear ramp up with sinusoidal vibration) (22)

$$\dot{\theta}_4(t) = 100 + 5\sin((2\pi)30t) + 4\sin((2\pi)45t) + 2\sin((2\pi)20t) \text{ (Mix of sinusoids)}$$
 (23)

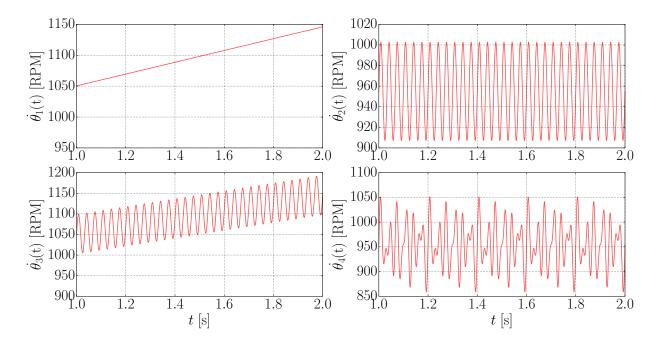


Figure 3: Four different speed profiles for numerical validation.

Graphical representations of the four angular speed profiles are presented in Fig. 3. The choice of constant speed offset, 100 radians/second, corresponds to approximately 955 RPM. This rotational speed was chosen as it is within the operating speed of many turbomachines. The choice of the four different transient speed conditions is quite arbitrary. The only criterion is that they represent transient conditions. The simulations were performed for a duration of 10 seconds. The sampling rate of the numerical shaft encoder was 1 MHz. This sampling rate was thought to be high enough to attenuate the quantisation noise in the signal. This is permitted as the focus of this article is geometry compensation. For employing the proposed method, the

two hyperparameters were chosen as in Eq. (24).

$$\sigma = 10$$

$$\beta = 1 \times 10^{10}$$
(24)

The standard deviation of the prior probabilities, σ , is wide, meaning little confidence is ascribed to the prior probabilities. The precision parameter, β , is very large, which means the observed values are thought to be approximately noise free. The combination of these parameters was found to give satisfactory results, and therefore the authors advocate using these two parameters when starting an analysis.

The geometry of the encoder was chosen so as to be similar to the encoder geometry from the laboratory experiments. Once again, the choice of geometry is arbitrary, the only requirement is that the encoder sections differ in distance from one another. The encoder has 79 unequal section increments. A graphical illustration of the encoder geometry is given in Fig. 4.

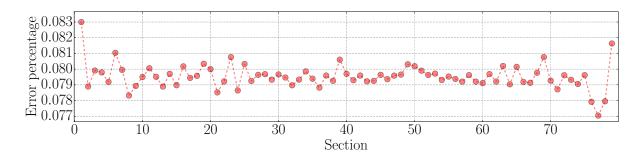


Figure 4: Simulated encoder geometry for numerical validation

If one were to assume all sections in Fig. 4 as equal one would be making an error for each section as shown in Fig. 5.

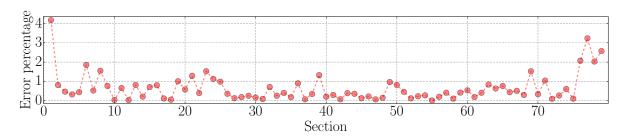


Figure 5: Error in encoder geometry estimate if one assumes constant angular sections.

3.1.2. Solution philosophy

The number of revolutions used to perform calibration, M, is an important factor to take into consideration. Note that one dimension of matrix \mathbf{A} scales linearly with the number of revolutions. Consequently,

the solution time of Eq. (17) scales with M. The accuracy of the calibration as well as the solving time of the system is therefore tested as a function of the number of revolutions used. The calibration is conducted for M = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55. Two philosophies are used to calibrate the section distances:

- 1. At the start of each calibration, the prior probabilities pertaining to the encoder increment distances are assigned as in Eq. (13) and (14), i.e. all shaft increments are assumed to be equal.
- 2. Each calibration run uses as a prior probability the result of the previous calibration run. The first run (M=5) therefore uses the assumption of equal increments and the second run uses the calculated increment distances from the first run. From there on the results are passed on for each calibration run. The effect of the two philosophies on the accuracy of each successive calibration run will be investigated.

