Approaches to nonlinear cointegration with a view towards applications in SHM

E J Cross and K Worden
Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.
e.j.cross@sheffield.ac.uk

Abstract. One of the major problems confronting the application of Structural Health Monitoring (SHM) to real structures is that of divorcing the effect of environmental changes from those imposed by damage. A recent development in this area is the import of the technique of cointegration from the field of econometrics. While cointegration is a mature technology within economics, its development has been largely concerned with linear time-series analysis and this places a severe constraint on its application - particularly in the new context of SHM where damage can often make a given structure nonlinear. The objective of the current paper is to introduce two possible approaches to nonlinear cointegration: the first is an optimisation-based method; the second is a variation of the established Johansen procedure based on the use of an augmented basis. Finally, the ideas of nonlinear cointegration will be explored through application to real SHM data from the benchmark project on the Z24 Highway Bridge.

1. Introduction
The field of econometrics is commonly concerned with the analysis of time series that exhibit nonstationarity and/or nonlinear behaviour. As the younger research field of structural health monitoring (SHM) also takes an interest in analysis of this nature, it is natural that SHM practitioners should look to econometrics for potentially useful analysis tools. Cointegration is one import from econometrics previously explored by the authors that appears to be very promising for use in SHM [1].

Cointegration is in fact a property of multivariate nonstationary time series, which if present, implies that some linear combination of the variables in question would be stationary. Econometricians are often interested in testing for cointegration; if two variables are cointegrated they share common trends and the stationary combination of them that one is able to find will define the long run equilibrium between them [2]. For econometricians attempting to define and understand relationships between economic variables, cointegration is also seen as a solution to the problem of spurious regression [3].

Stepping away from the world of economics towards the arguably more deterministic world of engineering, questions about the nature of time series and how variables interrelate may seem less pertinent. Instead, more interest may lie in the creation of a stationary variable from multiple nonstationary ones, i.e. in the removal of common trends from a multivariate data set. In fact, the ability to remove common trends from multivariate data sets becomes very desirable when considering the data normalisation problem in SHM. Often, in SHM, many of the features that are monitored for their sensitivity to damage are also very susceptible to any operational or environmental
conditions (e.g. temperature) that a structure is subject to. When a structure is monitored outside of laboratory conditions, often changes in monitored structural response caused by these innocent operational/environmental variations can mask any indication of damage. However, if a number of monitored variables are all affected by the same changing condition, (temperature say), cointegration could be used to remove this dependency and create a variable independent of environmental/operational influences that still maintains its sensitivity to damage.

For SHM then at least, the interest of cointegration lies in how to combine nonstationary variables to create a stationary result purged of the common trends in the original set. As one may imagine, a large body of literature is already available from econometrics on this very topic (see [4] for a good overview). One of the most common approaches used to find a stationary linear combination of multivariate nonstationary data is the Johansen procedure [5], which is a maximum likelihood based method for combining nonstationary variables whose first difference is stationary.

Despite this wealth of information available on nonstationary variable analysis, little of it is relevant to variables which exhibit nonlinear behaviour. This may be, in part, due to the fact that many econometric times series are considered to be sufficiently modelled by difference stationary models, that is to say models with a unit root, which linear cointegration is ideally suited to cope with. Nonlinear cointegration has been developed to analyse those time series that are considered to include nonlinear trends, or a group of variables which are believed to be nonlinearly related. Much of the research carried out into nonlinear cointegration by econometricians has been summarised in [6], where the focus is largely on identification of relationships between econometric variables.

After introducing cointegration theory in a slightly more rigorous way, three different approaches for coping with nonlinearity when using cointegration will be addressed in this paper, the last of which will be considered in the context of real SHM data collected from the Z24 Highway Bridge.

2. Linear Cointegration

To start, a formal definition of cointegration is necessary.

Definition: A set of nonstationary variables, say \( \{y_i\} \), are cointegrated if some linear combination \( z_i \) of them is stationary,

\[
z_i = \{\beta\}^T \{y_i\}
\]

If some \( \{\beta\}^T \) in (1) can be found such that \( z_i \) is stationary, then \( \{\beta\} \) is called the cointegrating vector.

Some constraints fall on the nonstationary variables in \( \{y_i\} \) if they are to be cointegrated; they must share common trends and they must also be ‘integrated to the same order’. A nonstationary time series \( y_i \) is integrated order \( d \), denoted \( y_i \sim I(d) \) if, after differencing the series \( d \) times it becomes stationary. In essence, each time series must have the same degree of nonstationarity.

