ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Sound and Vibration 271 (2004) 365–390

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

Instability of systems with a frictional point contact. Part 1: basic modelling

P. Duffour, J. Woodhouse*

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

Received 16 July 2002; accepted 24 February 2003

Abstract

The linear stability is investigated of systems which contain a sliding frictional contact at a single point. A condition for instability is found, in terms of the transfer functions of the two systems at the point of contact. This condition is explored for generic systems, to establish the circumstances under which instabilities might be expected. A major conclusion is that if the coefficient of friction is assumed to be constant, then at least one mode of one or other of the contacting systems must have a displacement at the contact with a particular pattern of signs. If such a mode exists then instability is possible, depending on the value of the coefficient of friction and on the frequencies and mode shapes of the other modes of the system. Stability boundaries are shown to be extremely sensitive to distribution of damping in the system, suggesting that damping might be one of the causes of the typical "capricious" behaviour of friction-instability experiments. Systems consisting of three modes are studied in detail. This is shown to be an important case since much of the behaviour of a system consisting of many modes can be understood by breaking it down into clusters of three. In a subsequent paper, some of the assumptions made here will be relaxed so as to catalogue systematically all the possible routes to instability within linear theory.

1. Introduction

Noise and vibration arising in mechanical systems containing sliding frictional contacts is commonplace. Usually it is undesirable, as in squeaking of hinges or squeal of vehicle disc or drum brakes. There is significant technological interest in understanding the phenomenon, or phenomena, well enough that one could design a brake, for example, which never

*Corresponding author. Fax: +44-1223-332662.
E-mail address: jw12@eng.cam.ac.uk (J. Woodhouse).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2003.02.002

squealed. However, despite research over many years which has produced a considerable literature, this goal has yet to be achieved. Research can be divided into two categories. First is experimental work, either on simplified laboratory systems (e.g. Refs. [1,2]) or on real disc or drum brakes [3–7]. The second category is theoretical modelling (e.g. Refs. [8–10]). A variety of physical mechanisms, and of different styles of analysis, are presented. What they all have in common is that each model applies rather specifically to a particular system, and it is not clear how one might generalize the conclusions to other systems.

However, generalization seems to be what is called for by the experimental evidence. The most striking feature of the results, agreed on by all authors whether studying real brakes or idealised pin-on-disc systems, is what might be termed the 'capriciousness' of the phenomenon of self-excited vibration in such a system. The rig will be switched on and might immediately make a noise with a given frequency. If it is turned off and on again an hour later, perhaps it will be silent, or produce a noise at a different frequency. This capriciousness is often discouraging for the experimentalist, and it is tempting to blame this feature on the numerous parameters potentially influencing the phenomenon (temperature, humidity, normal load, geometrical details, contacting materials and so on). However, a more constructive interpretation is that the capriciousness is surely indicating something fundamental about the nature of the phenomenon. Within very broad limits, *any* system embodying a sliding frictional contact seems to be prone to self-excited vibration and when it happens it generally exhibits this capriciousness. Squeal, or whatever word may be used to describe the phenomenon, is a 'problem waiting to happen', always on the brink of occurring, and often with several different types of squeal available to be 'chosen' by the system. This apparently generic behaviour cannot be investigated in a very illuminating way by analyzing specific systems, because the details obscure any possible generality. The aim of this paper is to present a unified analysis of a more general class of systems. Some conclusions can be drawn from this analysis about which characteristics of a system may make it prone to squeal, and about the origin of capriciousness. For definiteness this account will be presented in terms of noise in disc braking systems, although the theory to be presented applies equally well to self-excited vibration in a variety of other systems containing a sliding contact.

The systems to be studied, although general in some respects, are still very particular in other respects. The aim is to study a class of systems which is as simple as possible, while retaining an essential aspect of generality. Two important restrictions apply:

1. this is a linearized stability analysis,
2. contact only occurs at a single point.

The mathematical result underlying this approach is the so-called "centre manifold theorem" [11]. Only squeal events which can be initiated via linear instability of the state of steady sliding are considered. There may well be some types of squeal which rely on intrinsically non-linear phenomena such as parametric resonance, without the necessity of instability when the underlying system is linearized. Equally, squeal may be initiated by a linear instability, but it may change its character significantly as it grows to large amplitude. These are all issues worthy of investigation. However, a clear understanding of the linear problem is surely a prerequisite for any further study.

2. Governing equations

The system to be studied is shown diagrammatically in Fig. 1. One linear system, the 'disc', is being driven in uniform motion. Another linear system, the 'brake', is stationary apart from vibration, and makes contact with the first system through a single point. All motion is assumed to take place in the plane of the diagram, which is also the plane of the imposed frictional sliding. At the contact point, there is a pair of equal and opposite normal reactions N acting on the two systems, and similarly a pair of equal and opposite friction forces F. These forces are both composed of a steady component (N_0 or F_0) plus a fluctuating component N' or F', which will be assumed to be small:

$$N = N_0 + N', \quad |N'| \ll |N_0|,$$
$$F = F_0 + F', \quad |F'| \ll |F_0|. \tag{1}$$

Denote the normal and tangential displacements of the disc by u_1 and v_1, respectively, and the normal and tangential displacements of the brake by u_2 and v_2, respectively. The tangential displacement of the disc needed here is that due to the vibration alone, disregarding the contribution from steady rotation. All fluctuating quantities will be considered in the frequency domain, as Fourier transforms of the associated time-varying quantities. The Fourier frequency variable is denoted ω. The goal of this analysis is to study the initiation of self-excited vibration from a state of steady sliding and all motion will be assumed to be sufficiently small for linear theory to be used.

The displacements of each system are related to the normal and tangential forces via matrices of transfer functions, defined by

$$\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} G_{11}(\omega) & G_{12}(\omega) \\ G_{21}(\omega) & G_{22}(\omega) \end{bmatrix} \begin{bmatrix} N' \\ F' \end{bmatrix}, \quad \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{bmatrix} \begin{bmatrix} N' \\ F' \end{bmatrix}. \tag{2}$$

The matrices are both symmetric from the standard principle of reciprocity (e.g. Ref. [12]). In practice, violation of symmetry could arise from the rotation of the disc, but this rotation will be assumed to be sufficiently slow for the effect to be neglected. To obtain a closed system of equations, two further relations are required. First, it is assumed that the brake and disc remain in contact, but some input to the linear system is allowed via surface roughness of the disc. This

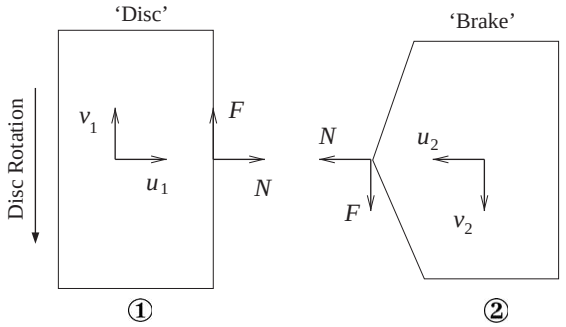


Fig. 1. Diagram showing two linear systems in sliding contact at a single point. Displacements u_i and v_i apply at the contact point but are shown shifted apart for clarity.

means that the two normal displacements are equal and opposite except for the effect of roughness, defined by a function (again in the frequency domain) r which is assumed known. Then

$$u_2 = r - u_1. \tag{3}$$

Finally, a frictional constitutive law is needed. In this paper, the simplest possible assumption is made whereby the friction force F is proportional to the normal force N with a constant coefficient of friction μ_0:

$$F = \mu_0 N. \tag{4}$$

In a companion paper [1], the effect of a more complicated friction law will be investigated. The sign convention in Fig. 1 has been chosen so that a positive value of μ_0 is expected.

The set of equations including the simple friction law (4) can be straightforwardly rearranged to give the solution:

$$N' = \frac{r}{G_{11} + \mu_0 G_{12} + H_{11} + \mu_0 H_{12}} \tag{5}$$

and

$$\begin{aligned} u_1 &= (G_{11} + \mu_0 G_{12})N', \\ v_1 &= (G_{12} + \mu_0 G_{22})N', \\ u_2 &= (H_{11} + \mu_0 H_{12})N', \\ v_2 &= (H_{12} + \mu_0 H_{22})N'. \end{aligned} \tag{6}$$

Since all the individual transfer functions G_{ij}, H_{ij} are of stable systems, all poles are in the upper half of the complex Fourier plane.

It follows immediately that this system can be unstable if and only if the function

$$D(\omega) = G_{11} + \mu_0 G_{12} + H_{11} + \mu_0 H_{12} \tag{7}$$

has at least one zero in the lower Fourier half-plane.

So far, ω has been used as a complex variable, as opposed to the Laplace variable s, more familiar in control. The domain of stability for the Laplace variable s is the left hand side complex plane, so that $t \mapsto e^{st}$ remains bounded. For $t \mapsto e^{i\omega t}$ to remain bounded, the Fourier variable ω must be in the *upper half complex plane*.

The statement above is the key theoretical result of this paper. Much of what follows will be based on exploring this equation from different points of view. First, this condition for instability will be analyzed in general terms, as some useful information can indeed be gained from a purely formal inspection. Then, the criterion will be tested by simulations of increasing complexity.

3. Some general observations

If the first subsystem is indeed a model of a brake disc, or a brake drum, then to a good approximation there is a plane of symmetry through the contact point and the centre of the disc/drum. It follows immediately that

$$G_{12} = 0, \tag{8}$$

since tangential forces and displacements are anti-symmetric with respect to this plane, while normal forces and displacements are symmetric so that they cannot interact. If this is indeed the case, then the function $D(\omega)$ reduces to

$$D(\omega) = G_{11} + H_{11} + \mu_0 H_{12}. \tag{9}$$

Henceforth, G_{12} will be assumed to be zero. Eq. (9) shows that, within the linear theory, the coefficient of friction only influences the stability via the dynamic cross-coupling of the 'brake'. If the 'brake' subsystem is also symmetrical so that $H_{12} = 0$, then the coefficient of friction no longer enters the formulation. D is then further reduced to $D(\omega) = G_{11} + H_{11}$, which simply expresses the passive normal coupling of the two subsystem. One expects such a system to be unconditionally stable. This is confirmed if the transfer functions are expressed in terms of modal parameters, which will be carried out next.

