A new bipenalty formulation for ensuring time step stability in time domain computational dynamics

Jack Hetherington[1,*], Antonio Rodríguez-Ferran[2] and Harm Askes[1]

[1] *University of Sheffield, Department of Civil and Structural Engineering, Mappin Street, Sheffield S1 3JD, UK*
[2] *Universitat Politècnica de Catalunya, Laboratori de Càlcul Numèric, Jordi Girona 1–3, 08034 Barcelona, Spain*

SUMMARY

It is well known that use of standard penalty methods can decrease the critical time step of time domain dynamic finite element analyses. The bipenalty method utilises both stiffness and mass penalties in order to impose constraints that have a minimal effect on the eigenfrequencies of the finite element system. One way of achieving this goal is to find a ratio of stiffness and mass penalty parameters—the critical penalty ratio (CPR)—that does not affect the maximum eigenfrequency (and therefore, for conditionally stable solution schemes, the critical time step) of a system. In this contribution, we develop a new method of calculating the CPR associated with a finite element formulation by examining the eigenvalue problem in detail. Advantages of the method compared with previous solutions include increased simplicity and generality, and the ability to consider multiple constraints. The method is demonstrated by deriving CPRs for a few finite element formulations, which are then verified using simple numerical examples. The superiority of the bipenalty method over standard mass penalty methods is also demonstrated.

KEY WORDS: penalty method; finite element method; constraints; critical time step; explicit time integration; bipenalty method

1. INTRODUCTION

Penalty methods are a popular technique for imposing constraints in numerical analysis. The most common method (referred to as the *large stiffness* or *stiffness penalty method*) consists of adding a large stiffness value, the *stiffness penalty parameter*, to the structural stiffness matrix of a system. In this way, we can set a solution variable to a certain value (absolute constraint) or link together more than one degree of freedom (relative constraint). The method is flexible, simple to implement and the size of the system is unaltered. The most obvious disadvantage of the method, in constrast with direct imposition or the Lagrange multiplier method, is that constraints are only approximately imposed; however, for large penalty parameters the errors introduced are usually well within acceptable bounds [1]. In addition, methods utilising negative penalties of large magnitude have recently provided ways of quantifying and reducing the modelling errors introduced by standard penalty methods, although they can introduce further problems regarding stability and conditioning of the system [2].

In dynamics, however, using the stiffness penalty method alone has a further disadvantage: introducing large stiffness values into the system can radically alter the eigenfrequencies associated with the problem; more specifically, the maximum eigenfrequency is greatly increased. This has a

---

*Correspondence to: cip09jeh@sheffield.ac.uk

few important implications. Firstly, the 'stiffness' of the system of equations (the word 'stiffness' in this case is used in the mathematical sense and defined as $S = \omega_{\max}/\omega_{\min}$, where ω_i are the associated eigenfrequencies) can be radically altered, which has implications for the stability and accuracy of the solution [3]. In time domain dynamic analyses using conditionally stable time integration schemes there is a related problem: the critical time step of an analysis can be greatly decreased by the use of penalty methods [4]. The critical time step Δt_{crit} is the time step above which a conditionally stable solution procedure (such as the popular central difference time integration scheme) becomes unstable and it is a function of the maximum eigenfrequency, $\omega_{\max}$. Decreasing the critical time step is significant because if the decrease is not accounted for the solution may become unstable, leading to meaningless results. Furthermore, even if it *is* accounted for when selecting a time step the computational cost of the analysis can increase significantly.

In dynamic analysis, penalty methods can also be applied to the mass matrix of a system [5, 6]. This technique is known as the *large mass*, *mass penalty* or *inertia penalty* method. Again, constraints imposed using mass penalties are only approximate in nature but, instead of increasing the maximum eigenfrequency of the system, one or more eigenfrequencies are greatly reduced, tending to zero for large penalty parameters [5]. This implies that when used simultaneously, the stiffness and mass penalties can be tuned such that the maximum eigenfrequency (along with the order of the stiffness, S, of the system equations) can be preserved. This technique will be referred to here as the *bipenalty method*, and in the present work we derive a method for ensuring the maximum eigenfrequency of a bipenalised element—and therefore the critical time step—is unaffected by the addition of the penalty. A secondary motivation for using both penalty types at once is a reduction in the modelling error that is inherent to all penalty methods; by applying stiffness *and* mass penalties we are effectively constraining both displacement and acceleration, resulting in a smaller overall violation of the displacement constraint.

The proposed method builds on the work of Askes et al. [7], who developed a method based on the calculation of a *critical penalty ratio* (CPR) for various finite element formulations. The CPR gives the highest ratio of stiffness and mass penalties that can be used without increasing the maximum eigenfrequency of an element. However, since no knowledge is assumed regarding how bipenalties in general affect the eigensolutions of the system, the method of Askes et al. [7] requires an analysis of the full, bipenalised problem. This limits the complexity of elements that can be considered, and increases the difficulty of obtaining closed form solutions for the CPR. In addition, only single absolute constraints are considered, restricting the usefulness of the results for practical applications.

A similar bipenalty formulation (referred to as a "consistent form of the penalty method") has also been examined by Paraskevopoulos et al. [3]. They propose using ratios that depend on the degree of freedom being constrained, and provide a simple method for selecting ratios that do not significantly affect the stiffness of the system. However, a penalty ratio that ensures that the critical time step of an analysis is unaffected is not provided, since their focus is the frequency domain. Asano [8, 9, 10] also used simultaneous stiffness, mass and damping penalties in a specialised contact-impact formulation. However, these works focus on computational accuracy and the effects of the penalties on the eigensolutions of a system are not explored.

In this paper, we derive a simple relationship between the CPR of an element and its maximum unpenalised eigenfrequency. This is achieved through a detailed examination of the effects of bipenalisation on the eigensolutions associated with a generic finite element. As such, the method of CPR calculation is simpler and more general than that proposed by Askes et al. [7], requiring knowledge of the maximum eigenfrequency of the unpenalised element only. This means that multiple constraints and more complex element formulations can be accounted for, and, as will be shown, the method also leads to a refinement of some of the results given previously by Askes et al.

2. THE BIPENALISED PROBLEM

In this paper we concentrate our attention on the partial differential equations for elastodynamics, the standard FE discretisation for which is written

$$\mathbf{M\ddot{u} + Ku = f} \tag{1}$$

where $\mathbf{M}$ and $\mathbf{K}$ are the mass and stiffness matrices of the system, $\mathbf{u}$ is the vector of displacements and $\mathbf{f}$ is the vector of external forces. To set a displacement to zero using a penalty of stiffness type, the potential energy of the system is modified so that

$$\mathcal{U} = \frac{1}{2}\mathbf{u}^T\mathbf{Ku} - \mathbf{u}^T\mathbf{f} + \frac{1}{2}\alpha_{\mathrm{s}}u_i^2 \tag{2}$$

where α_{s} is the stiffness penalty parameter and u_i is the DOF to be prescribed. To include an inertia-type penalty, the kinetic energy of the system is also modified [7], so that

$$\mathcal{T} = \frac{1}{2}\dot{\mathbf{u}}^T\mathbf{M}\dot{\mathbf{u}} + \frac{1}{2}\alpha_{\mathrm{m}}\ddot{u}_i^2 \tag{3}$$

where α_{m} is the mass penalty parameter. As α_{s} and α_{m} tend to infinity, the constraints added to (2) and (3) are more accurately enforced, respectively. The modified dynamic equations of equilibrium are then given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{T}}{\partial\dot{\mathbf{u}}^T} + \frac{\partial\mathcal{U}}{\partial\mathbf{u}^T} = \left[\mathbf{M} + \mathbf{M}^{\mathrm{P}}\right]\ddot{\mathbf{u}} + \left[\mathbf{K} + \mathbf{K}^{\mathrm{P}}\right]\mathbf{u} = \mathbf{f} \tag{4}$$

where the penalty stiffness and mass matrices are zero except for $K_{ii}^{\mathrm{P}} = \alpha_{\mathrm{s}}$ and $M_{ii}^{\mathrm{P}} = \alpha_{\mathrm{m}}$. The penalty parameters have the same units as the entries to which they are added, K_{ii} and M_{ii} (in this case $\mathrm{N/m}$ and kg respectively).