3.1.3. Numerical validation results

For illustrative purposes, the estimated geometries from the first and fourth speed profiles for M=55 are shown in Fig. 6. These were found to be the most and least accurate estimates, respectively, for M=55. From Fig. 6 it can be seen that the estimated encoder geometry for $\dot{\theta}_1$, the most accurate estimate, is

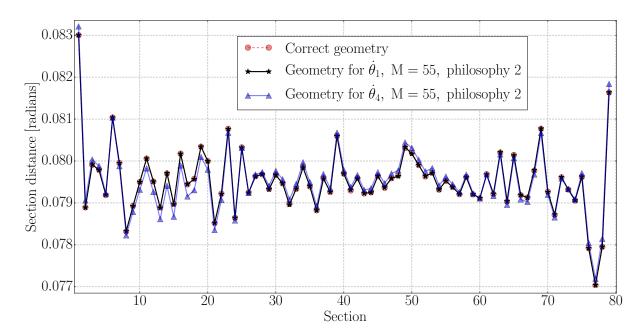


Figure 6: Estimated encoder geometry for the first and fourth speed profiles solving with M=55 and using philosophy 2.

nearly identical to the real geometry. The least accurate calibration, $\dot{\theta}_4$, performs less accurately than for $\dot{\theta}_1$ but still captures the underlying structure of the geometry well. The percentage errors for each section as obtained for all four speed profiles and both solution philosophies are shown in Fig. 7 on a linear scale

and in Fig. 8 on a log scale.

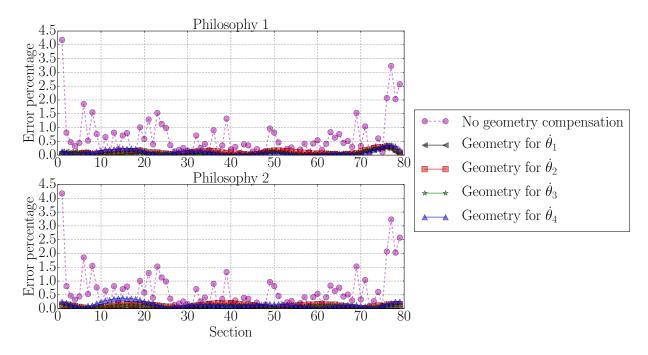


Figure 7: Section percentage errors for M=55 as calculated from all four numerically tested shaft speeds. Results shown on linear scale.

By investigating Fig. 7 and Fig. 8 it can be seen that for all solved cases, the total estimated geometry is more accurate than with no geometry compensation, even for highly nonlinear speed profiles. The most accurate estimated geometry is for $\dot{\theta}_1$ using philosophy 2. It is difficult to compare the accuracies of the other calibration results from the figures. To investigate this, the RMS error is shown for all test cases in Fig. 9.

A few observations can be made from examining Fig. 9. In all cases where $M \ge 10$ the geometry compensation improved the estimated geometry. The most accurate geometry compensation was performed for the linear ramp up by using philosophy 2, i.e. continually updating the prior probabilities for the geometry.

When no geometry compensation is performed, it results in the most inaccurate estimate. The performance of the proposed method can be evaluated by comparing the RMS values from the proposed method to the RMS value of the no compensation estimate. The least accurate estimate for M=55 is 82% more accurate than the no compensation estimate. The most accurate estimate is 99% more accurate than the no compensation estimate.

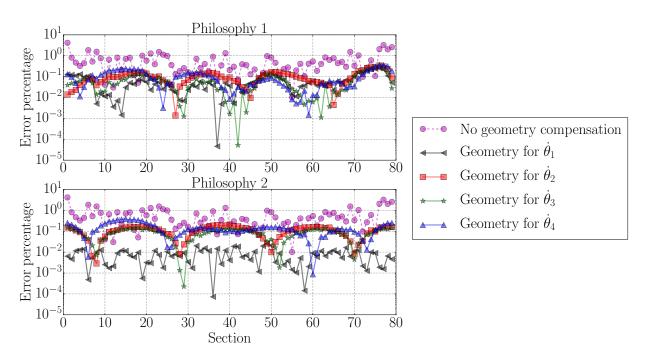


Figure 8: Section percentage errors for M=55 as calculated from all four numerically tested shaft speeds. Results shown on log scale.

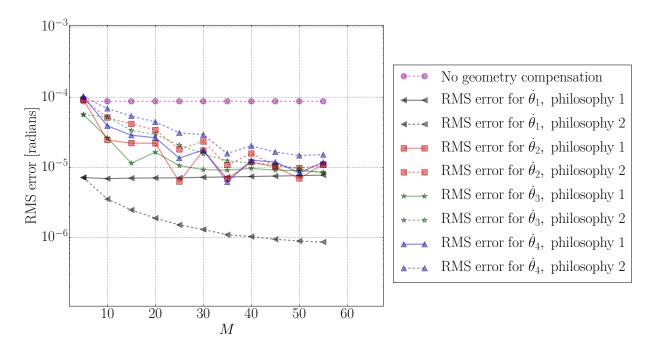


Figure 9: RMS error values for all four numerical speed profiles as a function of ${\cal M}.$

It is interesting to note, however, that solution philosophy 2 is not more accurate for all speed profiles. For $\dot{\theta}_2$, $\dot{\theta}_3$ and $\dot{\theta}_4$ philosophy 1 performs more accurately than philosophy 2. This is as a result of the highly nonlinear nature of these speed profiles. The first estimate (M=5) for these three cases is relatively inaccurate. Using this inaccurate estimate for the prior probability of the next inference seems to bias the result in a less beneficial way than just using equal shaft increments. However, it must be noted that for large M values the two philosophies converge. It is entirely possible that for even larger M values philosophy 2 will outperform philosophy 1. This was not experimented with for the purposes of this article as it was prohibitively expensive.