The transfer function G_{11} is a driving-point receptance, which can be expressed in terms of the mode shapes ϕ_n, their natural frequencies ω_n^d and their modal damping factors δ_n^d using the standard formula

$$G_{11}(\omega) = \sum_n \frac{\phi_n^2(x)}{(\omega_n^{d2} + 2\mathrm{i}\omega\omega_n^d\delta_n^d - \omega^2)}, \tag{10}$$

where $\phi_n(x)$ denotes the value of the (mass-normalized) mode shape at the position of the contact point, in the normal direction. Proportional damping has been assumed here, for simplicity. The effect of non-proportional damping will be discussed in the companion paper [1].

In a very similar way, the transfer functions H_{11} and H_{12} can be expressed in terms of the brake mode shapes ψ_n, their natural frequencies ω_n^b and their damping factors δ_n^b:

$$H_{11}(\omega) = \sum_n \frac{\psi_n^2(x)}{(\omega_n^{b2} + 2\mathrm{i}\omega\omega_n^b\delta_n^b - \omega^2)}, \quad H_{12}(\omega) = \sum_n \frac{\psi_n(x)\psi_n(y)}{(\omega_n^{b2} + 2\mathrm{i}\omega\omega_n^b\delta_n^b - \omega^2)}, \tag{11}$$

where $\psi_n(x)$ denotes the value of the (mass-normalized) mode shape at the position of the contact point, in the normal direction, and $\psi_n(y)$ denotes the corresponding mode shape in the tangential direction. Thus

$$D(\omega) = \sum_n \frac{\phi_n^2(x)}{(\omega_n^{d2} + 2\mathrm{i}\omega\omega_n^d\delta_n^d - \omega^2)} + \sum_n \frac{\psi_n(x)[\psi_n(x) + \mu_0\psi_n(y)]}{(\omega_n^{b2} + 2\mathrm{i}\omega\omega_n^b\delta_n^b - \omega^2)}. \tag{12}$$

The function $D(\omega)$ has the functional form of a transfer function, although it is not the transfer function of any obvious physical system. However, the zeros of D may, in some sense, be thought of as the complex eigenvalues of the coupled system. The fact that D has the form of a transfer function means that its characteristics can be deduced from standard arguments about the distribution of peaks and anti-resonances [12]. The frequencies ω_n must be appropriately interleaved to give the full set of peaks in this function. Between an adjacent pair of peaks, there will either be a shallow dip or a sharp anti-resonance. If the coefficients of the resonant terms in the expansion (12) have the same sign, an anti-resonance generally occurs, while if they have opposite signs, a shallow dip occurs. This distinction is important for the present investigation: to predict the threshold of instability, one is interested in the zeros of D and particularly in zeros which are very close to the real ω-axis, either just above (and thus stable) or just below (and thus

unstable). Such zeros, which are near the stability threshold, will occur close to anti-resonance frequencies of D.

To investigate the formal properties of $D(\omega)$, it is useful to consider the resonant terms in expansion (12) as belonging to a single pool of "modal blocks", regardless of which subsystem they come from. It is also convenient to write the combinations of mode shape coefficients appearing at the numerators in Eq. (12) as single coefficients a_i. The function $D(\omega)$ can then be written as

$$D(\omega) = \frac{a_1}{(\omega_1^2 + 2\mathrm{i}\omega_1\delta_1\omega - \omega^2)} + \frac{a_2}{(\omega_2^2 + 2\mathrm{i}\omega_2\delta_2\omega - \omega^2)} + \cdots. \tag{13}$$

This form will later be used as the basis for simulations, because it reduces slightly the large number of parameters involved. For convenience, the coefficients a_i in Eq. (13) will be referred to as "modal amplitudes", although they may not represent the amplitude of any physical variable of the system.

In Eq. (13), $D(\omega)$ appears as the sum of rational fractions of degree -2. Putting these fractions to the same denominator, D can be written as a single ratio of two polynomials, say P and Q, so that $D(\omega) = P(\omega)/Q(\omega)$. The roots of $Q(\omega)$, which are the poles of $D(\omega)$, are the roots of the quadratic denominators appearing in expansion (12). For small damping $\delta_i \ll 1$, these poles have a simple approximate expression: $\omega_i(1 + \mathrm{i}\delta_i)$ and $\omega_i(-1 + \mathrm{i}\delta_i)$. The zeros of $D(\omega)$ are the roots of $P(\omega)$. P has a number of noteworthy properties:

1. If D consists of N_{tot} resonant terms in total, the degree of P will be $2(N_{tot} - 1)$. Therefore P will have $2(N_{tot} - 1)$ roots and D has as many zeros.
2. $P(\mathrm{i}\omega)$ has real coefficients. Therefore the roots of $D(\omega)$ are either purely imaginary or appear as "$-$i" times a complex conjugate pair. In geometrical terms, this means that the pattern of zeros is symmetrical with respect to the imaginary axis in the complex ω-plane. Given the expression for the poles of D given above, it is clear that these poles will have the same symmetry property.
3. The leading coefficient of $P(\omega)$ is always the sum of the a_i. When this sum equals zero, the degree of P drops by one and so does the number of roots. Therefore, one can expect the condition $\sum a_i = 0$ to appear as a remarkable event. The odd power coefficients of P are a linear combination of damping factors, therefore P only possesses even-power coefficients when the system is undamped.
4. The root-coefficient relationships ensure that if the coefficients of $P_s(s) = P(-\mathrm{i}\omega)$ have different signs, the system is unstable (see e.g. Ref. [13]).

Property (4) is clearly of interest for this study, as it links the modal properties of the two subsystems, to the stability of the coupled system. This property can be refined by using complex analysis theory. In principle, the number of unstable zeros of $D(\omega)$ could be obtained using the Nyquist criterion, familiar from control theory (see e.g. Ref. [13]). This would amount to defining a contour consisting of a large portion of the real axis, closed by a semi-circle in the lower half complex plane (unstable region). A typical situation is represented diagrammatically in Fig. 2, where the crosses represent the poles and the circles, the zeros of $D(\omega)$.

This plot shows two unstable zeros in the lower half-plane, encircled by the contour labelled Γ_0. Nyquist's criterion states (among other things) that the number of zeros inside the semi-circular

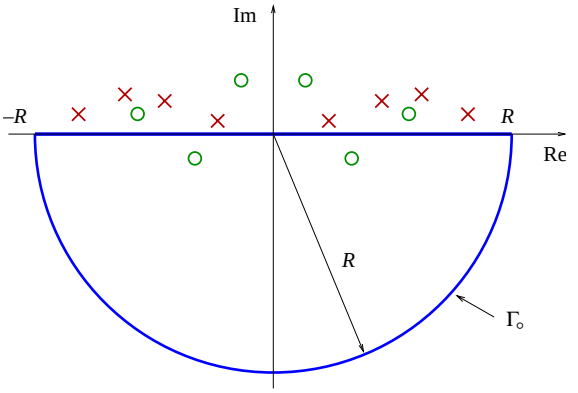


Fig. 2. Plot showing diagrammatically a typical contour Γ_0, in the Fourier complex plane. The crosses represent the poles of D, while the circles represents the zeros. The poles are always in the upper half-plane. The zeros can be in the upper (stable) or lower (unstable) half complex plane. The patterns of zeros and poles are both symmetrical with respect to the imaginary axis. The two unstable zeros are encircled by the semi-circular contour of radius R, labelled Γ_0.

contour is equal to the number of times the curve obtained by plotting $D(\omega)$ as ω moves along Γ_0, encircles the origin.

In practice, counting these encirclements when D contains many modes can be difficult, because the image curve $D(\Gamma_0)$ usually follows a very intricate path. However, the criterion can be used to prove an interesting result: the function $D(\omega)$ cannot have any zero in the lower half complex plane if all the "modal amplitude" coefficients a_i are positive. This statement can be proved without difficulty by noticing that if the a_i are all positive, then, as ω moves along the path Γ_0, the image path $D(\Gamma_0)$ cannot take *real negative* values. Therefore, it cannot encircle the origin. This points towards the importance of the negative compound coefficients "a_i". From expansion (12), it appears that only those compound coefficients coming from the brake can be negative. Their expression is then

$$a_i = \psi_i^2(x) + \mu_0 \psi_i(x)\psi_i(y). \tag{14}$$

Therefore, for instability to occur, it is necessary that the brake possesses modes such that $\psi_i(x)\psi_i(y) < 0$. If there are no such modes, then this theory predicts that the system cannot be unstable when the disc rotates in the direction shown in Fig. 1. However, reversing the rotation of the disc amounts to changing the sign of the coefficient of friction in D, so that those modes of the brake such that $\psi_i(x)\psi_i(y) > 0$ can now induce instability. Disregarding whether this is practically feasible or not, Eq. (14) also shows that any brake mode such that $\psi_i(x)\psi_i(y) < 0$ could result in the corresponding compound a_i being negative providing the coefficient of friction μ_0 is sufficiently large.

This completes the general observations that could be made on the properties of criterion (7). In the next section, a number of approximate systems will be investigated using this formalism. These analyses will now be local: a system will approximated by a number of neighbouring modes. Most of the general comments made in this section will re-emerge, often as elementary mathematical properties of the particular system under investigation.

4. Approximate analysis of generic systems

The previous sections point towards the study of the zeros of transfer functions. It is more customary to study poles of transfer functions. Coming from a different perspective, Lyon and co-workers [14–16] have made some useful observations on the location of transfer function zeros and their relation to the phase of the frequency response. The results presented in this section can be thought of as an extension of Lyon's work. The approach adopted here is based on the assumption that however complicated the function $D(\omega)$ for a particular system, one can usually expect to obtain a reasonable approximation in the vicinity of a given frequency by considering only the nearby resonances. Therefore, this approach is only expected to give meaningful results where this assumption is valid, that is, in the vicinity of a given particular frequency, for instance that of a squeal event. In this section, various approximations of increasing complexity will be investigated. The simplest useful approximation which can lead to a prediction about zeros of $D(\omega)$ is to consider just two terms, and neglect all others. This case is so simple that it can be dealt with in some detail. The influence of additional terms from distant resonances will then be considered.