In the present work we consider an arbitrary finite element with n degrees of freedom subject to k absolute constraints imposed using the bipenalty method, as described above. Our goal is to preserve the critical time step of the analysis. To this end, we consider only conditionally stable Newmark time integration schemes, which have a critical time step given by

$$\Delta t_{\mathrm{crit}} = \frac{\Omega_{\mathrm{crit}}}{\omega_{\mathrm{max}}} \tag{5}$$

where Ω_{crit} is the 'critical sampling frequency' and ω_{max} is the maximum eigenfrequency of a system, sub-system, or finite element [11]. The critical sampling frequency is dependent only on damping parameters and the time integration scheme being employed. For undamped systems $\Omega_{\mathrm{crit}} = (\gamma/2 - \beta)^{-1/2}$ where γ and β are the standard Newmark parameters. For the explicit central difference scheme used in Sections 6 and 7, $\gamma = 1/2$ and $\beta = 0$ so that $\Delta t_{\mathrm{crit}} = 2/\omega_{\mathrm{max}}$. Since penalties do not affect the critical sampling frequency, we turn our attention to the maximum eigenfrequency of the system, the value of which can be greatly affected by the use of penalty methods.

The maximum eigenfrequency is proportional to wave speed, which is in turn related to the ratio of stiffness to mass (e.g., for a 1D bar the elastic wave speed is given by $c_e = \sqrt{E/\rho}$). Hence, stiffness penalties tend to *increase* the wave speed and maximum eigenfrequency and therefore *decrease* the critical time step, while the inverse is true for mass penalties [6, 7]. The goal here is to develop a method of calculating the critical penalty ratio for any finite element; that is, the ratio of stiffness and mass penalty parameters that preserves the maximum eigenfrequency (and therefore the critical time step) of that element.

The unconstrained, unpenalised eigenvalue problem (UP) for a finite element is given by

$$\left(\mathbf{K} - \omega_i^2\mathbf{M}\right)\mathbf{u}_i = \mathbf{0} \tag{6}$$

where $\mathbf{K}$ and $\mathbf{M}$ are the stiffness and mass matrices, respectively, and the eigenvectors $\mathbf{u}_i$ and corresponding eigenfrequencies ω_i form the n solutions (ordered so that $\omega_1 \leq \omega_2 \leq \ldots \leq \omega_n \equiv$

ω_{max}). The bipenalised problem (BP), in which absolute constraints have been applied to the first k DOF, is given by

$$\left([\mathbf{K} + \mathbf{K}^{\text{p}}] - \tilde{\omega}_i^2 [\mathbf{M} + \mathbf{M}^{\text{p}}]\right) \tilde{\mathbf{u}}_i = \mathbf{0} \tag{7}$$

where $\mathbf{K}^{\text{p}}$ and $\mathbf{M}^{\text{p}}$ contain the stiffness and mass penalty parameters and the n solutions are given by $\tilde{\mathbf{u}}_i$ and $\tilde{\omega}_i$ (once again ordered so that $\tilde{\omega}_1 \leq \tilde{\omega}_2 \leq \ldots \leq \tilde{\omega}_n \equiv \tilde{\omega}_{\text{max}}$). Since we will be considering only single finite elements in this work, we assume that the penalty parameters for all DOF have the same magnitude so that

$$\mathbf{K}^{\text{p}} = \alpha_{\text{s}} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \qquad \mathbf{M}^{\text{p}} = \alpha_{\text{m}} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{8}$$

where $\mathbf{I}_k$ is the identity matrix of size $k \times k$.

The penalty ratio is defined as

$$R = \frac{\alpha_{\text{s}}}{\alpha_{\text{m}}} \tag{9}$$

and has units of s^{-2} for structural systems. This is a different definition to that used by Askes et al., where the ratio is defined in terms of *penalty factors* and written as

$$r = \frac{p_{\text{s}}}{p_{\text{m}}} \tag{10}$$

where p_{s} and p_{m} are the stiffness and mass penalty factors, respectively, defined in terms of the existing entries in the matrix they are being added to by

$$p_{\text{s}} = \frac{\alpha_{\text{s}}}{K_{jj}}; \qquad p_{\text{m}} = \frac{\alpha_{\text{m}}}{M_{jj}} \tag{11}$$

where j is the number of the penalised DOF. Penalty factors are commonly used in practice because they better describe the effectiveness of a penalty (since the accuracy depends on the *relative* size of a penalty, not its absolute value), and because they are dimensionless quantities. However, in the present work we have found it advantageous to consider penalty parameters α_{s} and α_{m}, rather than penalty factors p_{s} and p_{m}. The penalty ratio given in (10) will be referred to as as the *dimensionless penalty ratio*. It can be calculated from the penalty ratio, R, by

$$r = \left(\frac{M_{jj}}{K_{jj}}\right) \cdot R \tag{12}$$

The CPR is the highest ratio of penalty parameters that does not increase the maximum eigenfrequency of the element, and is denoted by R_{crit} (or r_{crit} for the corresponding dimensionless penalty ratio). It may be calculated (for a given element formulation) by the following process [7]:

1. Solve Equation (6) for the unpenalised eigenfrequencies, ω_i.
2. Select the largest eigenfrequency, $\omega_{\text{max}} = \max(\omega_i)$.
3. Substitute ω_{max} for $\tilde{\omega}_i$ in Equation (7) and solve for the penalty ratio, R, which has ω_{max} as a solution.

A simpler and more robust method has been developed here by directly linking the CPR to the properties of the unpenalised system. Previous results of CPR calculations (which will be revisited in Section 4), along with numerical experiments and study of the underlying eigenvalue problem, led to the hypothesis that R_{crit}, the critical penalty ratio, is equal to the maximum eigenvalue of the unpenalised elements. That is,

$$R_{\text{crit}} = \omega_{\text{max}}^2 \tag{13}$$

This suggests that the CPR can be calculated by finding only the maximum eigenfrequency of the *unpenalised* system, negating the need for Step 3 above in which the bipenalised eigenvalue problem is solved. Discarding this step greatly simplifies the calculations, since it is the most time consuming part of the process and becomes increasingly complex for multiple constraints and more sophisticated finite elements. In addition, penalty parameters may be calculated with no knowledge of how many constraints are added, which DOF are being constrained, etc. In the next section, a proof of the observation given in (13) is presented for k absolute constraints imposed on an arbitrary finite element.