The solving time for each speed profile using the two philosophies is shown in Fig. 10. From looking at

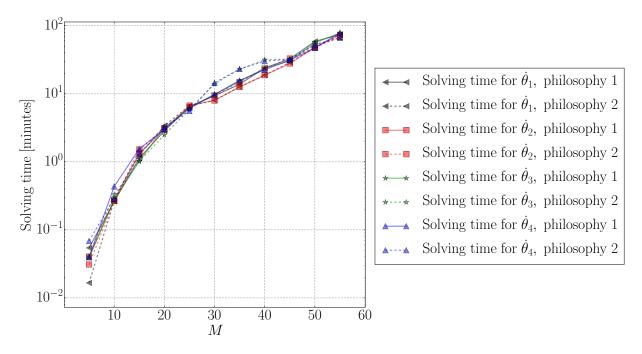


Figure 10: Solving times for all calibration methods.

Fig. 10 it can be seen that the solving time starts out relatively short at 0.05 minutes and becomes rather long at approximately 90 minutes. It can also be seen that there is no considerable difference between the solving times of philosophies 1 and 2. A general trend can be established that the solving time doubles for each 5 revolutions added. The computations were performed on a desktop computer with eight 3.5 GHz cores and 32 GB memory. The Python programming language was used and run using a single core. A sparse solver was used to solve for \mathbf{x} in each case.

Two comments on the simulated results must be noted. Firstly, highly nonlinear shaft speed profiles were

used in the numerical validation. In practice, the shaft speed profile will generally be approximately linear with small deviations or steady with small deviations. It is therefore reasonable to expect the proposed method to be able to handle all shaft speed profiles encountered in practice. Secondly, it can also be noted by looking at Fig. 9 that using increasingly large values of revolutions will result in increasingly accurate geometry estimates. It is therefore feasible to use many revolutions for calibration. The solving time will be long but one will have a very accurate estimate for the geometry which can henceforth be used for the calculation of IAS.

3.2. Laboratory experiments

3.2.1. Experimental setup

An experimental setup was used to test the proposed method in a practical situation. The setup consisted of a 50 mm diameter rotor supported on two self-aligning ball bearings. The rotor was connected to a three phase motor through a flexible shaft coupling. The variable speed motor has a controller accepting input signals in the range 0 - 10 V. The theoretical maximum speed of the rotor is 1500 RPM. A zebra-tape based shaft encoder was utilised to measure the shaft speed. The zebra tape has 79 sections and a stationary optical probe. It is assumed that the zebra tape sections are not equidistant and it is the goal of the experiments to determine this geometry.

The rotor was also instrumented with an additional optical shaft encoder. A Heidenhain optical encoder with 5000 pulses per revolution was fastened to the shaft. The signal from the Heidenhain shaft encoder was passed through a divide-by-10 counter, effectively changing the pulses per revolution of the encoder to 500. This was necessary as the fine angular resolution of the encoder resulted in large amounts of quantisation noise. The rotor system is shown in Fig. 11. The zebra tape encoder and the Heidenhain encoder signals were sampled with an OROS OR35 signal analyser. The maximum sampling rate of the analyser is 102.4 kHz. LabView was used to generate input signals for the rotor system by inserting the desired speed profile as a mathematical formula. A National Instruments NI9264 analogue output card sent the signals from LabView to the motor controller. The experimental setup allowed for the versatility of creating many shaft speed profiles to thoroughly test the proposed method. The experimental setup also allowed for independent calibration of the zebra tape geometry, which is the subject of the next paragraph.

3.2.2. Shaft encoder reference geometry calibration

The zebra tape geometry was determined independently and used as a reference geometry for evaluating the accuracy of the proposed method. The reference geometry was established through using two different methods. Both methods approximate the correct encoder geometry. To the best of the authors' knowledge,

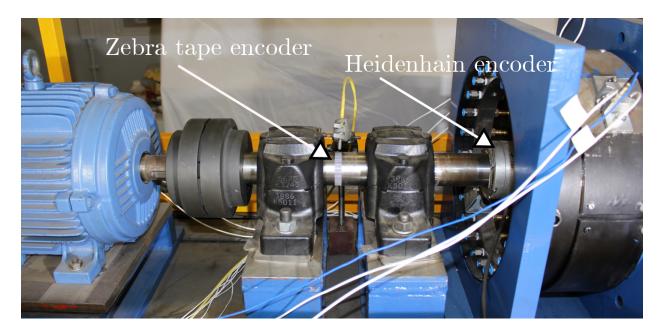


Figure 11: Rotor system for experimental validation.

there are no published methods that claim to be able to perfectly calibrate a shaft encoder geometry.