The first step in any cointegration analysis is therefore to determine the order of integration of each variable in question. This will be achieved here with the Augmented Dickey Fuller (ADF) test [7, 8], which determines whether a variable will exhibit stationary or nonstationary behaviour by examining its characteristic roots when fitted to an error-correction model. This error-correction model takes the form

\[
\Delta y_i = \rho y_{i-1} + \sum_{j=1}^{p-1} b_j \Delta y_{i-j} + \epsilon_i
\]

where the difference operator \( \Delta \) is defined as \( \Delta y_{i-j} = y_{i-j} - y_{i-j-1} \). A suitable number of lags \( p \) should be included in this model to insure that \( \epsilon_i \) becomes a white noise process [4]. In this form, the characteristic roots of the process in question, and therefore the stationarity of the process, is determined by the value of \( \rho \). If \( \rho = 0 \), the process has a unit root and will be nonstationary. The ADF
test therefore evaluates whether \( \rho \) is statistically close enough to zero to deem the time series nonstationary. In much the same way one would conduct a t-test, the test statistic
\[
t_p = \frac{\hat{\rho}}{\sigma_\rho} \tag{3}
\]
where \( \hat{\rho} \) is the least squares estimate of \( \rho \), and \( \sigma_\rho \) the variance of the parameter, should be compared to critical values from ADF statistic tables (see [9]); an hypothesis of nonstationarity is rejected at level \( \alpha \) if \( t_p < t_\alpha \). If the hypothesis is accepted, the time series has a unit root and is \( I(1) \). If the hypothesis is rejected, the test should be repeated for \( \Delta y_i \), if the hypothesis is then accepted \( y_i \) is an \( I(2) \) nonstationary sequence. This can be continued until the integrated order of the time series is ascertained.

Upon ascertaining the order of integration of each of the variables under consideration, any that are of the same order are eligible for cointegration analysis. As touched upon on the introduction, this is done using the Johansen procedure, which is designed to work with \( I(1) \) variables. The premise of the Johansen procedure is to use a maximum likelihood approach to estimate the parameters of a Vector Error-Correction Model (VECM) of the variables under consideration. A VECM takes the form
\[
\{\Delta y_i\} = [\Pi]\{y_{i-1}\} + \sum_{j=1}^{p-1}[B_j]\{\Delta y_{i-j}\} + [\phi]\{D(t)\} + \{\varepsilon_i\} \tag{4}
\]
where \( \{y_i\} \) denotes an n-vector including all \( n \) variables to be analysed, with the subscript \( i \) relating to time, \( i = 1, \ldots, N \), \( p \) represents the model order, or the number of lags to be included in the model, and \( \{\varepsilon_i\} \) is a normally distributed noise process; \( \{\varepsilon_i\} \sim N(0, [\Sigma]) \). A term to describe a deterministic trend \( \{D(t)\} \) has also been included. In this form, the matrix \( [\Pi] \) contains the combinations that will create the most stationary residuals, i.e. \( [\Pi] \) contains the cointegrating vectors. Unfortunately, if the variables are cointegrated, \( [\Pi] \) must be of reduced rank, which makes estimates of the parameters in (4) harder to obtain - a reduced rank regression must be used. Details of this reduced rank regression are not included here as they are in depth and complicated. Instead, for more details readers should refer to [1, 5].

3. Nonlinear Cointegration

The above theory has been developed for linear time series, and will not always be compatible where nonlinearities are concerned. Situations that could cause problems are if variables include nonlinear deterministic trends or are nonlinearly related to each other in some other way. In these cases, the way in which the variables are combined has to be modified in order to still be able to create a stationary residual.

One situation that may commonly arise is where two (or more) different variables from the same system exhibit nonlinear dependencies on some external disturbance, such as temperature fluctuation, for example. To demonstrate this idea theoretically, suppose there are two different variables \( \{x_i, y_i\} \) from the same system, one which reacts linearly with respect to some external disturbance, \( t \), and one which reacts nonlinearly, in a quadratic way, say, to that same external disturbance. Suppose these variables take the form
\[
x_i = \alpha t_i + \varepsilon_i \tag{8}
\]
\[
y_i = \beta t_i^2 + \varepsilon_i \tag{9}
\]
where \( \alpha, \beta \) are constants, \( t_i \) is some deterministic trend caused by the external disturbance and \( \varepsilon_i \) are random normally distributed processes.

It is clear that a linear combination of \( x \) and \( y \) could not result in a stationary sequence. However, some combination of \( y \) and the square of \( x \) should produce a comparatively stationary signal:
\[
z_i = a_1 x_i^2 + a_2 y_i \tag{10}
\]
If the parameters $a_1$ and $a_2$ can be found so that $z_i$ is stationary, the vector $[a_1 \ a_2]$ will be analogous to the linear cointegrating vector, and $x$ and $y$ will be nonlinearly cointegrated. It is then a matter of finding the parameters $a_1$ and $a_2$.