4.1. Two-mode approximation

The two-mode approximation can be useful if the system squeals at a frequency in the vicinity of which there are two relatively isolated modes of the uncoupled subsystems. In this context, the function D will have the form

$$D(\omega) \approx \frac{a_1}{(\omega_1^2 + 2\mathrm{i}\omega_1\delta_1\omega - \omega^2)} + \frac{a_2}{(\omega_2^2 + 2\mathrm{i}\omega_2\delta_2\omega - \omega^2)}. \tag{15}$$

Suppose, for example, that the first term comes from mode n of the disc, while the second comes from mode m of the brake. Then using the notation defined in the previous section,

$$a_1 = \phi_n^2(x), \quad \omega_1 = \omega_n^d, \quad \delta_1 = \delta_n^d, \tag{16}$$

and

$$a_2 = \psi_m^2(x) + \mu_0\psi_m(x)\psi_m(y), \quad \omega_2 = \omega_m^b, \quad \delta_2 = \delta_m^b.$$

4.1.1. Positive-frequency pole approximation

As well as neglecting all but two terms in Eq. (13), further simplification can be obtained by factorizing the denominator expressions, which are quadratic in ω, expressing each term as a sum of two partial fractions and retaining only the one with a resonance at a positive value of $\mathrm{Re}(\omega)$. The second term, with a negative value of $\mathrm{Re}(\omega)$, is typically more distant than the other neglected resonances of the system. This leads to the approximation

$$D(\omega) \approx \frac{c_1}{(\omega - \bar{\omega}_1)} + \frac{c_2}{(\omega - \bar{\omega}_2)}, \tag{17}$$

where

$$\bar{\omega}_k \approx \omega_k(1 + \mathrm{i}\delta_k) \quad \text{and} \quad c_k = -\frac{a_k}{2\omega_k} \quad (k = 1 \text{ or } 2) \tag{18}$$

are, respectively, the positive-frequency poles of the modes retained, and their corresponding residues. By convention, whenever the term "residue" is used in an unspecified way in what follows, it will always be understood as the residue associated with a positive-frequency pole.

The one complex zero $\omega = \omega_z$ resulting from this approximation is simply

$$\omega_z \approx \frac{(c_1\bar{\omega}_2 + c_2\bar{\omega}_1)}{(c_1 + c_2)}. \tag{19}$$

From Eq. (19), it is clear that, whatever the values of c_1 and c_2, ω_z always lies on the straight line passing through $\bar{\omega}_1$ and $\bar{\omega}_2$ in the complex plane. The slope of this line depends on the ratio of damping factors of the two modes (this slope depends on their natural frequencies as well, but these are assumed to be close). If the poles have similar imaginary parts, the pole line will be almost parallel to the real axis and will only cross it far away from the poles. In that case, the present approximation is not expected to hold since the influence of other poles may no longer be negligible. On the other hand, if the two damping factors are very different, the line will have a large slope and will cross the real axis to produce potential instability not too far from the poles. This suggests that if two neighbouring modes have very different damping factors, then instability is more likely to arise in the vicinity of these two modes.

If c_1 and c_2 both have the same sign, then ω_z is simply a weighted average of the two complex poles and therefore lies on the line segment joining them. Since the two poles are stable, they both lie in the upper half complex plane and so do all the points on the segment joining them. Thus, this combination cannot lead to an unstable zero.

More interesting is the case when c_1 and c_2 have opposite signs. Note that given the sign reversal between the residues c_i and the modal amplitudes a_i (see Eq. (18)), the residues of the disc can only be negative, while those of the brake can be positive or negative (provided the mode shapes are real, as implied by the assumption of proportional damping).

Two cases may be distinguished. Suppose first that $c_1 < 0, c_2 > 0$ and $c_1 + c_2 < 0$. Then

$$\omega_z = \bar{\omega}_2 + \alpha(\bar{\omega}_2 - \bar{\omega}_1) \quad \text{with } \alpha = -\frac{c_2}{c_1 + c_2}. \tag{20}$$

The zero occurs on the opposite side of $\bar{\omega}_2$ from $\bar{\omega}_1$, and if α is reasonably small it will be close to $\bar{\omega}_2$. It may, of course, be influenced by the next resonance on that side of $\bar{\omega}_2$, but it is supposed that $\bar{\omega}_1$ and $\bar{\omega}_2$ are close together compared to the interval to the next resonance, so that this possibility can be temporarily ignored. This zero can be unstable if

$$-\alpha\,\mathrm{Im}(\bar{\omega}_1) + (1 + \alpha)\,\mathrm{Im}(\bar{\omega}_2) < 0 \tag{21}$$

or, using Eq. (16)

$$\delta_m^b < \left(\frac{\alpha}{1 + \alpha}\right)\left(\frac{\omega_n^d}{\omega_m^b}\right)\delta_n^d. \tag{22}$$

Instability can occur if the damping of the brake mode is sufficiently *small*. Fig. 3 shows a plot of the function $D(\omega)$ for a typical example of the behaviour just discussed, with parameter values producing an unstable zero.

The second case occurs when $c_1 < 0$, $c_2 > 0$, $c_1 + c_2 > 0$ and yields very similar results so only the main points are given here: this time the zero occurs on the opposite side of $\bar{\omega}_1$ from $\bar{\omega}_2$ and it can

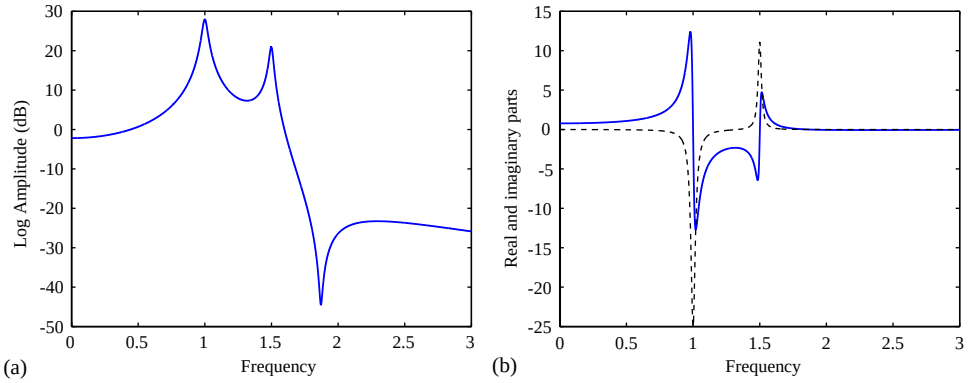


Fig. 3. Plot of D showing a zero (antiresonance) beyond the two fixed poles at frequencies 1 and 1.5 rad/s. The damping factors were 0.02 and 0.01, respectively, and the residues -1 and 0.5 in the same order. (a) Magnitude plot of D (in dB). (b) Plot of the real (solid line) and imaginary (dashed line) parts of D (linear scale).

be unstable if

$$\delta_m^b > \left(\frac{1+\beta}{\beta}\right)\left(\frac{\omega_n^d}{\omega_m^b}\right)\delta_n^d, \quad \text{with } \beta = -\frac{c_1}{(c_1+c_2)} > 0. \tag{23}$$

In this case, instability can occur if the damping of the brake mode is sufficiently *large*.

This two-pole analysis suggests that

1. a prerequisite for the occurrence of instability is the presence of consecutive modes having the amplitudes 'a_i' of opposite signs;
2. the stability of the system is significantly influenced by the damping. More precisely, stability is more likely when the two damping factors are very different.

4.1.2. Influence of the negative-frequency poles

To verify how much these conclusions are specifically linked to the "positive-pole" approximation, the stability of the system is now studied with the negative poles added back in. To compare the two approximations, it is more convenient to use the modal amplitudes a_i, rather than residues c_i. The function D for this system was expressed in Eq. (15). Its zeros are the roots of the following quadratic equation:

$$(a_1+a_2)(\mathrm{i}\omega)^2 + 2(a_1\omega_2\delta_2 + a_2\omega_1\delta_1)(\mathrm{i}\omega) + a_1\omega_2^2 + a_2\omega_1^2 = 0. \tag{24}$$

Written as a polynomial in $s = \mathrm{i}\omega$, it is possible to use the Routh–Hurwitz criterion [13] to determine the stability of the roots of Eq. (24). For a quadratic, the set of conditions for the system to be stable is simply that all the coefficients should have the same sign, either positive or negative:

$$\begin{array}{lll} (1) & a_1+a_2>0 & \\ (2) & a_1\omega_2\delta_2 + a_2\omega_1\delta_1 > 0 & \text{or} \\ (3) & a_1\omega_2^2 + a_2\omega_1^2 > 0 & \end{array} \qquad \begin{array}{lll} (1) & a_1+a_2<0, & \\ (2) & a_1\omega_2\delta_2 + a_2\omega_1\delta_1 < 0, & \\ (3) & a_1\omega_2^2 + a_2\omega_1^2 < 0. & \end{array} \tag{25}$$

From these two sets of inequalities, it is clear that the system is always stable if the two modal amplitudes have the same sign. This conclusion confirms a general result from Section 3. From the definitions of a_1 and a_2 given in Eq. (16), a_1 is always positive (disc 'mode'), and only a_2 can be positive or negative (brake 'mode'). Since it is the only interesting case, a_2 will be assumed to be negative (i.e. $-a_2 > 0$). In order to analyze the interaction of the two modes in a somewhat systematic way, one mode, say mode 1, has its characteristics kept fixed, while mode 2 will be varied in frequency and amplitude. This will also set the scene for the subsequent approximations, where this procedure will be used extensively.