For both methods, a constant voltage input signal was generated and sent to the motor controller. This resulted in a nearly constant shaft speed of approximately 406 RPM. Lowering the shaft speed further resulted in a less steady shaft speed. As both calibration methods work best for steady speeds, the shaft speed was not lowered further. Both the zebra tape encoder signal and the Heidenhain encoder signal were measured with the signal analyser for 10 seconds. This resulted in 67 complete revolutions of the shaft. All complete revolutions were used in the calibration of the reference geometry.

The first calibration method exploited the fact that the Heidenhain encoder has a finer resolution than the zebra tape encoder. For every pulse of the zebra tape encoder, several pulses from the Heidenhain encoder occurred. A section distance of the zebra tape encoder was therefore estimated by counting the number of Heidenhain encoder pulses occurring inside one zebra tape pulse. The number of pulses were then multiplied by a factor of $2\pi/500$, converting it to radial angular distance.

The underlying assumption here is that all sections of the Heidenhain encoder are equal in width. It has already been mentioned in the literature review that errors of up to 5% in section width can be present [15]. This is one reason why the current method is only approximate. Within one zebra tape encoder pulse, several whole Heidenhain encoder pulses and two partial pulses (one at the start and one at the end of the zebra tape pulse) occur. It was assumed that the shaft speed in every partial pulse is constant, which also

contributes to the approximate nature of the calibration method. The percentage of the Heidenhain encoder pulses occurring inside the zebra tape pulse could then easily be calculated. Fig. 12 shows a zebra tape pulse with the corresponding whole and partial pulses from the Heidenhain encoder.

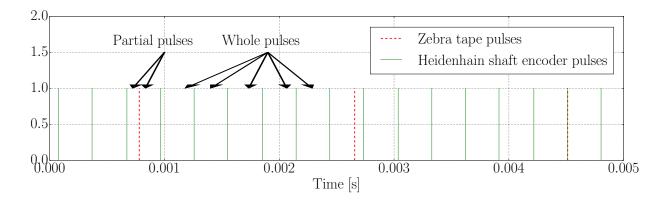


Figure 12: Illustration of whole and partial pulses of the Heidenhain encoder in a single pulse from the zebra tape encoder.

The second reference calibration method, proposed by Resor et al. [6], synchronously averages the encoder section distances over many revolutions by assuming the shaft speed within every revolution remains constant. Resor et al. [6] did not explicitly evaluate the accuracy of the method in determining the real shaft encoder geometry. They did, however, show that application of the method dramatically reduced the high order content in the torsional vibration frequency spectrum. Since uncompensated-for geometric inaccuracies manifest as high order content, it can be assumed that the method performs well in calibrating the shaft encoder geometry.

Fig. 13 shows the average encoder section distances for both methods over the 67 measured revolutions. From Fig. 13 it can be seen that both methods obtain similar reference geometries. This gives one confidence that the reference geometry is an accurate representation of the true geometry. The question now arises as to which of the two reference geometries should be used as the absolute reference geometry. Fig. 14 shows the standard deviation of every section for all revolutions for both calibration methods.

It can be seen from Fig. 14 that the method proposed by Resor et al. [6] results in a much smaller standard deviation per section than the first method. This result is counter intuitive, as the literature consulted indicates that optical shaft encoders are generally very precise [9]. Note that the optical encoder has roughly six times more pulses per revolution than the zebra tape encoder, making it much more susceptible to quantification errors. This occurred even though the pulses were divided by 10. It is therefore noted that the optical encoder signal needs to be captured at much higher acquisition rates to be used in the current application.

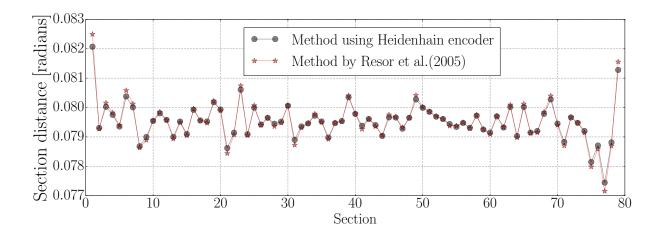


Figure 13: Average encoder section distances for both methods of calibration

It is reasonable to assume that the most accurate calibration method is the one with the smallest standard deviation. The reference geometry as calculated with the method by Resor et al. [6] was therefore used. It is assumed that this reference geometry is sufficiently accurate to allow for a meaningful accuracy evaluation of the proposed method.