3. RELATING CPR TO MAXIMUM UNPENALISED EIGENFREQUENCY

The proof consists of two parts. Firstly, we show that if the penalty ratio R is set to ω_n^2 (the square of the maximum unpenalised eigenfrequency) then ω_n is the largest eigenfrequency of the BP, as well as the UP. Interestingly, this proof does not require any assumptions about the magnitude of the penalty parameters. Secondly, we show that as the penalty parameters tend to infinity $\sqrt{R}$ becomes an eigenfrequency of the BP, irrespective of its value; hence, for $R > \omega_n^2$, the maximum eigenfrequency of the BP is greater than that of the UP. From this we may conclude that the critical penalty ratio is equal to the maximum eigenfrequency squared of the unpenalised element; i.e., $R_{\mathrm{crit}} = \omega_{\mathrm{max}}^2$.

In this section we utilise the Rayleigh quotient, defined as

$$\rho(\mathbf{v}) = \frac{\mathbf{v}^T \mathbf{K} \mathbf{v}}{\mathbf{v}^T \mathbf{M} \mathbf{v}} \tag{14}$$

where $\mathbf{v}$ is any non-zero vector. Two useful properties of the Rayleigh quotient are [1]:

- The Rayleigh quotient of an eigendirection $\beta \mathbf{u}_i$ (for any $\beta \neq 0$) is the corresponding eigenfrequency squared,

$$\rho(\beta \mathbf{u}_i) = \omega_i^2 \tag{15}$$

- The Rayleigh quotient of any non-zero vector $\mathbf{v}$ is bounded by the minimum and maximum eigenfrequencies squared,

$$\omega_1^2 \leq \rho(\mathbf{v}) \leq \omega_n^2 \tag{16}$$

We would like to note also that the following proofs are valid for positive or negative penalty parameters. In addition, although the focus of this work is elastodynamics, we note that we can proceed in a similar way for the discretisation of any second order hyperbolic PDE.

Lemma 1
If $R = \omega_n^2$, the eigenpair $(\omega_n, \mathbf{u}_n)$, an eigensolution of the unpenalised problem, is also an eigensolution of the bipenalised problem.

Proof
For the eigenpair $(\omega_n, \mathbf{u}_n)$ to be a solution of the BP as well as the UP, we must make sure that the solution satisfies (7) as well as (6). Since, from (6),

$$\left(\mathbf{K} - \omega_n^2 \mathbf{M} \right) \mathbf{u}_n = \mathbf{0} \tag{17}$$

and, for $R = \omega_n^2$,

$$\mathbf{K}^{\mathrm{p}} - \omega_n^2 \mathbf{M}^{\mathrm{p}} = \mathbf{0} \tag{18}$$

then it is clear that we also have

$$\left(\mathbf{K} - \omega_n^2 \mathbf{M} \right) \mathbf{u}_n + \left(\mathbf{K}^{\mathrm{p}} - \omega_n^2 \mathbf{M}^{\mathrm{p}} \right) \mathbf{u}_n = \mathbf{0} \tag{19}$$

and, therefore, $(\omega_n, \mathbf{u}_n)$ is an eigensolution of the BP for $R = \omega_n^2$. $\square$

Lemma 2
If $R = \omega_n^2$, there is no eigenfrequency of the bipenalised problem, $\tilde{\omega}_i$, that exceeds the maximum eigenfrequency of the unpenalised problem, ω_n.

Proof
Here we consider the Rayleigh quotient associated with the BP,

$$\rho_{\mathrm{BP}}(\mathbf{v}) := \frac{\mathbf{v}^T \left(\mathbf{K} + \mathbf{K}^{\mathrm{p}} \right) \mathbf{v}}{\mathbf{v}^T \left(\mathbf{M} + \mathbf{M}^{\mathrm{p}} \right) \mathbf{v}} = \frac{\mathbf{v}^T \mathbf{K} \mathbf{v} + \alpha_{\mathrm{s}} \|\mathbf{v}^{\mathrm{p}}\|^2}{\mathbf{v}^T \mathbf{M} \mathbf{v} + \alpha_{\mathrm{m}} \|\mathbf{v}^{\mathrm{p}}\|^2} \tag{20}$$

where $\mathbf{v}^{\mathrm{p}}$ is a vector containing only the first k entries of $\mathbf{v}$ (i.e., the entries associated with the bipenalised DOF only). Substituting in the Rayleigh quotient of the UP, $\rho_{\mathrm{UP}}(\mathbf{v})$, and taking into

account (16) we have

$$\rho_{\mathrm{BP}}(\mathbf{v}) = \frac{\rho_{\mathrm{UP}}(\mathbf{v})\mathbf{v}^T\mathbf{M}\mathbf{v} + R\alpha_{\mathrm{m}}\|\mathbf{v}^{\mathrm{p}}\|^2}{\mathbf{v}^T\mathbf{M}\mathbf{v} + \alpha_{\mathrm{m}}\|\mathbf{v}^{\mathrm{p}}\|^2} \leq \omega_n^2\left(\frac{\mathbf{v}^T\mathbf{M}\mathbf{v} + \alpha_{\mathrm{m}}\|\mathbf{v}^{\mathrm{p}}\|^2}{\mathbf{v}^T\mathbf{M}\mathbf{v} + \alpha_{\mathrm{m}}\|\mathbf{v}^{\mathrm{p}}\|^2}\right) = \omega_n^2 \qquad (21)$$

Therefore, $\rho_{\mathrm{BP}}(\mathbf{v})$ is bounded from above by ω_n^2. $\qquad\square$

Theorem 1
If the penalty ratio is selected such that $R = \omega_n^2$, the maximum eigenfrequency of the bipenalised problem is equal to that of the unpenalised problem.

Proof
Since, $\rho_{\mathrm{BP}}(\mathbf{v})$ is bounded from above by ω_n^2 (Lemma 2), we conclude that BP has no eigenfrequency larger than ω_n and therefore, since ω_n is an eigenfrequency of BP (Lemma 1),

$$\tilde{\omega}_n = \omega_n \qquad (22)$$

for the case where $R = \omega_n^2$. $\qquad\square$

Theorem 2
If the penalty ratio is selected such that $R > \omega_n^2$, the maximum eigenfrequency of the bipenalised problem is greater than that of the unpenalised problem, for large α_{m}.