If a_2 and ω_2 are varied, the two sets of conditions (25) can be represented as areas in the (ω_2, a_2)-plane. These areas of stability are delimited by three curves whose equations are given by setting each condition to zero. The first condition is simply a horizontal line at $-a_2 = a_1$, cutting the plane into two parts. Condition (2) also defines two regions, delimited by the straight line $-a_2 = a_1(\delta_2/(\delta_1\omega_1))\omega_2$, while the third condition defines two areas delimited by a parabola of equation $-a_2 = a_1\omega_2^2/\omega_1^2$. The domain of stability is the intersection of all these areas. Note that only condition (2) involves the damping factors. This means that conditions (1) and (3) remain unchanged whatever the system damping.

Fig. 4 shows the stability domain as a shaded area for values of the damping factors such that $\delta_2/\delta_1 = 3/2$. For this plot, Mode 1 has both its frequency and modal amplitude set to 1. The damping factor is 0.01. Mode 2 is varied in natural frequency from 0 to 2, and in modal amplitude from -5 to 0. This plot is particularly useful to understand how the damping effects the stability regions. As already noted, the damping factors are only involved in condition (2), through their ratio. This ratio governs the slope of the dashed straight line. If they are identical, the three curves meet at a single point (ω_1, a_1), so that the line resulting from condition (2) does not modify the regions of stability defined by conditions (1) and (3) alone. The stability domain is then the same

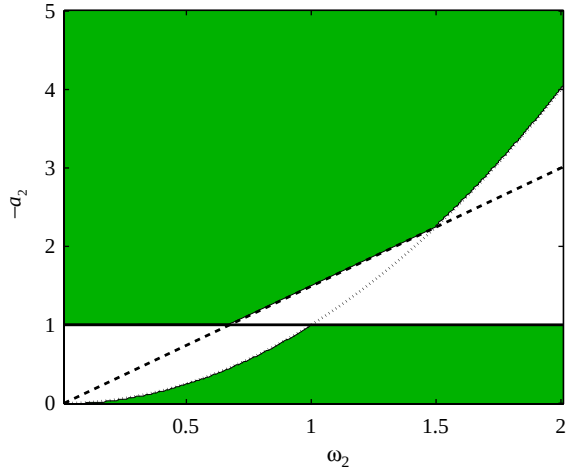


Fig. 4. Plot showing the domain of stability of a two-mode system. The properties of Mode 1 are fixed at frequency 1, modal amplitude 1 and damping factor 0.01. The natural frequency of mode 2 is varied from 0 to 2, while its modal amplitude is varied from -5 to 0. Its damping factor is $\delta_2 = 0.015$. The shaded areas are the zones where the system is stable. The horizontal line is $-a_2 = a_1 = 1$, the dashed line is $-a_2 = a_1(\delta_2/(\delta_1\omega_1))\omega_2$, while the dotted line is $-a_2 = a_1\omega_2^2/\omega_1^2$.

as that of the undamped system. This feature will recur many times in the subsequent sections: if all the modes have the same damping factor, then the stability region is the same as in the undamped case. When the damping factors are different, the line defined by condition (2) encroaches either on the upper "undamped" stability region if $\delta_2/\delta_1 > 1$ (Fig. 4), or on the lower "undamped" region, if $\delta_2/\delta_1 < 1$. The more different the damping factors, the more the stability region is reduced. This behaviour was also observed in the two-pole analysis. In contrast to the latter, the present analysis predicts the existence of instability, even if the system is undamped (the single zero predicted by the two-pole analysis is always marginally stable if the system is undamped). Subsequent analysis will throw some light on this particular point.

It may be objected at this point that the range of variation of the second mode frequency may be too large for a local analysis. If a normalized frequency equal to 1 represents 1 kHz, then varying the second frequency between 0 and 2 means varying it between 0 and 2 kHz! A real system would probably have many other modes within this range, which makes the assumption that the two frequencies are relatively isolated very implausible. This objection is probably fair. The reason for choosing such a wide range is that it gives a good understanding of how plots showing thresholds of stability (Fig. 4) are constructed. These plots will recur throughout this work. Most of the features described here will still be present, but the complexity of the algebra will prevent the clear analytical understanding reached with this simple case.

4.2. Two poles plus a constant residual

The conclusions of the two-mode analysis are likely to be affected by the presence of other poles. With more than two poles, analytical investigation is difficult because the algebra becomes rapidly unwieldy, and with six or more poles it is in principle impossible in most cases, because it requires the roots of a polynomial of degree 5 or higher. However some further progress is possible by supposing that around a frequency of interest, the function D can be appropriately approximated by keeping the two nearest poles and assuming that the contribution from more remote poles can be equated to a *constant* residual. Further analytical progress can be achieved by ignoring the contribution from the negative-frequency poles.

With the notation introduced in the previous sections, the approximate expression for D then becomes

$$D(\omega) \approx \frac{c_1}{(\omega - \bar{\omega}_1)} + \frac{c_2}{(\omega - \bar{\omega}_2)} + R, \tag{26}$$

where R is the contribution from remote poles. Denote by $\bar{\omega}_3 = \omega_3(1 + \mathrm{i}\delta_3)$ one of these remote poles. The exact contribution of this pole to D would be a term proportional to $1/(\omega - \bar{\omega}_3)$. As ω moves further away from $\bar{\omega}_3$, the imaginary part of $1/(\omega - \bar{\omega}_3)$ decays like $1/(\omega - \omega_3)^2$, whereas its real part decays like $1/(\omega - \omega_3)$, so that in the vicinity of $\bar{\omega}_1$ or $\bar{\omega}_2$, the real part of $1/(\omega - \bar{\omega}_3)$ will dominate its imaginary part. Therefore, R will be assumed to be real.

If D given by Eq. (26) is rearranged as a single rational fraction, its numerator is a quadratic in ω, whose roots are

$$\omega_{z\pm} = \bar{\omega} - \frac{c_1 + c_2}{2R} \pm \sqrt{\left(\frac{c_1 + c_2}{2R}\right)^2 + \bar{A}^2 + \frac{c_1 - c_2}{R}\bar{A}}, \tag{27}$$

where

$$\bar{\omega} = \frac{\bar{\omega}_1 + \bar{\omega}_2}{2} \quad \text{and} \quad \bar{\Delta} = \frac{\bar{\omega}_2 - \bar{\omega}_1}{2}.$$

Eq. (27) shows that $\omega_{z\pm}$ can only have a negative imaginary part when the square root possesses a negative imaginary part large enough to overcome that of $\bar{\omega}$. In order to carry this analysis further, the system will be assumed to be undamped. Then, all the quantities become real and it is possible to study the sign of the expression under the square root. The influence of damping will be discussed at the end of this section.

4.2.1. The undamped case

With no damping, Eq. (27) can be conveniently rewritten as

$$\omega_{z\pm} = \omega - \frac{c_1 + c_2}{2R} \pm \sqrt{\left(\Delta + \frac{c_1 - c_2}{2R}\right)^2 + \frac{c_1 c_2}{R^2}}, \tag{28}$$

where ω and Δ are the (real) values of $\bar{\omega}$ and $\bar{\Delta}$ when the damping is set to zero. One of these zeros has a negative imaginary part if and only if the expression under the square root is negative. It is immediately clear that this can only occur if $c_1 c_2 < 0$. Assume first that $c_1 < 0$ and $c_2 > 0$. Then

$$\left(\Delta + \frac{c_1 - c_2}{2R}\right)^2 + \frac{c_1 c_2}{R^2} = \Delta^2 \left(1 - \frac{(\sqrt{-c_1} - \sqrt{c_2})^2}{2R\Delta}\right)\left(1 - \frac{(\sqrt{-c_1} + \sqrt{c_2})^2}{2R\Delta}\right).$$

If $R\Delta < 0$, both brackets are positive, so that the zeros cannot be complex. If $R\Delta > 0$, the product of the two brackets is negative whenever $R\Delta$ satisfies

$$\frac{(\sqrt{-c_1} - \sqrt{c_2})^2}{2} \leqslant R\Delta \leqslant \frac{(\sqrt{-c_1} + \sqrt{c_2})^2}{2}. \tag{29}$$

Similarly, when $c_1 > 0$ and $c_2 < 0$, instability can only arise if $R\Delta < 0$ and satisfies

$$-\frac{(\sqrt{c_1} + \sqrt{-c_2})^2}{2} \leqslant R\Delta \leqslant -\frac{(\sqrt{c_1} - \sqrt{-c_2})^2}{2}. \tag{30}$$

In the $(c_1, c_2, R\Delta)$ space, these inequalities represent a volume bounded by two surfaces of identical shape and symmetrical about the origin. Fig. 5 shows one of these two surfaces, confined in one octant.

Note that the apex of this "cone-like" surface is at the origin, and the surface touches the plane $R\Delta = 0$ along the line $c_1 = -c_2$. This means that in principle, for some values of the ratio c_1/c_2 (in particular when, $c_1/c_2 = -1$, the only case studied by Lyon), any non-zero value of R, however small, can produce instability. This suggests that even very remote modes could tip a zero into the unstable region. It is also interesting to note that only the difference between the frequencies matters for the stability, not their individual values.

Fig. 6 shows how the zeros move in the complex plane when ω_1 and ω_2 are fixed at 1 and 1.2, respectively (i.e., $\Delta = 0.2$) and the value of R is gradually increased from nearly zero to 50. The residue values are kept constant at $c_1 = -1$ and $c_2 = 2$. Increasing R in these conditions corresponds to moving up along the black vertical line shown in Fig. 5.

It is reassuring that as R decreases to zero, one of the roots tends towards minus infinity along the real axis, while the other root tends towards the value of the single zero given by Eq. (19) in

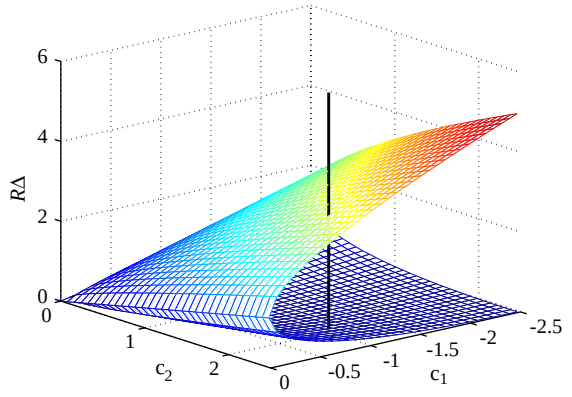


Fig. 5. Surface plot showing half of the stability threshold surface for a system made of two modes plus a constant residual. The system is unstable inside the cone-like shape. The vertical line passing through marks the section explored in Fig. 6.