3.2.3. Experimental procedure

To validate the proposed method, thorough testing on a variety of shaft speed profiles was conducted. For continuity within the article, four speed profiles similar to those in the numerical validation were tested. The speed profiles were generated by inserting the mathematical formula representing the profile into the LabView signal generator. The transient signals were then sent to the motor controller through the NI9264 digital to analogue card. In general, the shaft speeds did not exceed 600 RPM. This was done to minimise quantisation noise in the measurements. This is permitted as the focus of this article is not to investigate quantisation noise. The four shaft speeds as inserted into the LabView program are given in Equations (25-28) in units of radians/s.

$$\dot{\theta}_1(t) = 4.3t \text{ (Linear ramp up)}$$
 (25)

$$\dot{\theta}_2(t) = 15.65 + 4.9\sin((2\pi)1.5t) \text{ (Sinusoidal speed with offset)}$$
(26)

$$\dot{\theta}_3(t) = 4.3t + 4.9\sin((2\pi)1.5t)$$
 (Linear ramp up with sinusoidal vibration) (27)

$$\dot{\theta}_4(t) = 15.65 + 4.9\sin((2\pi)1.5t) + 3.92\sin((2\pi)2.25t) + 1.96\sin((2\pi)t)$$
 (Mix of sinusoids) (28)

The speed profiles are similar in form as those used in the numerical validation. One significant difference is the frequencies of the sinusoidal terms. It was necessary to lower these, as the rotational system did not respond desirably to the same frequencies as in numerical validation. This does not affect the integrity

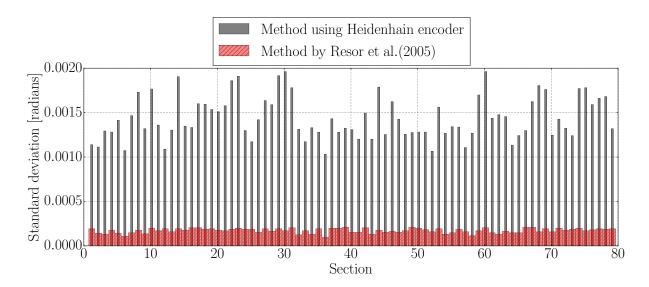


Figure 14: Standard deviations for every encoder section for both reference geometry compensation algorithms.

of the experimental validation as those speed profiles are still highly nonlinear. Fig. 15 shows the four speed profiles as calculated with Eq. 1 and the reference shaft encoder geometry. All shaft speed signals were sampled for a period of 10 seconds at a sampling rate of 102.4 kHz. For $\dot{\theta}_1$ and $\dot{\theta}_3$ the start of the acquisition was triggered when the shaft speed reached 200 RPM. This was accomplished with a built in tachometer in the OROS software. The start of the acquisition for $\dot{\theta}_2$ and $\dot{\theta}_4$ was triggered manually. The same two solving philosophies as in the numerical validation were used with the method. The calibration for each speed profile was again performed for a variety of different M values, the same as in the numerical experiment. The model hyperparameters β and σ were left unchanged from the numerical validation. The proposed method was used on all cases to obtain an estimate for the shaft encoder geometry.

3.2.4. Experimental results

Four different transient shaft speeds were used to calibrate the shaft encoder geometry using the proposed method. For illustrative purposes, the calibrated encoder geometry for profiles $\dot{\theta}_1$ and $\dot{\theta}_4$ is shown in Fig. 16 along with the reference geometry and its standard deviation values. These results are for philosophy 2 and M=55. It can be seen that the proposed method succeeds in calculating the encoder geometry to within the standard deviation of the reference geometry for all but one section ($\dot{\theta}_1$ for section 75). In general, the estimated encoder geometry corresponds to the reference encoder geometry. Fig. 17 shows the percentage errors of each section estimate for M=55 and both solution philosophies, the same information is conveyed in Fig. 18 on a log scale to better discern between the methods.

It can be seen in all cases that the application of the proposed method improves the accuracy of the encoder geometry considerably. From Fig. 18 it is observed that the estimated geometry for all speed profiles

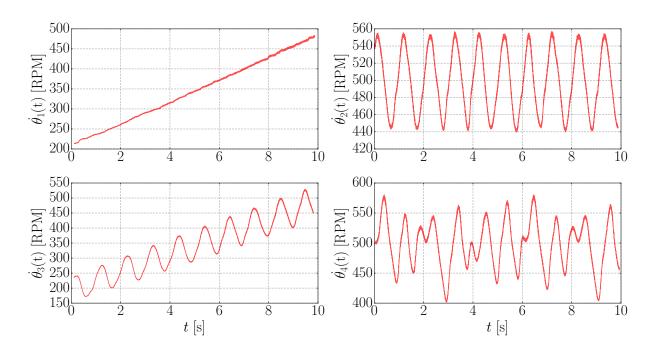


Figure 15: Four speed profiles used by the proposed method to calibrate the shaft encoder geometry.