Proof
Again we consider the Rayleigh quotient of the bipenalised problem, given in (20). This time we substitute for $\mathbf{v}$ the n eigenvectors of the BP, giving

$$\rho_{\mathrm{BP}}(\tilde{\mathbf{u}}_i) = \frac{\tilde{\mathbf{u}}_i^T\mathbf{K}\tilde{\mathbf{u}}_i + R\alpha_{\mathrm{m}}\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\|^2}{\tilde{\mathbf{u}}_i^T\mathbf{M}\tilde{\mathbf{u}}_i + \alpha_{\mathrm{m}}\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\|^2} \qquad (23)$$

where $\tilde{\mathbf{u}}_i^{\mathrm{p}}$ is a vector containing only the first k entries of $\tilde{\mathbf{u}}_i$. Since $\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\|$ cannot be zero for n linearly independent eigenvectors, we may assume there are eigenmodes for which $\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\| \neq 0$. In this case, taking limits for $\alpha_{\mathrm{m}} \to \infty$ results in

$$\lim_{\alpha_{\mathrm{m}}\to\infty}\rho_{\mathrm{BP}}(\tilde{\mathbf{u}}_i) = \lim_{\alpha_{\mathrm{m}}\to\infty}\frac{\tilde{\mathbf{u}}_i^T\mathbf{K}\tilde{\mathbf{u}}_i + R\alpha_{\mathrm{m}}\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\|^2}{\tilde{\mathbf{u}}_i^T\mathbf{M}\tilde{\mathbf{u}}_i + \alpha_{\mathrm{m}}\|\tilde{\mathbf{u}}_i^{\mathrm{p}}\|^2} = R \qquad (24)$$

Therefore, in the limit, and using Equation (15), $R > \omega_n^2$ is an eigenfrequency squared of the BP. $\quad\square$

Theorems 1 and 2 provide all we require to ensure that we can calculate the CPR of an element with the simple equation $R_{\mathrm{crit}} = \omega_{\max}^2$, assuming that the penalty parameters are large. Further observations concerning the effect of bipenalisation on the eigenfrequencies and eigenvectors of a system are given in A.

4. REVISITING EARLIER RESULTS USING THE NEW FORMULATION

In order to verify the new method of CPR calculation, we will compare results to those given in [7]. To do this, we will return to the dimensionless penalty ratio, r, as defined in Section 2. For each element type (see the original work for details of the element formulations) the maximum eigenfrequency is determined analytically, and, using Equations (12) and (13), the dimensionless penalty ratio is then calculated.

A comparison of results is given in Table I. The final two columns give the previously calculated penalty ratios and results as calculated by the new method, respectively. The results agree exactly for the majority of cases.

For beam elements with bipenalised displacement *and* rotation, there is always some ambiguity concerning how to determine r_{crit}. For example, Askes et al. report that use of their method results

Element type	DOF	Mass	ν	r_{crit} (Askes et al.)	$r_{\text{crit}} = \left(\frac{M_{jj}}{K_{jj}}\right) \cdot \omega_{\text{max}}^2$ (new method)
Bar	displacement	lumped	-	2	2
		consistent	-	4	4
Beam	displacement	lumped	-	8	8
	rotation			2	2
	displ. & rot.			$\downarrow 2$	2
	displacement	consistent	-	260	260
	rotation			20	20
	displ. & rot.			$\downarrow 20$	20
Square (plane stress)	displacement	lumped	all	$\frac{6(1+\nu)}{3-\nu}$	$\frac{6(1+\nu)}{3-\nu}$
		consistent	$< \frac{1}{3}$	$\geq \frac{8(1+\nu)}{3-\nu}$ $(*)$	4
			$\geq \frac{1}{3}$	$\frac{8(1+\nu)}{3-\nu}$	$\frac{8(1+\nu)}{3-\nu}$
Square (plane strain)	displacement	lumped	all	$\frac{6}{3-4\nu}$	$\frac{6}{3-4\nu}$
		consistent	$< \frac{1}{4}$	$\geq \frac{8}{3-4\nu}$ $(*)$	4
			$\geq \frac{1}{4}$	$\frac{8}{3-4\nu}$	$\frac{8}{3-4\nu}$

Table I. Comparison of dimensionless CPRs given in [7] and those calculated using the new method. Values marked with an asterisk $(*)$ were postulated and verified numerically.

in an expression for r_{crit} with two solutions in the limit $\alpha_{\text{m}} \to \infty$; one corresponds to the value obtained for a single displacement constraint, while the other corresponds to the single rotational constraint. When calculating the CPR in terms of penalty parameters there is no ambiguity since ω_{max} does not depend on which DOF is being constrained. However, the ratio M_{jj}/K_{jj} in Equation (12) does depend on which constrained DOF is being considered; thus, there is more than one possible value of r_{crit}. Essentially, there are two options: use a different dimensionless ratio r for each DOF, leading to a constant R for all DOF and therefore a crisp bound on stability; or choose the lowest calculated value, which gives a safe but conservative result. For simplicity, the latter option was chosen when compiling the results in Table I.

The results obtained for square elements with low Poisson's ratio also differ between the two methods. By the previous method, for the plane stress square element with consistent mass, we have $0 < r_{\text{crit}} < 4$ over the range $-1 < \nu < \frac{1}{3}$. This result was actually postulated and verified numerically, since an analytical solution could not be found with the earlier method. By the new method, we find that r_{crit} is actually independent of ν and equal to the upper bound of the result previously reported; thus, it does not contradict the results of the previously conducted stability analysis.

5. NEW RESULTS OF CPR CALCULATIONS

We will now derive new closed form expressions for the CPR for a few more common 2D element types. The process in each case is the same:

1. Derive the element stiffness and mass matrices.
2. Solve $\det\left(\mathbf{K} - \omega_i^2 \mathbf{M}\right) = 0$ for ω_i.
3. Select $\max(\omega_i)$ and calculate CPR from $R_{\text{crit}} = \omega_{\text{max}}^2$.

The element stiffness and consistent mass matrices are formulated in each case by

$$K = \int_A B^T DB \; dA \tag{25}$$

$$M = \rho \int_A H^T H \; dA \tag{26}$$

where D is the matrix of constitutive relations, ρ is the density, and A is the area of the element. The m shape functions are given by $N = [N_1, N_2, \ldots, N_m]$, where m is the number of nodes per element, and

$$H = \begin{bmatrix} N_1 & 0 & N_2 & 0 & \cdots & N_m & 0 \\ 0 & N_1 & 0 & N_2 & \cdots & 0 & N_m \end{bmatrix} \tag{27}$$

$$B = \begin{bmatrix} \dfrac{\partial N_1}{\partial x} & 0 & \dfrac{\partial N_2}{\partial x} & 0 & \cdots & \dfrac{\partial N_m}{\partial x} & 0 \\ 0 & \dfrac{\partial N_1}{\partial y} & 0 & \dfrac{\partial N_2}{\partial y} & \cdots & 0 & \dfrac{\partial N_m}{\partial y} \\ \dfrac{\partial N_1}{\partial y} & \dfrac{\partial N_1}{\partial x} & \dfrac{\partial N_2}{\partial y} & \dfrac{\partial N_2}{\partial x} & \cdots & \dfrac{\partial N_m}{\partial y} & \dfrac{\partial N_m}{\partial x} \end{bmatrix} \tag{28}$$

For each element under examination we will state the shape functions and the resulting B matrix that we use in the formulation. The constitutive matrix D is defined according to the assumptions of plane stress or plane strain in the usual way—see, for example, [1, Table 4.3]. Once each element has been formulated, closed form expressions for the eigenfrequencies of the element are computed using the symbolic mathematical operations software Maple. Since the dominant mode of deformation (and therefore the maximum eigenfrequency) often depends on the Poisson's ratio, ω_{max} for each formulation is determined by inspection. The process in general is similar to that found in [12], where eigenvalues are calculated for the purpose of finding accurate expressions for the critical time step for some 4- and 8-noded element formulations.