One way to do this would be to compute $x_i^2$ and to include it as variable in its own right in the Johansen procedure, i.e. the input to the Johansen procedure would be $\{x_i^2, y_i\}$. In this way, the only difference in the approach to the Johansen procedure is a manipulation of the form of the variables that are linearly combined.

A different approach again is to treat (10) as an optimisation problem, where the aim is to choose parameters $[a_1 \ a_2]$ such that $z_i$ is as stationary as possible. For this purpose, the authors suggest using a nonlinear optimisation routine based on differential evolution. The following subsection will briefly describe differential evolution but readers are referred to [10] and [11] for more details. Section 3.2 will follow with results using techniques on data simulated to represent the theoretical situation above.

3.1. Differential Evolution
Differential evolution, first introduced by Storn and Price in 1997 [11], is an evolutionary algorithm that begins with an initial population of trial solutions to some problem and reaches an optimal set of solutions through successive cycles of mutation, crossover and selection. The suitability of trial solutions are determined by some objective function, set according to the individual problem in hand. For this application, the trial solutions take the form of a vector of parameters guesses $[a_1 \ a_2]$ that satisfy (10) with $z_i$ stationary.

The optimisation routine is summarised in Figure 1. To begin with an initial population of parameter vectors are randomly generated. To each parameter vector in the initial population, a cost value is specified according to the objective function chosen. A new generation of solutions is created from this initial population as follows. Firstly, a target vector is chosen from the initial population. Next, a trial vector is created by ‘mutation’; from the initial population, two parameter vectors are randomly chosen (A and B in figure 1), their difference (A-B) is multiplied by some scaling factor, to which finally a third randomly chosen parameter vector (C) from the initial population is added. The resultant is called the mutated trial vector.

![Figure 1. A schematic of differential evolution](image-url)
A new parameter vector is now created through ‘cross-over’ of the mutated trial vector and the target vector. Cross-over creates a new vector by choosing individual elements from the mutated trial vector and the target vector by a series of binomial experiments (see [10] for details). This newly created vector will then be selected for the next generation if its cost value is lower than that of the target vector. If it is higher, the target vector will be placed in the next generation population. The process is repeated for each vector in the initial population. As the process evolves through the generations the population will eventually become full of suitable parameter vectors with low cost values.

For the purposes of nonlinear cointegration, a suitable objective function must be chosen on the basis of the stationarity of the cointegrated signal (such as (10)). Several options are available, the simplest being to choose the objective function to minimise the variance of the cointegrated signal. Another suitable option would be to use the ADF statistic described above, which has a larger negative value the more stationary the time series is. Both cost functions will be trialled for the theoretical example above, and the results will be compared in the next section along with the results of using an altered basis for the Johansen procedure.

3.2. Results

The nonlinear cointegration procedures suggested here are trialled for combining time series of type (8) and (9). Namely, differential evolution using the two different cost functions discussed and a slightly modified version of the Johansen procedure are used to choose the parameters in (10) that produce the most stationary combination of (8) and (9).

The results are shown in Figure 2. Figure 2(a) and (b) show the results using differential evolution with the variance-based cost function and the ADF statistic-based cost function respectively. For this particular trial a scaling factor of 0.9 and a crossover ratio of 0.5 were used in the differential evolution step, for more details on the choice of such parameters readers are referred to [10, 11]. Figure 2(c) shows the results when using \( \{x_1^2, y_1\} \) as a basis for the Johansen procedure.

On inspection, Figure 2 shows that a stationary residual has successfully been found for combination (10) by all three methods. In all cases, the methods are providing successful nonlinear cointegrating vectors for the times series.

An interesting property and a possible drawback, however, of the kinds of combinations used to cointegrate these nonlinear trends (10), is that on closer inspection, the variance of each of the combined signals is increasing with time, although each cointegrated signal is mean stationary. This growing variance is small in Figure 2(a) where variance was used a cost function, it grows at a faster rate, however, in figures 2(b) and (c) where the ADF statistic and the Johansen procedure were used. To understand this one expands (10):

\[
z_i = a_1 x_i^2 + a_2 y_i
= a_1 (a^2 t^2_i + t_i \varepsilon_i + \varepsilon_i^2) + a_2 (\beta t^2_i + \varepsilon_i)
\]

(11)

The differential evolution, and indeed the Johansen procedure, will choose the parameters \( a_1 \) and \( a_2 \) so that the quadratic deterministic trends cancel each other out. The remaining terms will include randomly distributed noise but also a term depending on \( t_i \varepsilon_i \). This term is responsible for the increasing variance. Although the residuals created shown in figure 2 are stationary in comparison to the original signals, with the initial trends removed, they are not strictly stationary on closer inspection. This issue may not necessarily, however, be such a problem in practice if the trends are much stronger sources of variation than ambient noise. Principled means of removing terms like the one discussed above are under consideration at the moment and will also be the subject of further work.