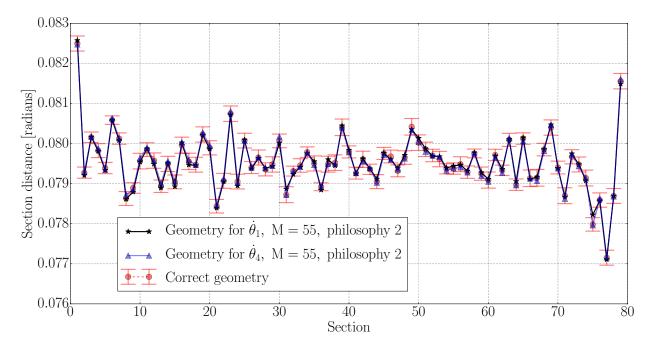


Figure 16: Estimated shaft encoder geometry as calculated from the first and fourth experimental shaft speeds using philosophy 2 and M=55.

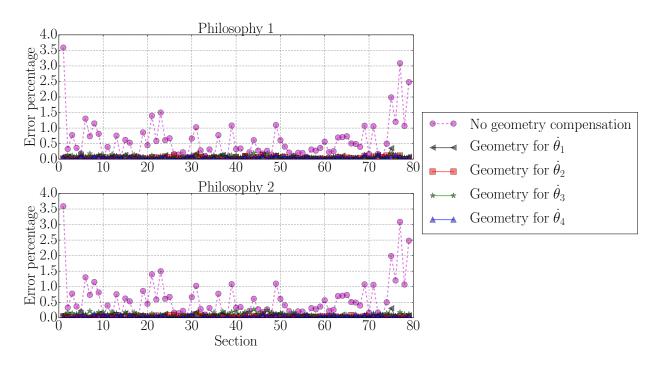


Figure 17: Section percentage errors for M=55 as calculated from all four experimental shaft speeds. Results shown on linear scale.

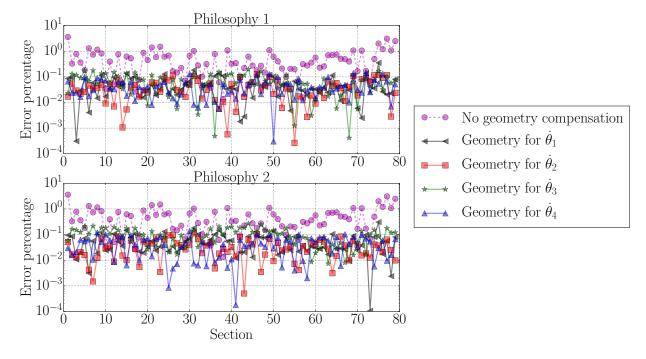


Figure 18: Section percentage errors for M=55 as calculated from all four experimental shaft speeds. Results shown on log scale.

are of similar accuracies. To investigate this further, Fig. 19 shows the RMS errors of all calibration runs.

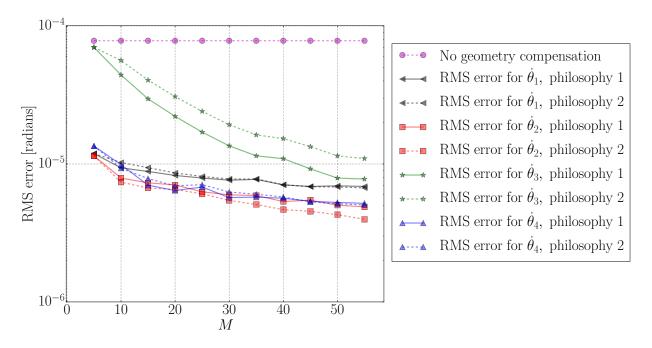


Figure 19: RMS error values for all four experimental speed profiles as a function of M.

It can be seen from Fig. 19 that the RMS error decreases as the number of revolutions increase for all speed profiles and both philosophies. This observation is consistent with the observations from the numerical validation. It is also observed that the geometry estimates for $\dot{\theta}_2$ and $\dot{\theta}_4$ are more accurate than for the other two profiles. The least accurate estimates occur with $\dot{\theta}_3$. Even so, the difference between all methods decrease as the number of revolutions increases. It is reasonable to assume that using more revolutions will yield even more accurate geometry estimates in all cases.

It is also observed that, in general, philosophy 2 does not yield significantly improved results over philosophy 1. From Fig. 19 it can be seen that the least accurate geometry estimate at M=55 is $\dot{\theta}_3$ using philosophy 2. The RMS error is 1.09×10^{-5} , which represents an increase in accuracy of 86% over an uncompensated-for geometry. The most accurate estimate represents an increase in accuracy of 95%. This validates the proposed method as being able to calculate the shaft encoder geometry. More accurate estimates are possible if more revolutions are used.

Finally, the times taken to solve the different calibration runs are shown in Fig. 20. It can be seen from Fig. 20 that the previously established rule of a doubling of solving time for each additional 5 revolutions holds.