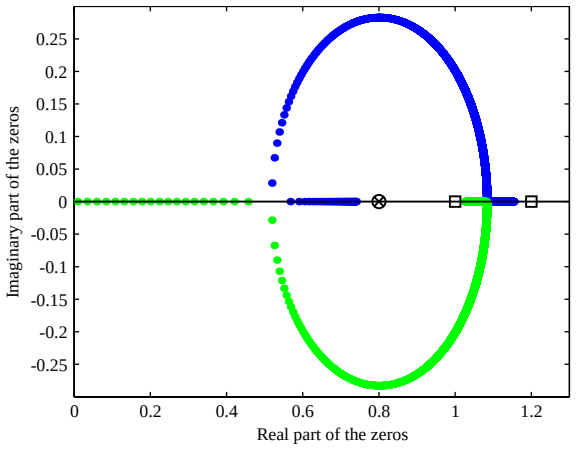


Fig. 6. Root locus plot showing how the zeros move when the value of R is varied from nearly zero to 50. The squares (□) are the two fixed poles, while the crossed circle (⊗) is the position of the zero when $R = 0$. For this run $c_1 = -1$ and $c_2 = 2$.

the previous section. As R increases and passes 0.85 in the run shown, the roots meet and split to become a complex conjugate pair. One is in the upper half plane (stable) and the other in the lower half (unstable). Each root goes symmetrically around an oval shape and they merge again on the real axis (when $R \approx 30$) between the two poles. They subsequently remain real, and each converges towards a different fixed pole. Provided the parameter values are such that the surface of Fig. 5 is crossed as R varies, such behaviour is always observed. Without damping, either the roots are real and the system is marginally stable or they are complex and the system is unstable since one of the roots will always be in the lower half complex plane. Therefore, the merging/splitting of the roots represents bifurcation points regarding stability. A Taylor expansion of the expressions of the zeros given by Eq. (28), for very large or very small values of R, shows that the roots always tend towards the limits described, whether the surface is crossed or not.

4.2.2. The influence of damping

When damping is added, it becomes very difficult to obtain analytical results for the stability thresholds, but an extensive range of simulations has shown that the pattern outlined above remains recognizable as long as the damping is light (i.e. δ_i of the order of a few percent). In Fig. 7, the damping factors have been set to $\delta_1 = 0.03$ and $\delta_2 = 0.02$, everything else being unchanged.

To a good approximation, it can be shown (using Eq. (28)) that this plot corresponds to the undamped plot 'shifted up' by $\mathrm{Im}(\bar{\omega})$. $\mathrm{Im}(\bar{\omega})$ being always positive, this suggests that the damped system has become stable when the undamped one was only marginally stable. The main difference between the damped and undamped case is that the two mergings of the roots in the undamped case have been replaced by two kinds of 'veerings'. The distortion of the merging/splittings into veerings means that there is no longer a direct correspondence between those events and stability thresholds. However, the oval shape is still recognizably present, so that the description made of the undamped case still holds good and one can expect the surface plot in Fig. 5 to provide a good first approximation of the system stability.

Since these mergings can be thought of as bifurcation points through which the system becomes unstable, they merit further examination. A suitable way of characterizing this veering would be to determine the length and orientation of the minimal distance between the two curves. From Eq. (27), these properties are given by the minimum modulus and the corresponding argument of the complex number:

$$\omega_{z+} - \omega_{z-} = 2\sqrt{\left(\bar{\Delta} + \frac{c_1 - c_2}{2R}\right)^2 + \frac{c_1 c_2}{R^2}}. \tag{31}$$

If the system is unstable, the modulus of this complex quantity will be minimized for two values of R. The corresponding values of the quantity "$\omega_{z+} - \omega_{z-}$" will be denoted σ_1 and σ_2, respectively. A simple analysis shows that the orientation of the veerings is, in most cases, governed by the sign

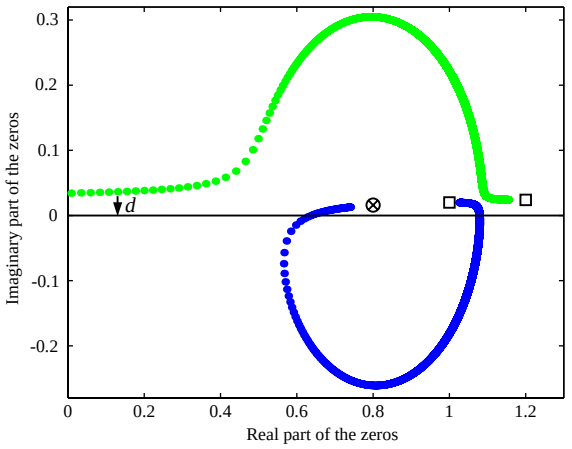


Fig. 7. Root locus plot showing how the zeros move when the value of R is varied from nearly zero to 50 with non-zero damping. As before, the squares (□) are the two fixed poles, while the crossed circle (⊗) is the position of the zero when $R = 0$. For this run $c_1 = -1$ and $c_2 = 2$ and the damping is 0.02 for the two poles.

of a parameter δ, defined by

$$\delta = \frac{(\omega_2\delta_2 - \omega_1\delta_1)}{(\omega_2 - \omega_1)}.$$

If $\delta > 0$, $Arg(\sigma_1) \approx \pi/4$ and $Arg(\sigma_2) \approx 3\pi/4$, whereas if $\delta < 0$, $Arg(\sigma_1) \approx 3\pi/4$ and $Arg(\sigma_2) \approx \pi/4$.

In the case shown in Fig. 7, $\delta > 0$ and the inclination of σ_1 is approximately $\pi/4$ as expected. Modifying slightly the parameter values so that δ becomes negative has the unexpected result of swapping the role of the roots so that the bottom root plotted in Fig. 7 is flipped to the top and vice versa. Therefore, a slight variations of the damping factors can induce drastic rearrangements of the roots in the complex plane.

4.2.3. Conclusions

In this section, a detailed study of the roots of a system made of two modes and a constant residual showed that

- Without damping, the system is either marginally stable or unstable. The emergence of instability is equivalent to a splitting of two previously real roots into a complex conjugate pair.
- With damping, broadly speaking, the system becomes stable when it was previously marginally stable. To a good approximation, the damped system is unstable when it was already unstable without damping, although there is no longer an exact correspondence between splitting of the roots and the emergence of instability.
- The effect of damping was shown to be subtle. Small variations of the damping factors can result in drastic rearrangements of the roots in the complex plane. This point confirms previous observations, made for the two-mode systems.

This system can be thought of as a useful intermediate case between the two-mode system studied in Section 4.1 and the next one, where the influence of a third resonant term in D is studied.

4.3. Stability of a three-mode system

In this section, the stability of a system consisting of three modes is investigated. As in Section 4.1, it would be possible to ignore the negative-frequency poles at first, then investigate their influence. However, the positive-frequency pole approximation is not very useful, because little can be done analytically. Therefore, the influence of a third pole will mainly be studied using numerical simulations of the complete three-mode system.

The rationale for the sequence of numerically calculated cases to be considered is as follows. As already mentioned, the 'composite transfer function' $D(\omega)$ will have, for most of its resonant terms, amplitudes of positive sign, corresponding to the driving-point terms in Eq. (12). It was shown in Section 3, that for the system to be unstable, at least one mode with a negative 'amplitude' must be introduced. Therefore, the system investigated in this section will have two modes with positive amplitudes, say modes 1 and 2, and a third mode, labelled 3, with a negative amplitude. Following the procedure adopted in Section 4.1.2, the two modes with positive amplitudes will have fixed frequencies and amplitudes, while these two modal properties will be varied for mode 3. Recall that according to expansion (12), the expression for the amplitude of

mode 3 is $\psi^2(x) + \mu_0\psi(x)\psi(y)$, so that varying it can also be thought of as *varying the coefficient of friction* μ_0.

Using the same notation as before, the expression for $D(\omega)$ is

$$D(\omega) \approx \frac{a_1}{(\omega_1^2 + 2i\omega_1\delta_1\omega - \omega^2)} + \frac{a_2}{(\omega_2^2 + 2i\omega_2\delta_2\omega - \omega^2)} + \frac{a_3}{(\omega_3^2 + 2i\omega_3\delta_3\omega - \omega^2)}. \tag{32}$$

First, the system is considered without damping.

4.3.1. The undamped case

This simplification makes it possible to obtain a glimpse of analytical insight. The zeros of $D(\omega)$ are the roots of the biquadratic equation:

$$\begin{aligned}(a_1 + a_2 + a_3)(i\omega)^4 + [a_1(\omega_2^2 + \omega_3^2) + a_2(\omega_1^2 + \omega_3^2) + a_3(\omega_1^2 + \omega_2^2)](i\omega)^2 \\ + a_1\omega_2^2\omega_3^2 + a_2\omega_1^2\omega_3^2 + a_3\omega_1^2\omega_2^2 = 0.\end{aligned} \tag{33}$$

In principle, it is possible to express the roots of this equation, but this is not very illuminating. In order to link this section with previous results, it is more interesting to apply a similar analysis using the Routh–Hurwitz criterion. Inspection of the possible stability cases for a biquadratic equation shows its roots are in the stable region if and only if

(i) all the non-zero coefficients have the same sign, and
(ii) the discriminant of the equation, regarded as a quadratic in $(i\omega)^2$, is positive.