Three different ways to approach nonlinearity and cointegration have been explored above using theoretical examples. The following section looks at a real world example where nonlinear cointegration is relevant. This example comes from the benchmark SHM study carried out on the Z24 Highway Bridge in Switzerland before its demolition. A brief introduction to this benchmark study is given to begin with.
Figure 2. Combination of signals with linear and quadratic deterministic trends, combination found using (a) differential evolution with a variance based cost function, (b) differential evolution with an ADF statistic-based cost function, (c) the Johansen procedure with modified basis.
4. Nonlinear Cointegration and the Z24 Bridge

The Z-24, a pre-stressed concrete highway bridge in Switzerland, was subject to a comprehensive monitoring campaign under the ‘SIMCES project’ [12], prior to its demolition in the late 1990s, it has since become a benchmark study in SHM. The monitoring campaign, which spanned a whole year, tracked modal parameters and included extensive measurement of the environmental factors affecting the structure, such as air temperature, soil temperature, humidity etc.

The Z24 monitoring exercise has been an important study in the history of SHM developments because towards the end of the monitoring campaign researchers were able to introduce a number of damage scenarios to the structure. To simulate a settlement of the pier foundation one of the support piers was lowered in stages by up to 95mm, which resulted in cracks in the bridge girder above the pier. The data set collected is therefore very useful for analysis as it includes data collected from a ‘normal’ and a damaged condition.

Of interest here are the natural frequencies of the bridge which were tracked over the period of a year and additionally over the period where the support pier was lowered. Modal properties of the bridge were extracted from acceleration data [13]. Figure 3 shows a time history of the four natural frequencies between 0-12Hz of the bridge. The solid vertical line marks the start of the period where different levels of damage (pier lowering) were introduced. Gaps where the monitoring system failed have been removed.

On inspection of Figure 3, the natural frequencies of the bridge are by no means stationary. There are some large fluctuations in the first half of the time history before the introduction of any damage. These fluctuations occurred during periods of very cold temperatures and have been associated with an increase in stiffness caused by freezing of the asphalt layer on the bridge deck. The natural frequency time histories are, therefore, a good illustrative example of damage sensitive parameters also sensitive to environmental variations, in this case temperature.

![Figure 3](image)

**Figure 3.** Time Histories of the extracted natural frequencies of the Z-24 Bridge, monitored over one year including a period when damage was introduced.

As the natural frequencies in their current form would not be suitable to monitor as a damage sensitive feature some action must be taken to remove the variable set’s sensitivity to temperature. If each variable (natural frequency) is linearly related to temperature, cointegration is an ideal tool to remove the temperature induced trends. In this case, however, the modal properties of the bridge are
First Natural frequency nonlinearly dependent on temperature (as an example, Figure 4 plots how the first natural frequency changes with temperature), which means that the Z24 provides an excellent example with which to explore the ideas of nonlinear cointegration.

The first sensible step when exploring the ideas of cointegration in this nonlinear context is to look at the results of using the linear Johansen procedure. Figure 5 shows the four natural frequencies of the Z24 bridge projected onto the ‘best’ cointegrating vector found by the Johansen procedure, when trained on data points 1-500. In this figure, the dotted horizontal lines indicate confidence intervals at 3σ of the residual from the training period, while the vertical solid line indicates the beginning of the period where damage was introduced. The cointegrating vector has successfully de-trended the data set and furthermore some indication of damage is visible towards the end of the data. It is perhaps unexpected that the linear Johansen procedure is able to remove nonlinear trends; looking at the relationships between the natural frequencies and how the Johansen procedure has combined them, however, sheds light on how this has been achieved. Although each frequency is related nonlinearly to temperature, which drives the frequency fluctuations, some of the frequencies (although not all) are linearly related to each other, which means that the Johansen procedure can successfully combine them to remove their common trends. The way in which the natural frequencies from the Z24 relate to each other is noted in Table 1. On studying the cointegrating vector in question, the residual in Figure 5 predominantly results from a combination of the first and third frequencies, which effectively removes the temperature dependent trends, the combination only contains very small contributions from the second and fourth frequencies. Although this is a successful removal of the temperature dependent trends, two of the variables have effectively not been included in the analysis. This is not an ideal situation as the loss of two variables reduces the chances of being able to successfully detect damage.