5.1. 4-noded square with reduced integration

In practical applications of the finite element method stiffness and mass matrices are generally obtained using numerical, rather than analytical, integration methods. The resulting matrices are accurate only if a sufficient number of integration points are used. However, reducing the number of integration points, while reducing the accuracy of the numerical integration, can increase the accuracy of the overall solution, as well as increasing the computational efficiency of the formulation. To investigate how using reduced integration may affect the CPR associated with a finite element, we consider a 4-noded square element, as shown in Figure 1.
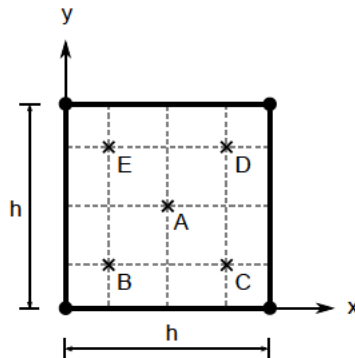

Figure 1. Square 4-noded quadrilateral element with side length h and integration points A to E. For exact position of integration points see [1, Table 5.7].

The shape functions for the square element can be stated directly in terms of the global coordinates, x and y:

$$N_1 = \left(1 - \frac{x}{h}\right)\left(1 - \frac{y}{h}\right) \tag{29}$$

$$N_2 = \frac{x}{h}\left(1 - \frac{y}{h}\right) \tag{30}$$

$$N_3 = \frac{xy}{h^2} \tag{31}$$

$$N_4 = \frac{y}{h}\left(1 - \frac{x}{h}\right) \tag{32}$$

For reduced integration we use only one integration point (point A in Figure 1) which is in the centre of the element at position $\left(\frac{h}{2}, \frac{h}{2}\right)$. To obtain the $\mathbf{B}$ matrix we substitute these x and y values into the standard formulation defined by (28), giving

$$\mathbf{B}^{\text{red}} = \frac{1}{2h}\begin{bmatrix} -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 \end{bmatrix} \tag{33}$$

However, this results in an element with two spurious zero-energy modes which may not be suitable for use in practice. A more reliable formulation may be obtained using selective integration; that is, by reducing the order of the integration for the shear terms of the $\mathbf{B}$ matrix only, while using exact integration for all other terms. This results in

$$\mathbf{B}^{\text{sel}} = \frac{1}{h^2}\begin{bmatrix} y-h & 0 & h-y & 0 & y & 0 & -y & 0 \\ 0 & x-h & 0 & -x & 0 & x & 0 & h-x \\ -\dfrac{h}{2} & -\dfrac{h}{2} & -\dfrac{h}{2} & \dfrac{h}{2} & \dfrac{h}{2} & \dfrac{h}{2} & \dfrac{h}{2} & -\dfrac{h}{2} \end{bmatrix} \tag{34}$$

The calculated R_{crit} values are given in Table II. The calculated values of R_{crit} were found to be the same for both fully and selectively reduced integration. The eigenvalues for a fully reduced, lumped mass element are given by Ling and Cherukuri [12] and they agree with the results presented here.

	Lumped mass	Consistent mass
Plane stress	$\frac{4}{1-|\nu|}\left(\frac{c_e}{h}\right)^2$	$\frac{12}{1-|\nu|}\left(\frac{c_e}{h}\right)^2$
Plane strain	$\frac{4}{1+\nu}\left(\frac{c_e}{h}\right)^2 \quad$ for $\nu < 0$	$\frac{12}{1+\nu}\left(\frac{c_e}{h}\right)^2 \quad$ for $\nu < 0$
	$\frac{4}{(1+\nu)(1-2\nu)}\left(\frac{c_e}{h}\right)^2 \quad$ for $\nu \geq 0$	$\frac{12}{(1+\nu)(1-2\nu)}\left(\frac{c_e}{h}\right)^2 \quad$ for $\nu \geq 0$

Table II. Calculated R_{crit} for a square element formulated using fully reduced/selectively reduced integration.

Since positive and negative values of Poisson's ratio ν lead to different expressions for the maximum eigenvalue, the expression used to calculate R_{crit} depends on the sign of ν. For plane stress this can be expressed using the absolute value $|\nu|$. For plane strain analyses, different expressions must be used. It is worth noting, however, that materials with $\nu < 0$ are rare in most fields of engineering analysis.

Finally, we note that in all cases the results for lumped and consistent mass are of the same form, differing only by a constant factor of three.

5.2. Right-angled triangle

The next element we consider is a 3-noded right-angled triangle, as shown in Figure 2. The shape functions for this element are again stated directly in terms of the global co-ordinates, x and y:

$$N_1 = 1 - \frac{x}{h} - \frac{y}{h} \tag{35}$$

$$N_2 = \frac{x}{h} \tag{36}$$

$$N_3 = \frac{y}{h} \tag{37}$$

Triangular elements need only one integration point in order to be integrated exactly, therefore reduced integration is not possible here. The B matrix is given by

$$\mathbf{B} = \frac{1}{h} \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \\ -1 & -1 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{38}$$

and the expressions for R_{crit} are given in Table III. Again, the expressions for lumped and consistent mass differ only by a constant factor; this time by a factor of four. However, for this formulation the dominant deformation mode is the same for all ν, so there is only one expression for R_{crit} in each case.
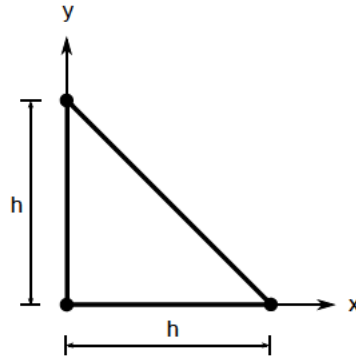

Figure 2. Right-angled 3-noded triangular element with equal height and base width, h.

	Lumped mass	Consistent mass
Plane stress	$\frac{3\left(2+\sqrt{3\nu^2+1}\right)}{1-\nu^2}\left(\frac{c_e}{h}\right)^2$	$\frac{12\left(2+\sqrt{3\nu^2+1}\right)}{1-\nu^2}\left(\frac{c_e}{h}\right)^2$
Plane strain	$\frac{3\left(2-2\nu+\sqrt{4\nu^2-2\nu+1}\right)}{(1+\nu)(1-2\nu)}\left(\frac{c_e}{h}\right)^2$	$\frac{12\left(2-2\nu+\sqrt{4\nu^2-2\nu+1}\right)}{(1+\nu)(1-2\nu)}\left(\frac{c_e}{h}\right)^2$

Table III. Calculated R_{crit} for a right-angled triangular element.

5.3. Parallelogram (distorted square)

In order to examine the effect of element distortion on the CPR we analyse the distorted square element shown in Figure 3. The additional variable, θ, somewhat complicates the calculation of eigenvalues; for simplicity, all sides are given a constant length of unity.
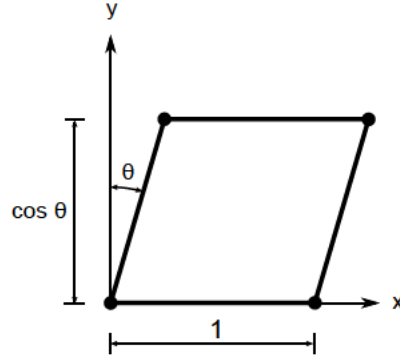
Figure 3. 4-noded quadrilateral element with sides of unit length and a simple transformation defined by θ.