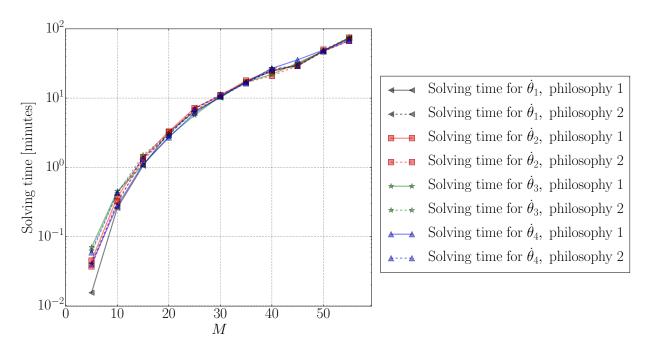


Figure 20: Solving times for experimental calibration runs.

3.3. Results discussion

The proposed method has been validated on four different transient speed profiles both numerically and experimentally. It has been demonstrated that the proposed method calibrates the shaft encoder geometry successfully and accurately. The least accurate estimate still poses a dramatic improvement over an uncompensated-for geometry. The proposed method is thus validated as the first published shaft encoder geometry compensation method that can be used on arbitrary shaft speeds and is able to determine every section distance of the encoder. A Python and Matlab implementation of the proposed method can be downloaded from http://caimlabs.co.za/index.php/other-links/software.

The proposed method addresses the problem of unequal encoder section distances in IAS determination. It must also be mentioned here that another important factor is addressed implicitly. The aliasing usually experienced by assuming a constant shaft speed between zero-crossing times is addressed by assuming the shaft speed as a second order polynomial between zero-crossing times. This undoubtedly mitigates the effect of aliasing if the proposed method is used to calculate the IAS. In future research a proper investigation must be undertaken to quantify the extent of the benefits obtained by this method. It should also be noted that it is entirely possible to use a higher order polynomial to express the shaft speed instead of a second order model. This should also be investigated in future.

3.4. Constant IAS calibration comparison

The goal of this article was to develop and demonstrate a method for performing shaft encoder geometry compensation that can be applied to signals acquired for any arbitrary shaft speed profile. This has been done and successfully demonstrated through numerical and laboratory experiments. The specific form of the IAS is therefore no longer a limiting factor when shaft encoder geometry compensation is required.

It is of interest to investigate briefly how the proposed method performs against current techniques where the shaft speed is constant. Since geometry compensation at constant shaft speeds fall outside the intended purpose of the proposed method, the investigation will be limited to a single simulated test.

For this numerical test, the rotor is simulated as running at 100 radians/second for 10 seconds. The sampling rate of the simulation and the true encoder geometry are the same as for the other numerical experiments.

The proposed method was then used to determine the true encoder geometry using philosophy 1 and M = 55. The geometry compensation algorithm as proposed by Resor et al. [6] is also used to determine the true encoder geometry. The results are shown in Fig. 21 a) and b).

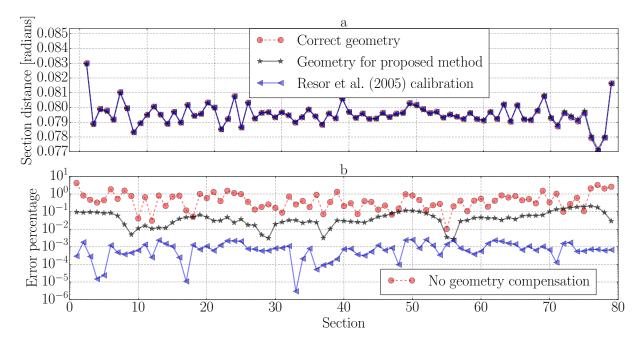


Figure 21: Shaft encoder geometry compensation for a constant IAS. A comparison between the propsed method and the method proposed by Resor et al. [6]. Figure a) shows the corrected geometry and figure b) shows the error percentages for each section.

From Fig. 21 a) it can be seen that both the proposed method and the method by Resor et al. [6] performs well to recover the true shaft encoder geometry. By looking at Fig. 21 b), the error values presented on a log scale reveal that the proposed method is the less accurate of the two.

It stands to reason that this should be the case. The proposed method does not assume any particular form of the IAS or the shaft encoder geometry, both of these quantities are treated as unknowns which must be solved. The solution shows that due to the flexibility of the IAS in the algorithm, the most probable solution is one in which the shaft encoder geometry is corrected for only until the IAS satisfied the conditions stated in Eq. 6 and Eq. 7. In short, the strength of the current method in being able to explicitly determine the IAS along with the encoder geometry turns out to be a weakness when the IAS is known beforehand as being constant. The method proposed by Resor et al. [6] assumes the IAS to be constant for each revolution and can therefore solve the shaft encoder geometry more accurately.

The simulated case reported above indicates that the proposed method should not be used if one is certain that the rotor IAS is constant. Further investigations have to be conducted to determine how much variety in rotor IAS is allowed before the proposed method will perform more accurately than the method proposed by Resor et al. [6]. This falls outside the scope of this article but should be investigated in future.