**Figure 4.** First Natural frequency nonlinearly dependent on temperature.

**Table 1.** Relationships between modal frequencies of the Z24 Bridge.

| Frequency | 1   | 2        | 3        | 4   |
|-----------|-----|----------|----------|-----|
| 1         |     | Nonlinear| Linear   | Linear |
| 2         | Nonlinear | -        | Nonlinear| Nonlinear |
| 3         | Linear | Nonlinear | -        | Linear |
| 4         | Linear | Nonlinear | Linear   | -    |
One way around the problem of nonlinearity in this circumstance could be to use ‘locally linear models’. Although overall the Z24 modal frequencies vary nonlinearly with temperature, each of them displays a linear relationship with temperatures above zero degrees, in other words there is a ‘locally’ linear relationship between temperature and each modal frequency. By only looking at data from the locale of the linear regime, cointegration principles once again become valid. To illustrate this, the Johansen procedure was re-implemented on the four Z24 variables, this time with all data points removed if they occurred at temperatures below 1°C. The first residual produced by the Johansen procedure for this modified data set was almost identical to that shown in Figure 5, in other words the Johansen procedure suggested the same combination to purge the temperature trends. As this particular combination is not very sensitive to damage a better candidate may come from the other cointegrating vectors calculated by the Johansen procedure (for n variables, the Johansen procedure will produce n-1 cointegrating vectors). Figure 6 shows the residual of the second cointegrating vector found by the Johansen procedure for the modified data set; it has successfully been purged of temperature dependence and also becomes nonstationary after the introduction of damage.

Despite the fact that looking at locally linear models has created a successful candidate for a feature sensitive to damage and insensitive to normal temperature variations, there is no escaping the fact that any new data points from very low temperatures projected onto the cointegrating vector would have to be discarded.

Ultimately what is needed is a way to combine the entire data set and purge the nonlinear trends entirely. Although pursued no further in this paper, the authors plan to adapt the theoretical ideas from Section three so that they can be applied to real data, where relationships between the variables are more complex and initially unknown. The authors’ current aim is to use an optimisation procedure to choose the necessary variable combinations to involve in the linear superposition. This is left to the next publication.

5. Conclusions

The idea of nonlinear cointegration has been introduced as potentially useful tool for SHM. A simple approach to nonlinear cointegration has been introduced whereby a straightforward multinomial combination of signals is chosen to remove any deterministic trends. The nonlinear optimisation routine of differential evolution, and also a variation of the Johansen procedure, have been suggested.
Data points occurring above 1°C combined using the second cointegrating vector from the Johansen procedure. Additionally the ideas of nonlinear cointegration have been explored in the context of real SHM data from the Z24 Bridge project.

Acknowledgements
The authors would like to express their gratitude to the researchers on the project SIMCES who conducted the benchmark study on the Z24 Bridge and also made their data freely available. Thanks should also go to Ville Lämsä for providing the authors with the data.

6. References

[1] Cross E J, Worden, K., Chen, Q. 2011 Cointegration; A novel approach for the removal of environmental trends in structural health monitoring data Preprint: submitted to Proceedings of the Royal Society - Series A
[2] Chen Q, Kruger U and Leung A 2009 Industrial & Engineering Chemistry Research 48(7) 3533-3543
[3] Yule G 1926 Journal of the royal statistical society 89(1) 1-63
[4] Perman R 1993 Journal of Economic Studies 18(3)
[5] Johansen S 1995 Likelihood-based inference in cointegrated vector autoregressive models (Oxford: Oxford University Press)
[6] Dufrénot G and Mignon V 2002 Recent developments in nonlinear cointegration with applications to macroeconomics and finance (Dordrecht: Kluwer Academic Publishers)
[7] Dickey D and Fuller W 1979 Journal of the American statistical association 427-431
[8] Dickey D and Fuller W 1981 Econometrica: Journal of the Econometric Society 49(4) 1057-1072
[9] Fuller W 1996 Introduction to statistical time series (Wiley-Interscience)
[10] Worden K, Manson G, Sohn H and Farrar C 2005 Proceedings of 23rd International Modal Analysis Conference
[11] Storn R and Price K 1997 Journal of global optimization 11(4) 341-359
[12] De Roeck G 2003 Journal of Structural Control 10(2) 127-134
[13] Peeters B and De Roeck G 2001 Earthquake Engineering & Structural Dynamics 30(2) 149-171