As in Section 4.1, this leads to a set of inequalities: one for each coefficient and an additional one for the discriminant. Each of these inequalities can be represented graphically in the (ω_3, a_3) plane thus defining the stability domain of the system. The curves of equation:

$$\begin{aligned}&(1) \quad -a_3 = a_1 + a_2, \\ &(2) \quad -a_3 = [a_1(\omega_2^2 + \omega_3^2) + a_2(\omega_1^2 + \omega_3^2)]/(\omega_1^2 + \omega_2^2), \\ &(3) \quad -a_3 = (a_1\omega_2^2\omega_3^2 + a_2\omega_1^2\omega_3^2)/\omega_1^2\omega_2^2\end{aligned} \tag{34}$$

can be expected to be salient features of this plot. Viewed as functions of ω_3, condition (1) is a straight horizontal line, while (2) and (3) are parabolas. The fourth condition, obtained by setting the discriminant to zero, is of a different nature: it is an implicit equation involving the square of a_3 and the fourth power of ω_3. Although this equation is not of a standard type, it can be factorised, so as to express a_3 in terms of the other parameters. This expression is cumbersome and not very enlightening in itself, except that it contains the following square root:

$$\sqrt{-4a_1a_2(\omega_1^2 - \omega_3^2)(\omega_2^2 - \omega_3^2)}.$$

If a_1 and a_2 have the same sign, so that their product is positive, the square root only exists if ω_3 is between the other two frequencies. Setting the discriminant to zero therefore yields a stability threshold curve, which will be confined in between the other two mode frequencies in the (ω_3, a_3) plane. Simulations show that this curve has broadly the shape of an ellipse. Further inspection of this equation reveals that it has some interesting geometrical properties which are summarized in Fig. 8.

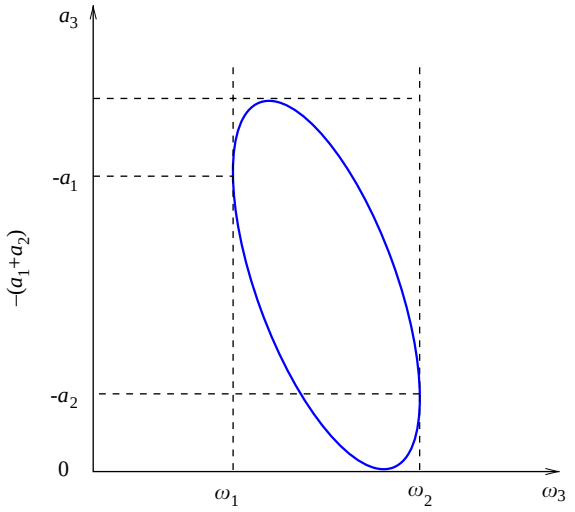


Fig. 8. Diagram summarizing some geometrical properties of the zero contour plot.

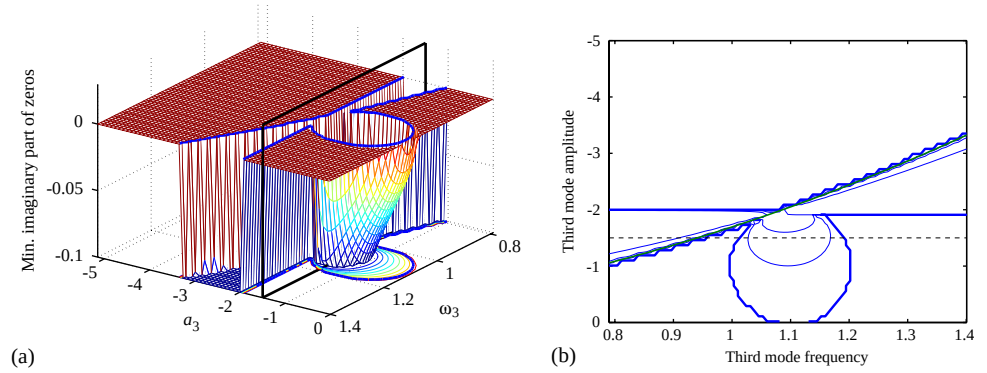


Fig. 9. (a) Surface plot showing the minimum imaginary part of the zeros of an undamped three-mode system where two are kept fixed whereas the third one is varied in amplitude and frequency. The zero contour line (stability threshold) is shown on the surface and on the base plane. (b) Contour plot of the surface plot shown in (a). The thick line is the zero contour. Thinner lines are the contours for values -0.1, -0.2, -0.5. Note that for better visibility, the surface plot has been rotated, so that the origin is the bottom right corner.

The behaviour of this system will now be simulated. Mode 1 has a frequency 1 and amplitude 1. Mode 2 has a frequency 1.2 and amplitude 1. Mode 3 has its amplitude varied from -5 to 0 and its frequency will range from 0.8 to 1.4. As in the two-mode section, the frequencies have been normalised. The range of variation of the third mode frequency spans a symmetrical interval around the fixed pole frequencies. As explained in the introduction to this section, only the poles around a given frequency of interest are assumed to have an influence on the stability.

In order to condense the maximum information into a single plot, the stability of the system will often be represented by a surface plot showing the minimum imaginary part of the zeros of $D(\omega)$, as a_3 and ω_3 are varied. Fig. 9(a) shows such a plot. The system is unstable whenever this plot shows a negative value.

Fig. 9(b) shows the contour plot corresponding to the surface shown in (a). The thick line is the zero contour: it is the threshold of stability of the system. Thinner lines are the contours for more negative values. This plot is the three-mode version of the plots shown in Fig. 4 with two modes. The horizontal line at $-a_3 = a_1 + a_2 = 2$ is clearly visible. The sloping curve across the plot corresponds to the quadratic condition (3) given above in Eq. (34). The curve of approximately circular shape corresponds to the discriminant condition. It is the new, higher order feature introduced by the third mode.

The zero contour is superimposed on the surface plot shown in Fig. 9(a). The surface plot gives some information regarding the "degree" of instability. Three main parts may be distinguished. First, in the front right of the plot, the surface has a cylindrical shape corresponding to the ellipse shape described above. Along and within this cylinder, the surface dips to relatively modest negative values. Second, the region near the line where the third pole amplitude is -2 shows a "canyon". The edges of this "canyon" correspond to conditions (1) and (3) above. Condition (1), which marks the passage through zero of the leading coefficient of the quadratic equation (33), causes one of the roots to move from plus to minus infinity at the crossing. This being well off the scale of this plot, the vertical scale has been limited to show the smooth part of the behaviour. Third, at the back, left hand side of the plot, there is a flat surface. This shows that the system is (marginally) stable for higher values of $|a_3|$.

To link these results to the behaviour observed when the third mode was approximated by a constant residual, it is useful to investigate the individual behaviour of the zeros. This also helps understand how the surface just shown is constructed.

Fig. 10 details the behaviour of the roots by showing (a) the real parts of the roots, (b) their imaginary parts. Plot (c) combines (a) and (b) into a root locus plot. For these three plots, a_3 is set to -1.5, while ω_3 is varied from 0.8 to 1.4. This amounts to looking at the roots within a vertical plane section of the surface plot. This section is marked by a black frame in Fig. 9(a), and a dashed line on the contour plot (b). In Fig. 10(b), the bottom curve is the minimum imaginary part within the section, therefore it is the curve that shapes the surface. The moderate dip between the two fixed frequencies represents a section through the cylinder. The splitting at the lower end of the frequency range appears as a cliff in the surface plot.

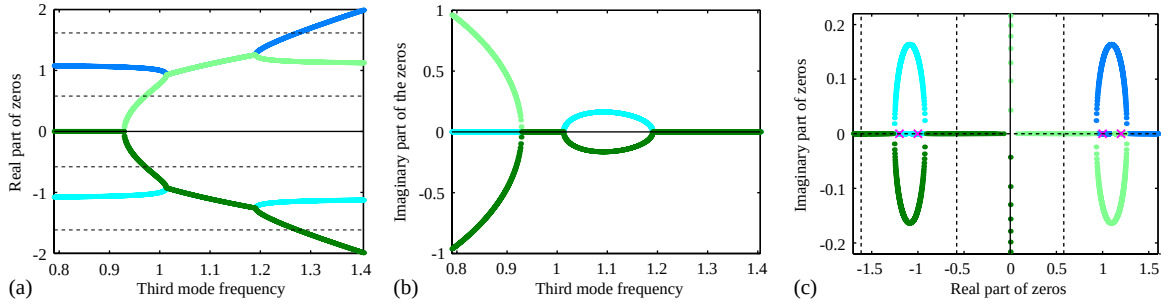


Fig. 10. Plots showing the behaviour of the zeros when the third mode frequency ω_3 is varied from 0.8 to 1.4, while the third mode amplitude is kept constant at -1.5: (a) shows the real part of the zeros, (b) their imaginary part, (c) combines (a) and (b) in a root-locus plot. The crosses in (c) show the position of the poles of the two fixed modes in the complex plane.

These results agree with the conclusions from the previous approximation, where the third mode was modelled as a constant residual. Fig. 10(c) is clearly similar to Fig. 6. Figs. 10(a) and (b) also confirm the previously observed relation between mergings of the real parts and the occurrence of instability. In the undamped case, the system becomes unstable whenever the real parts of two modes of the coupled system merge. Mathematically, this is a direct consequence of the fact that P is a polynomial in ω^2 .

Fig. 10 also reveals that for lower values of ω_3, the roots become *purely imaginary*. The presence of purely imaginary zeros is somewhat problematic. If the system had only 3 modes, the model would be exact and the purely imaginary zeros would simply imply that for those parameter values, the system variables could grow or decay exponentially. However, the objective of these simulations is rather to approximate locally a complex system by a few modes. It is possible that purely imaginary zeros may actually exist for a given system, but it is not possible to tell from the present analysis whether they constitute reliable predictions or not. Given that the analysis is intended to be local, credit should only be granted to those roots which lie in the vicinity of the original poles (the two fixed ones at 1.0 and 1.2 as well as the varying one). The frequency band in which the results can be considered meaningful has been chosen (somewhat arbitrarily) to range from 0.6 to 1.6 (i.e., ± 0.2 the limits of third pole frequency range). This 'range of validity' is indicated in Fig. 10(a) by two dashed lines.