Here we introduce a local co-ordinate system in terms of ξ and η, and write the shape functions as

$$N_1 = (1 - \xi)(1 - \eta) \tag{39}$$
$$N_2 = \xi(1 - \eta) \tag{40}$$
$$N_3 = \xi\eta \tag{41}$$
$$N_4 = \eta(1 - \xi) \tag{42}$$

which are essentially the same as the shape functions for a square given by (29) to (32) with a constant $h = 1$. This leads to

$$
\mathbf{B}^T =
\begin{bmatrix}
\eta - 1 & 0 & \dfrac{\xi - 1}{\cos\theta} - (\eta - 1)\tan\theta \\[2ex]
0 & \dfrac{\xi - 1}{\cos\theta} - (\eta - 1)\tan\theta & \eta - 1 \\[2ex]
1 - \eta & 0 & -\dfrac{\xi}{\cos\theta} - (1 - \eta)\tan\theta \\[2ex]
0 & -\dfrac{\xi}{\cos\theta} - (1 - \eta)\tan\theta & 1 - \eta \\[2ex]
\eta & 0 & \dfrac{\xi}{\cos\theta} - \eta\tan\theta \\[2ex]
0 & \dfrac{\xi}{\cos\theta} - \eta\tan\theta & \eta \\[2ex]
-\eta & 0 & \dfrac{1 - \xi}{\cos\theta} - \eta\tan\theta \\[2ex]
0 & \dfrac{1 - \xi}{\cos\theta} - \eta\tan\theta & -\eta
\end{bmatrix}
\tag{43}
$$

Values of R_{crit}, now a function of θ as well as ν and c_e, are presented in Table IV. The relationships are now more complex, but for high Poisson's ratios the difference between the expressions for lumped and consistent mass is once again simply a factor of three.

Here, as with other element formulations, decreasing the shortest length (or *characteristic length*) of the element causes an increase in the maximum eigenfrequency, which in turn leads to a smaller critical time step and a larger CPR. We can see this effect in the expressions given in Table IV by noting that the height of the element is $\cos\theta$; in all cases R_{crit} increases as $\cos\theta$ decreases. As θ approaches $\pi/2$ the shortest length tends to zero, causing the maximum eigenfrequency to tend to infinity; clearly, very heavy distortion levels have the potential to cause difficulties in terms of stability. However, it is already considered good practice to avoid such severe distortions in any finite element mesh for reasons of modelling accuracy.

2D Assumption	Mass Matrix	R_{crit}
Plane stress	Lumped	$\dfrac{4\left(1+\sqrt{\sin^2\theta+\nu^2\cos^2\theta}\right)}{(1-\nu^2)\cos^3\theta}c_e^2$
Plane stress	Consistent	$\dfrac{6\left[3-\nu+(1+\nu)\sin\theta\right]}{(1-\nu^2)\cos^3\theta}c_e^2 \quad \text{for } \nu < \frac{3\sin\theta+1}{5\sin\theta+3}$
		$\dfrac{12\left(1+\sqrt{\sin^2\theta+\nu^2\cos^2\theta}\right)}{(1-\nu^2)\cos^3\theta}c_e^2 \quad \text{for } \nu \geq \frac{3\sin\theta+1}{5\sin\theta+3}$
Plane strain	Lumped	$\dfrac{4\left(1-\nu+\sqrt{(v-1)^2-(1-2\nu)\cos^2\theta}\right)}{(1+\nu)(1-2\nu)\cos^3\theta}c_e^2$
Plane strain	Consistent	$\dfrac{6(3-4\nu+\sin\theta)}{(1+\nu)(1-2\nu)\cos^3\theta}c_e^2 \quad \text{for } \nu < \frac{3\sin^2\theta-2\sin\theta-1}{8\sin^2\theta-4\sin\theta-4}$
		$\dfrac{12\left(1-\nu+\sqrt{(v-1)^2-(1-2\nu)\cos^2\theta}\right)}{(1+\nu)(1-2\nu)\cos^3\theta}c_e^2 \quad \text{for } \nu \geq \frac{3\sin^2\theta-2\sin\theta-1}{8\sin^2\theta-4\sin\theta-4}$

Table IV. Calculated R_{crit} for the distorted square element

6. NUMERICAL VERIFICATION

In order to verify the results obtained in the previous section, a series of simple finite element analyses were performed using the various element formulations. The three meshes used are shown in Figure 4. In all cases, the boundary conditions (rigid supports) are imposed using the bipenalty method with a constant stiffness penalty parameter, $\alpha_s = 10^{13}\,\text{N/m}$. The force is applied constantly throughout each analysis, vertically downwards on the top-right node of each model, and has a magnitude of $10\,\text{kN}$. The material properties are taken as $E = 10^9\,\text{N/m}^2$, $\rho = 1000\,\text{kg/m}^3$ and $\nu = 0.3$ in all cases. The central difference time integration scheme is used to compute the solutions; since the central difference method has a critical sampling frequency $\Omega_{\text{crit}} = 2$, the time step used in each case is given by $\Delta t_{\text{crit}} = 2/\omega_{\text{max}}$ where ω_{max} is the maximum eigenfrequency of a single element. The results (given in Table V) show t_{unstable}, the time at which instability is first detected; here we assume instability has occured when the maximum displacement reaches $10^{10}\,\text{m}$.
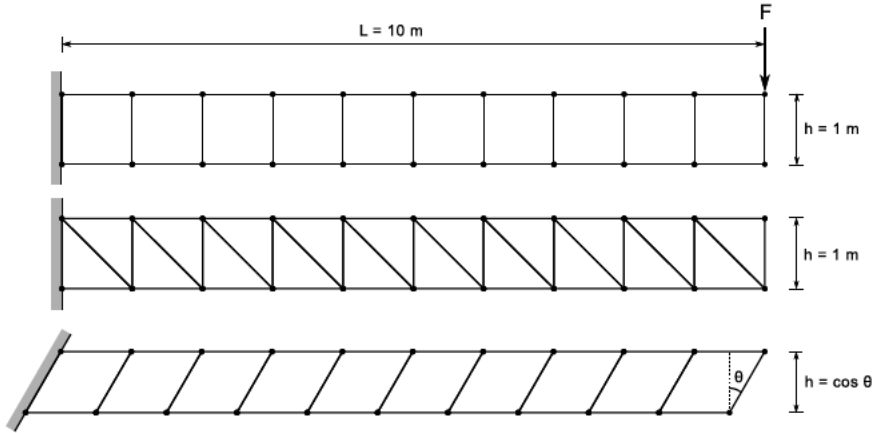

Figure 4. The three meshes used to test the square, triangle and parallelogram elements. All supported nodes have displacement fixed in both the x- and y-direction.