4. Conclusion

This article presents the derivation and validation of a novel technique to perform shaft encoder geometry compensation at arbitrary shaft speeds. The literature consulted indicated that there is a need for an online calibration procedure that can be applied to arbitrary shaft speeds for calibrating the entire encoder geometry. The proposed method satisfies these requirements. The method is derived and solved using Bayesian linear regression, and is validated on four transient shaft speeds, both numerically and experimentally. It is shown that the proposed method is capable of determining the encoder geometry for highly nonlinear shaft speeds to within a high accuracy. The proposed method increases in accuracy as more recorded revolutions are used for the calibration process.

Future research will focus on determining the ideal hyperparameters for each calibration run. Although the hyperparameters used in this article worked sufficiently well for all cases, increased accuracy might be possible if the choice of hyperparameters could be optimised. Additional future research can look into using the same approach as the proposed method for estimating the IAS once geometry compensation has been performed. This may lead to decreased effects of aliasing. The amount of deviation from a constant IAS that would render constant-speed calibration techniques less accurate than the proposed method should also

be determined in future research.

The proposed method is both easy to implement and a rigorous method of performing shaft encoder geometry compensation. The authors believe that it can be widely applied to many situations involving incremental shaft encoders and transient shaft speeds.

Acknowledgements

D. H. Diamond has been supported by a Skye Foundation postgraduate scholarship and the Eskom Power Plant Engineering Institute (EPPEI) during this research.

References

- [1] A. Y. B. Sasi, F. Gu, Y. Li, A. D. Ball, A validated model for the prediction of rotor bar failure in squirrel-cage motors using instantaneous angular speed, Mechanical Systems and Signal Processing 20 (2006) 1572–1589.
- [2] C. J. Stander, P. S. Heyns, Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions, Mechanical Systems and Signal Processing 19 (2005) 817–835.
- [3] J. Yang, L. Pu, Z. Wang, Y. Zhou, X. Yan, Fault Detection in a Diesel Engine By Analysing the Instantaneous Angular Speed, Mechanical Systems and Signal Processing 15 (2001) 549–564.
- [4] A. A. Gubran, J. K. Sinha, Shaft instantaneous angular speed for blade vibration in rotating machine, Mechanical Systems and Signal Processing 44 (2014) 47–59.
- [5] L. Renaudin, F. Bonnardot, O. Musy, J. Doray, D. Rémond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, Mechanical Systems and Signal Processing 24 (2010) 1998–2011.
- [6] B. R. Resor, M. W. Trethewey, K. P. Maynard, Compensation for encoder geometry and shaft speed variation in time interval torsional vibration measurement, Journal of sound and vibration 286 (2005) 897–920.
- [7] I. Alejandre, M. Artés, Method for the evaluation of optical encoders performance under vibration, Precision Engineering 31 (2007) 114–121.
- [8] H. André, F. Girardin, a. Bourdon, J. Antoni, D. Rémond, Precision of the IAS monitoring system based on the elapsed time method in the spectral domain, Mechanical Systems and Signal Processing 44 (2014) 14–30.
- [9] D. Rémond, Practical performances of high-speed measurement of gear transmission error or torsional vibrations with optical encoders, Measurement Science and Technology 9 (1998) 347–353.
- [10] A. Rivola, M. Troncossi, Zebra tape identification for the instantaneous angular speed computation and angular resampling of motorbike valve train measurements, Mechanical Systems and Signal Processing 44 (2014) 5–13.
- [11] Y. Li, F. Gu, G. Harris, A. Ball, N. Bennett, K. Travis, The measurement of instantaneous angular speed, Mechanical Systems and Signal Processing 19 (2005) 786–805.
- [12] M. El Badaoui, F. Bonnardot, Impact of the non-uniform angular sampling on mechanical signals, Mechanical Systems and Signal Processing 44 (2014) 199–210.
- [13] Q. Leclere, F. Girardin, D. Rémond, An analysis of Instantaneous Angular Speed measurement errors, in: Surveillance 7 International Conference, 2013, pp. 1–11.
- [14] K. Fyfe, E. Munck, Analysis of computed order tracking, Mechanical Systems and Signal Processing 11 (1997) 187–205.
- [15] Heidenhain, Rotary Encoders, 2011.
- [16] S. Braut, R. Zigulic, G. Stimac, A. Skoblar, Efficient approach for encoder geometry compensation in time interval

- measurement of torsional vibration, in: 10th International Conference on Vibrations in Rotating Machinery, 2012, pp. 1-10.
- [17] K. Janssens, P. Van Vlierberghe, W. Claes, B. Peeters, T. Martens, P. D'Hondt, Zebra tape butt joint detection and correction algorithm for rotating shafts with torsional vibrations, in: Proceedings of the ISMA2010, 2010, pp. 2869–2882.
- [18] C. Bishop, Pattern Recognition and Machine Learning, 1st ed., Springer, New York, 2009.