Bearing this in mind, the way the surface is plotted can be modified, so that the only roots included in the calculation of the minimum imaginary part, are those whose real parts lie within this assumed "range of validity". Fig. 11 shows the result when this procedure is applied.

This surface plot shows that the only remaining instability feature is the cylindrical sleeve. As a cross-check, the zeros have also been computed using a positive-pole approximation for $D(\omega)$. The result is not shown because the corresponding surface plot looks exactly the same as Fig. 11. This suggests that this cylindrical feature might indeed play an important role; an argument which becomes even stronger if one recalls that the rounded shape in the root-locus plot, characteristic of a section through the cylinder, was also apparent from the "constant residual" approximation.

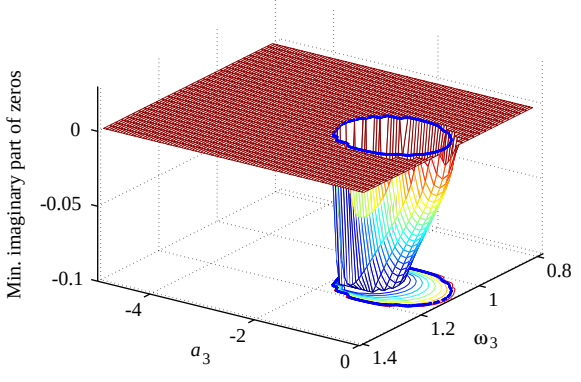


Fig. 11. Surface plot showing the minimum imaginary part of the zeros of a three-mode system where two are kept fixed whereas the third one is varied in amplitude between -5 and 0 and in frequency between 0.8 and 1.4. Only those roots whose real part lies between 0.6 and 1.6 were considered for the computation of the minimum. The zero contour line (stability threshold) is shown on the surface and on the base plane.

From now on, the zeros that have a real part lying outside the "range of validity" will always be ignored. The next task is to investigate the influence of various parameters on the stability. First, the influence of varying the values of the modal amplitudes will be discussed.

4.3.2. *Influence on stability of varying the modal amplitudes*

The expression for the function $D(\omega)$ given by expansion (12) is a linear combination of a_i. This means that only the relative magnitude and sign of these coefficients matters. In the simulation results shown so far, the two fixed modes had an amplitude of 1. When the three modes have three different modal amplitudes, the description given so far is not essentially modified. Simulation cases show that the circular sleeve becomes more "elliptical", in accordance with the properties described in Fig. 8. This suggests that a significant variation of the amplitudes does not add any new feature to the previous description.

4.3.3. *The influence of damping*

The study of the simpler approximations showed that the damping had a rather subtle but important effect on stability. Therefore, it is of interest to investigate the influence of damping on the present three-mode model. To show how damping affects the system, a particular distribution of damping will be described first, following a similar format to that used in the undamped case. Several distributions of damping will then be compared.

If modes 1 and 2 are indeed modes of the disc and mode 3, a mode of the brake, it is plausible that mode 3 has a higher damping than modes 1 and 2. Accordingly, the damping factors will be: $\delta_1 = 0.01$, $\delta_2 = 0.01$, $\delta_3 = 0.03$. To make comparisons with previous plots easier, other parameter values will be given the same values as in the undamped case. Mode 1 will have a frequency and amplitude of 1. The frequency and amplitude of mode 2 will be set at 1.2 and 1, respectively. Mode 3 will be varied in frequency from 0.8 to 1.4, while its amplitude will cover $[-5\ 0]$. Fig. 12 shows the surface plot resulting from simulating the behaviour of the three-mode system with these values.

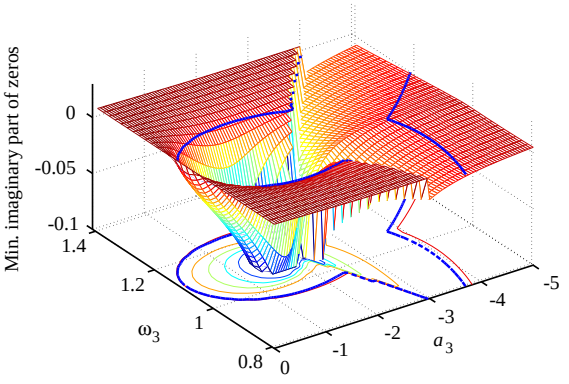


Fig. 12. Surface plots showing the minimum imaginary part of the zeros for a three-mode system with damping. The parameter values are as follow: $\omega_1 = 1$, $a_1 = 1$, $\delta_1 = 0.01$, $\omega_2 = 1.2$, $a_2 = 1$, $\delta_2 = 0.01$, $\delta_3 = 0.03$, a_3 and ω_3 are varied within the ranges shown. Only the zeros whose real part lies in the range of validity are taken into account. The thick line on the surface and bottom plane shows the zero-contour.

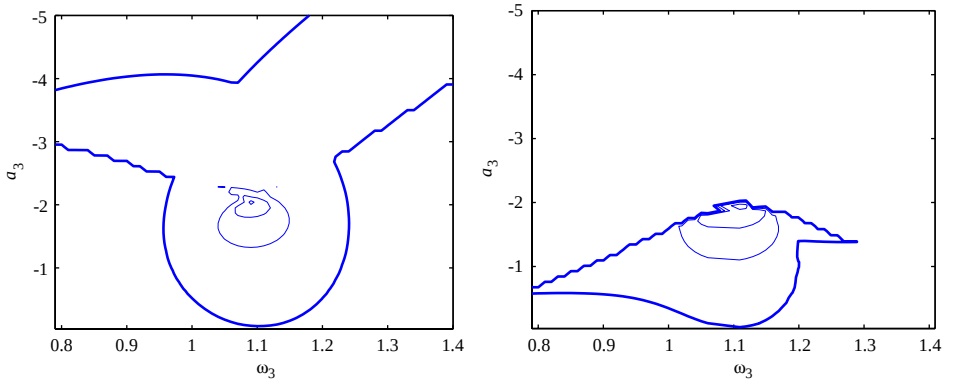
For this plot, only the zeros whose real parts lie within [0.6 1.6] are considered. Damping appears to have distorted the surface, but all the features previously described can still be recognised. Overall, damping has blunted most of the sharp edges, and made stable most of the areas previously marginally stable. The cylindrical sleeve is again an important feature of the plot. The main difference occurs for large negative values of a_3. In this area, the surface now slightly slopes down near the valley. It also has a ridge along the line $\omega_3 = 1.1$. As in the undamped case, ignoring the zero outside the "range of validity" caused the steep valley to disappear. However, a substantial area of the zone $a_3 < -2$ is now unstable. It was marginally stable in the undamped case.

A wide variety of damping distributions have been examined. It is easier to make comparisons by showing zero contour plots. Fig. 13 shows a typical comparison for two different distributions, described in the figure caption. The plots shown here are representative.

These particular cases are shown because one of them, (a), has just been studied in detail. Case (b) is simply a permutation of the damping factor values used in (a). As noted in the two-mode section, any uniform distribution of damping will give the same stability contour as in the undamped case. This contour is visible on the surface shown in Fig. 11. The last possible permutation of the damping factor values ($\delta_1 = 0.03$, $\delta_2 = 0.01$, $\delta_3 = 0.01$) gives a contour similar to that in Fig. 13(b), except that the shape is flipped with respect to the centre line $\omega_3 = 1.1$.

From these plots and others not shown, some important conclusions can be drawn:

1. As noted for the two-mode approximation, when the distribution of damping is uniform, the stability threshold is the same as in the undamped case.
2. Whatever the distribution of damping, the cylinder described in the undamped case appears to be a significant and extremely robust feature. It was shown in the undamped case that this feature occurs for values of a_3 between 0 and $-(a_1 + a_2)$. The stability thresholds for larger negative values of a_3 are more variable. This is probably due to the relative flatness of the



(a) $\delta_1 = 0.01$, $\delta_2 = 0.01$, $\delta_3 = 0.03$ (b) $\delta_1 = 0.01$, $\delta_2 = 0.03$, $\delta_3 = 0.01$

Fig. 13. Plots showing contour lines of the minimum imaginary part surface for two different distributions of damping. The thick line is the zero-contour. Thinner lines are for more negative values. The two fixed modes had parameters set as follows: $\omega_1 = 1$, $a_1 = 1$, $\omega_2 = 1.2$, $a_2 = 1$.

surface in this region: a small variation of the position of the surface can result in a large shift in the threshold line.

3. It seems that the unstable region is larger, whenever the damping is not uniformly distributed. This suggests that, from a stability point of view, a uniform distribution of damping is optimum.
4. The damping affects the stability boundary very significantly and in a non-systematic way. Slight modifications of the damping within a system can result in a very different stability map. This may be one of the reasons for the "capriciousness" of friction-induced vibration phenomena. It is difficult to believe that the natural frequencies or mode shapes of a particular system will change significantly if only minor modifications are made. However, it is plausible that tightening a bolt, or adding a shim, can slightly alter the damping mechanisms within the system, causing it to become stable or unstable.

The next section reviews the concept of "mode locking" in the light of the conclusions reached so far.

4.3.4. Mode-locking

In the brake noise context, the phrase "mode locking" is used by some authors who suggest that instability is more likely to occur when two natural frequencies of the uncoupled subsystems are very close [17]. The present analysis sheds a new light on this idea. The simulation results presented in this paper suggest that with only two modes, instability is not influenced by the closeness of the two natural frequencies. With three modes, it appears that the most robust feature is the shallow dip showing in the surface plots in the damped case, or the cylindrical sleeve in the undamped case. In all simulation results shown, the imaginary part reaches its most negative value when the third mode frequency equals 1.1 (that is mid-way between the two fixed mode frequencies) and when the third mode amplitude equals -2 (that is when the amplitudes add up to zero). This qualifies the concept of 'mode locking': the most general conclusion that can be drawn from this analysis is that instability is more likely to arise when *three consecutive modes have fairly close frequencies* $\omega_1 < \omega_2 < \omega_3$, *whose corresponding modal amplitudes have alternating signs*: $+ - +$ or, less probably, $- + -$. It might be helpful to recall here that the phrase "modal amplitude" refers to the numerator of the resonant terms in $D(\omega)$. The study of the three-mode approximation suggests that the coincidence of two frequencies is a significant property. However, it does not indicate that instability is more likely or any stronger. Rather, it often marks the limit between stability and instability. It also appears from observation of the real part plots (e.g., in Fig. 10(a)), that when the pattern just described is unstable, the unstable frequency of the combined system is very close to the middle frequency ω_2. It is also interesting to note that in this analysis, no distinction is made regarding the origin of a particular mode: once the system is coupled by friction, the modes form a pool of equivalent interacting "modal terms", regardless of which subsystem they originate from. However, as already mentioned, "modal terms" with negative amplitudes can only arise from the brake, if the disc is perfectly symmetrical.