We see from the results that using $R = R_{\text{crit}}$ is safe in all cases; no instability is detected in the first 60 s of any analysis. Increasing the penalty ratio to $1.01 R_{\text{crit}}$ causes an instability to occur within the first second of the analysis in every case. Most analyses also show instability at $R = 1.001 R_{\text{crit}}$ and

Element type	2D	Mass matrix	t_{unstable} (s)			
			$R = 1.01R_{\text{crit}}$	$1.001R_{\text{crit}}$	$1.0001R_{\text{crit}}$	R_{crit}
Square, selective integration	Stress	Lumped	0.225898	0.712834	2.702412	-
		Consistent	0.115931	0.363251	1.190708	-
	Strain	Lumped	0.202632	0.646115	2.790697	-
		Consistent	0.103251	0.321826	1.107863	-
Right-angled triangle	Stress	Lumped	0.163204	0.508923	1.833244	-
		Consistent	0.081602	0.261625	1.842588	-
	Strain	Lumped	0.148607	0.466262	1.883280	-
		Consistent	0.073199	0.238379	-	-
Parallelogram, $\theta = \pi/12$	Stress	Lumped	0.219030	0.693978	2.845849	-
		Consistent	0.097597	0.307217	1.001001	-
	Strain	Lumped	0.196518	0.627785	3.094658	-
		Consistent	0.091155	0.283982	0.917389	-
Parallelogram, $\theta = \pi/3$	Stress	Lumped	0.091038	0.411640	-	-
		Consistent	0.038435	0.138958	-	-
	Strain	Lumped	0.082409	0.440621	-	-
		Consistent	0.034932	0.130806	-	-
Parallelogram, $\theta = 5\pi/12$	Stress	Lumped	0.040997	-	-	-
		Consistent	0.015160	-	-	-
	Strain	Lumped	0.038708	-	-	-
		Consistent	0.013985	-	-	-

Table V. Time at which instability was detected in a simple time domain analysis for various values of R. Blank entries indicate that no instability was detected within 60 s.

$R = 1.0001R_{\text{crit}}$, which demonstrates the crisp boundary that exists between stability and instability. For the most heavily distorted square element (that with $\theta = 5\pi/12$) the boundary is not as well-defined, however, using a penalty ratio of R_{crit} results in a stable analysis and hence using the calculated R_{crit} is safe, as expected.

7. SUPERIORITY OF THE BIPENALTY METHOD OVER INERTIA PENALTY METHOD

The choice of penalty ratio, to which we have given most of our attention so far, is important mainly for controlling eigenvalues in order to ensure *stability*; the *accuracy* of the bipenalty method is largely controlled by the magnitude of the penalties. However, it is simple to demonstrate the superior accuracy of the bipenalty method over the mass penalty method by means of a simple 1D problem.

Consider a bar of length $L = 100\,\text{m}$ modelled by 100 2-noded finite elements. A tensile traction of $\sigma = 1\,\text{N/m}^2$ is applied to the left-hand end of the bar, while the right-hand end is fixed using a penalty method. The material parameters are $E = 1\,\text{N}\,\text{m}^{-2}$ and $\rho = 1\,\text{kg/m}^3$ and the cross-sectional area of the bar is $A = 1\,\text{m}^2$. The mass penalty is kept constant at $\alpha_{\text{m}} = 5000\,\text{kg}$ and the calculated CPR is $R_{\text{crit}} = 4\,\text{s}^{-2}$. The central difference time integration scheme is used with a lumped mass matrix.

The displacement time history of the fixed node for three different R values is shown in Figure 5, while the acceleration is shown in Figure 6. The most accurate results (i.e., displacement and acceleration closest to zero) are achieved with both stiffness and mass penalties, and a penalty ratio

of $R = R_{\text{crit}}$, although the acceleration constraint is reasonably well imposed for all values of R. However, when using only an inertia penalty ($R = 0$), the displacement constraint is very poorly imposed when compared with simultaneous stiffness and inertia penalties. While inertia penalties indirectly constrain displacements by creating an inertia force when accelerations are non-zero, violations of the displacement constraint are able to accumulate since no stiffness penalties are present to provide a correcting force. The consequence is that for $R = 0$ the final displacement of the fixed node at time $t = 150\,\text{s}$ is several orders of magnitude greater than the displacement calculated with $R = R_{\text{crit}}$.
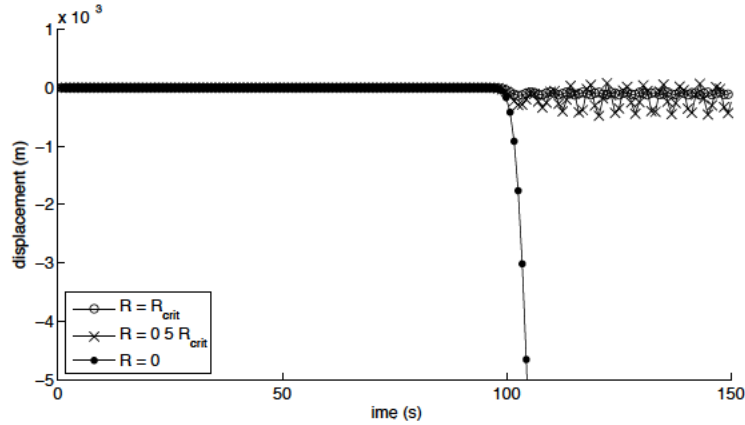

Figure 5. Displacement time history of the bipenalised node for three penalty ratios and $\alpha_{\text{m}} = 5000\,\text{kg}$. For $R = 0$ the displacement at time $t = 150\,\text{s}$ is approximately $-0.5\,\text{m}$.
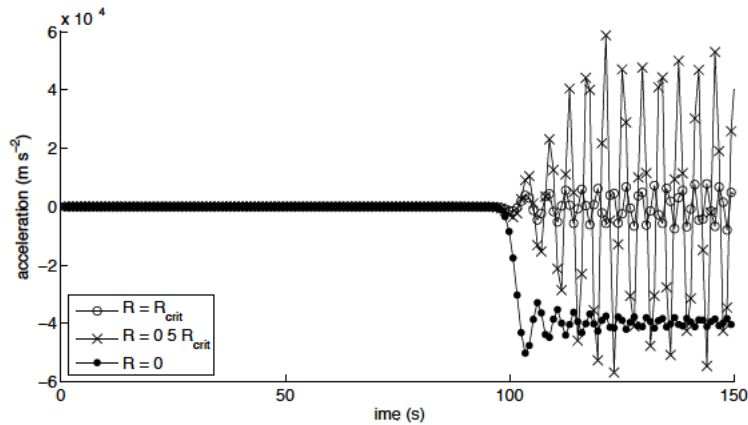

Figure 6. Acceleration time history of the bipenalised node for three penalty ratios and $\alpha_{\text{m}} = 5000\,\text{kg}$.

Looking beyond the simple absolute restraints demonstrated above, we may consider practical applications for which using mass penalties alone would be totally ineffectual. For example, when modelling contact-impact problems we may wish to model two bodies colliding, each with constant, non-zero velocities. In this case, the acceleration of both contact surfaces would be zero throughout the process. As a result, constraint violation via penetration would not be detected by a constraint condition expressed in terms of acceleration; they will have no effect unless stiffness penalties (i.e., displacement constraints) are also present to create a non-zero relative acceleration between the body surfaces as penetration is detected.

Essentially, mass penalties constrain only accelerations, not displacements. In some cases penalising accelerations is enough to ensure displacement constraints are also imposed with

satisfactory accuracy; however, a more accurate and more robust approach is to include stiffness penalties alongside mass penalties in the formulation.