This completes the study of the three-mode approximation. The array of behaviours resulting from the presence of a third mode proved to be much richer and *irreducible* to the behaviour exhibited by only two modes. The next section investigates whether the addition of other modes again results in such drastically different behaviours.

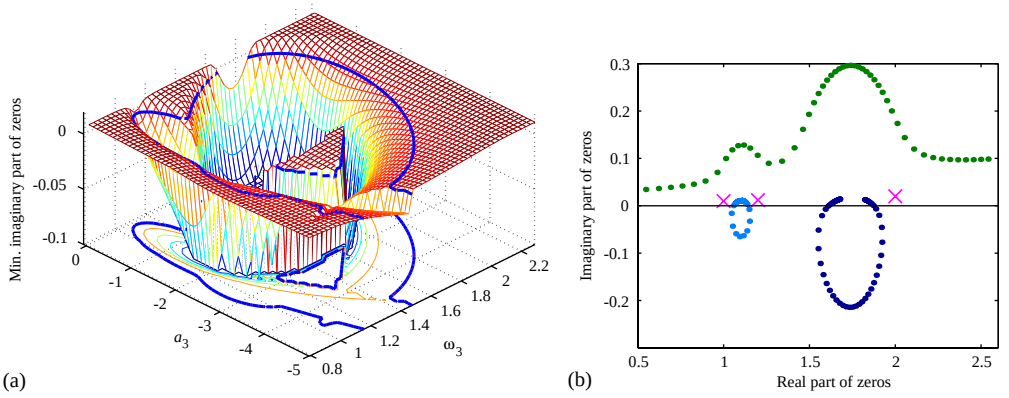


Fig. 14. (a) Surface plot showing the minimum imaginary part of the zeros for a four-mode system containing 2 fixed modes with frequencies 1, 1.2, amplitudes 1, while a third mode is varied in frequency and amplitude. The fourth mode is added with frequency 2 and amplitude $a_4 = +1$. (b) Root locus plot for the four-mode system described in (a) when the third mode amplitude is kept fixed at $a_3 = -1$. These plots show that the behaviour of a three mode system, as shown in Fig. 9(c), is essentially duplicated when the system contains more modes. The crosses represent the position of the fixed poles in the complex plane.

4.4. *Influence of additional modes*

The previous section points toward the important influence on stability of particular three-mode sequences. In order to investigate how many of the previous conclusions hold when the three modes are included in a more complex system, a fourth relatively remote mode is added to the system studied in the previous section.

Fig. 14(a) shows the minimum imaginary part of the (suitable) zeros for a four-mode system. Three modes are fixed. The first two are the same as before with frequencies of 1 and 1.2, and amplitudes of 1. The third mode is varied in frequency and amplitudes. The fourth mode is fixed at a frequency 2 and amplitude 1. In Fig. 14(b), the imaginary parts of the zeros is plotted against their real parts in the section taken from the surface at $a_3 = -1$, while ω_3 is varied from 0.8 to 2.4.

These plots clearly show that with more than three modes, the behaviour described in the three mode section (e.g., in Fig. 10(c)) is reproduced within each cluster of three modes with alternating signs. With more modes, there can be several possibilities to form such clusters, depending on the sign of each amplitude coefficient and the position of the varying frequency ω_3. The root locus plot shown in Fig. 14(b) also suggests that the dimensions of the loop that goes into the unstable half of the complex plane are connected to the interval separating these two fixed mode frequencies (this is clear in the undamped case, as seen from Fig. 8). Therefore, *instability might be stronger for clusters of 3 modes with amplitudes of alternating signs and fairly widely spaced frequencies.*

5. Conclusions

The modelling presented in this paper allows stability to be investigated for any system containing two linear subsystems in sliding contact at a single point. The mechanical behaviour of the two contacting subsystems is captured via a set of transfer functions at the contact point,

regardless of their physical realization. Assuming the damping is proportional, these transfer functions can in turn be expressed in terms of eigenfrequencies, damping factors and mode shapes. Using Coulomb's law with a constant coefficient of friction to model the friction behaviour at the interface, a number of conclusions can be drawn:

1. For such a system to be unstable, it is necessary that the system possesses some asymmetry so that, at least for one of the subsystems, the normal and tangential degrees of freedom are dynamically coupled.
2. It is also necessary that at least one of the "composite modal amplitudes" appearing in expansion (12) becomes negative. Larger values of the coefficient of friction help to reach such negative modal amplitudes.
3. For a three-mode undamped system, instability was shown to arise if and only if the composite modal amplitudes satisfy a certain pattern of signs. If this pattern occurs, instability will occur within a definite region in the parameter space.
4. For a damped three-mode system, damping will usually make the system stable when the corresponding undamped system is marginally stable. Provided damping is light, the stability boundaries of the undamped system are only moderately affected, so that the stability domain is usually recognizable. However, damping can also cause substantial new unstable regions of the parameter space to appear. The exact topography of these unstable regions is highly unpredictable and it strongly depends on the damping distribution. The less uniform the damping in the system, the wider the instability domain of the damped system is likely to be. This suggests that damping distribution is one source for the capriciousness of systems containing a sliding point contact.
5. Many features of systems with more than 3 modes can be understood by viewing them as consisting of fairly independent clusters of 3 modes. Therefore, the conclusions reached from the three-mode system analysis can be used for any other system to a reasonable extent. This justifies a posteriori the appropriateness of using a local approximation for the study of stability.

A major strength of the present modelling is that the theory proposed is inherently non-controversial. When instability is predicted by this linear theory, one would expect something to be observed. It is therefore important to review all the plausible features that might influence stability within linear theory. This is the object of the companion paper [1].

Acknowledgements

The authors thank Bosch Braking Systems for financial support and Professor K.L. Johnson for valuable discussions.

Appendix A. Nomenclature

$u_{1/2}$ displacement at the contact point on the "disc" in the normal/tangential direction
$v_{1/2}$ displacement of the contact point on the "brake" in the normal/tangential direction
F/N total friction/normal force

F_0/N_0 average value of the friction/normal force
F'/N' fluctuating component of the friction/normal force
$\mathbf{G}, \mathbf{H}$ (2×2) receptance matrices at the contact point for the "disc" and the "brake"
μ_0 constant coefficient of friction
ω frequency
ω_i natural frequency of mode "i"
$\bar{\omega}_i$ positive-frequency pole associated with mode "i"
δ_i damping factor of mode "i"
$\phi_i(x/y)$ mass normalised ith mode shape coefficient of the disc in the normal/tangential direction
$\psi(x/y)$ mass normalised ith mode shape coefficient of the brake in the normal/tangential direction
c_i residue associated with pole $\bar{\omega}_i$

References

[1] P. Duffour, J. Woodhouse, Instability of systems with a frictional point contact. Part 2: Model extensions, *Journal of Sound and Vibration* 271 (2004) 391–410, this issue.
[2] S.W.E. Earles, C.K. Lee, Instabilities arising from the frictional interaction of a pin-disc system resulting in noise generation, *Transactions of the American Society of Mechanical Engineers, Journal of Engineering for the Industry* 98 (1976) 81–86.
[3] R.A.C. Fosberry, Z. Holubecki, An investigation of the cause and nature of brake squeal, M.I.R.A. Report 1955/2, 1955.
[4] R.A.C. Fosberry, Z. Holubecki, Some experiments on the prevention of brake squeal, M.I.R.A. Report 1957/1, 1957.
[5] R.A.C. Fosberry, Z. Holubecki, Third experiments on the prevention of brake squeal, M.I.R.A. Report 1957/2, 1957.
[6] R.A.C. Fosberry, Z. Holubecki, Interim report on disc brake squeal, M.I.R.A. Report 1959/4, 1959.
[7] R.A.C. Fosberry, Z. Holubecki, Disc brake squeal, its mechanism and suppression, M.I.R.A. Report 1961/2, 1961.
[8] M.R. North, Disc brake squeal—a theoretical model, M.I.R.A. Research Report 1972/5, 1972.
[9] W.V. Nack, Brake squeal analysis by finite elements, *International Journal of Vehicle Design* 23 (2000) 236–275.
[10] M. Nishiwaki, Generalized theory of brake noise, *Proceedings of the Institution of Mechanical Engineers* 207 (1993) 195–202.
[11] P. Glendinning, *Stability, Instability and Chaos*, Cambridge University Press, Cambridge, 1994.
[12] E. Skudrzyk, *Simple and Complex Vibratory Systems*, Pennsylvania State University Press, Pennsylvania, 1968.
[13] R.C. Dorf, *Modern Control Systems*, Addison-Wesley, Reading, MA, 1980.
[14] R. Lyon, Progressive phase trands in multi-degree-of-freedom systems, *Journal of the Acoustical Society of America* 73 (1983) 1223–1228.
[15] R. Lyon, Range and frequency dependance of transfer function phase, *Journal of the Acoustical Society of America* 76 (1984) 1433–1437.
[16] M. Tohyama, R. Lyon, Zeros of a transfer function in a multi-degree-of-freedom vibrating system, *Journal of the Acoustical Society of America* 86 (1989) 1854–1862.
[17] R. Allgaier, L. Gaul, W. Keipler, K. Willner, Fiction induced vibration and mode lock-in, *Zeitschrift für Angewandte Mathematik und Mechanik* 81 (2001) S49–S52.