8. DISCUSSION

In this contribution, we have analysed the effect of bipenalisation on the eigensolutions of an arbitrary finite element in order to provide a simple method for calculating the critical penalty ratio, R_{crit}, for any element. This allows us to derive a bipenalty method that may be used to impose absolute constraints in time domain dynamics without affecting the critical time step of the analysis. We have also demonstrated numerically the superior accuracy of this method over the use of mass penalties alone.

The methods used in Section 5 are reliant on being able to find closed form solutions for the maximum eigenfrequency of individual elements. However, determining a closed form solution for R_{crit} becomes difficult when considering elements undergoing arbitrary distortions. The elemental stiffness and mass matrices become increasingly complex to formulate and the associated eigenvalue problem becomes more difficult (or even impossible) to solve. Even if closed form solutions are found, selecting the maximum eigenfrequency is not always straightforward, since it may depend on any number of variables (for example, the Poisson's ratio of the material). In addition, as more nodes are added to the elements (e.g., 8-noded 2D quadrilaterals, 8- or 20-noded 3D brick elements) simply finding closed form expressions for the eigenfrequencies is likely to be impossible. Hence, for many problems, especially those which use non-uniform/unstructured meshes, the analytical approach described in Section 5 may not be sufficient.

In order to develop a more practical approach, we may consider common methods of calculating stable time steps. Strictly, the critical time step of an analysis is set by the maximum eigenfrequency of the entire structural system; in order to obtain the smallest possible time step that can be used without instability, this eigenfrequency must be found. While direct iteration (also known as forward iteration or power iteration) provides a relatively cheap method of finding this maximum eigenfrequency, for large systems this may still be undesirable. In practice the time step is often estimated by computing (or estimating) only the maximum eigenfrequencies of the individual elements, since this gives a safe upper bound on the critical time step of the whole system [13]. The same technique can be used when calculating an appropriate penalty ratio, the main difference being that we only need to consider the elements affected by penalties. By finding the penalised element with the highest maximum eigenfrequency and basing our CPR calculation on that element, we can ensure that the maximum eigenfrequency of that element is not increased. It is safe to use the same penalty ratio for all elements once we have found this critical element because, once bipenalised, the maximum eigenfrequencies of all constrained elements will tend to the penalty ratio that we choose, thus ensuring that no elemental eigenvalue exceeds the chosen value of $R = \alpha_{\mathrm{s}}/\alpha_{\mathrm{m}}$. If the time step calculation is based on an eigenfrequency greater than or equal to the one used to calculate our penalty ratio then time step stability is ensured.

It may also be possible to develop approximate techniques, similar to those often used in practice to estimate the critical time step of an element. In any case, the relationship $R_{\mathrm{crit}} = \omega_{\max}^2$ is likely to be central to any approximate or numerical method of estimating a safe CPR for practical use. As well as exploring practical ways of calculating a good penalty ratio for a given problem, future work will attempt to expand the theory presented here to include relative constraints. This should allow interface elements and contact formulations—two applications where penalty methods are especially popular—to be formulated using the bipenalty method.

A. EIGENSOLUTIONS OF A BIPENALISED FINITE ELEMENT

The bipenalised problem (BP) may be written as

$$\left[\left(\mathbf{K} + \mathbf{K}^{\mathrm{p}}\right) - \tilde{\omega}_i^2 \left(\mathbf{M} + \mathbf{M}^{\mathrm{p}}\right)\right] \tilde{\mathbf{u}}_i = \mathbf{0} \tag{44}$$

where $\mathbf{K}^{\mathrm{p}}$ and $\mathbf{M}^{\mathrm{p}}$ are the stiffness and mass penalty matrices, and $\tilde{\mathbf{u}}_i$ and $\tilde{\omega}_i$ form the n eigensolutions. It is also useful to split the system into the following form:

$$\begin{bmatrix} \mathbf{K}^{\mathrm{pp}} - \tilde{\omega}_i^2 \mathbf{M}^{\mathrm{pp}} + \alpha_{\mathrm{m}}(R - \tilde{\omega}_i^2)\mathbf{I}_k & \mathbf{K}^{\mathrm{pf}} - \tilde{\omega}_i^2 \mathbf{M}^{\mathrm{pf}} \\ \mathbf{K}^{\mathrm{fp}} - \tilde{\omega}_i^2 \mathbf{M}^{\mathrm{fp}} & \mathbf{K}^{\mathrm{ff}} - \tilde{\omega}_i^2 \mathbf{M}^{\mathrm{ff}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_i^{\mathrm{p}} \\ \tilde{\mathbf{u}}_i^{\mathrm{f}} \end{bmatrix} = \mathbf{0} \tag{45}$$

where $R = \alpha_{\mathrm{s}}/\alpha_{\mathrm{m}}$, $\mathbf{I}_k$ is an identity matrix of size $k \times k$ and k is the number of penalised DOFs. The vectors $\tilde{\mathbf{u}}_i^{\mathrm{p}}$ (of dimension k) and $\tilde{\mathbf{u}}_i^{\mathrm{f}}$ (of dimension $n - k$) therefore contain the solution variables for the penalised and free DOF, respectively.

The unpenalised problem (UP) is given by

$$\left(\mathbf{K} - \omega_i^2 \mathbf{M}\right) \mathbf{u}_i = \mathbf{0} \tag{46}$$

The reduced (i.e., fully constrained) problem (RP) is equivalent to the unpenalised system with the first k rows and columns removed; it can be written as

$$\left[\mathbf{K}^{\mathrm{ff}} - \left(\omega_i^{\mathrm{red}}\right)^2 \mathbf{M}^{\mathrm{ff}}\right] \mathbf{u}_i^{\mathrm{red}} = \mathbf{0} \tag{47}$$

and has $n - k$ solutions. Eigenmodes $\mathbf{u}_i^{\mathrm{red}}$ are of dimension $n - k$.

A.1. Eigenmodes of the bipenalised problem

It is useful to examine first how bipenalisation affects the eigenmodes $\tilde{\mathbf{u}}_i$ of the system. To accomplish this we note that the eigenvectors $\tilde{\mathbf{u}}_i$ of a finite element system may be scaled such that they are $\mathbf{K}$-orthogonal and $\mathbf{M}$-orthonormal [1], that is,

$$\tilde{\mathbf{u}}_i^T (\mathbf{K} + \mathbf{K}^{\mathrm{p}}) \tilde{\mathbf{u}}_j = \tilde{\omega}_i^2 \delta_{ij} \tag{48}$$

$$\tilde{\mathbf{u}}_i^T (\mathbf{M} + \mathbf{M}^{\mathrm{p}}) \tilde{\mathbf{u}}_j = \delta_{ij} \tag{49}$$

Considering the case where $i \neq j$, from (49) we have

$$\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_j + \alpha_{\mathrm{m}} \left(\tilde{\mathbf{u}}_i^{\mathrm{p}}\right)^T \tilde{\mathbf{u}}_j^{\mathrm{p}} = 0 \tag{50}$$

and therefore

$$\lim_{\alpha_{\mathrm{m}} \to \infty} \left[\left(\tilde{\mathbf{u}}_i^{\mathrm{p}}\right)^T \tilde{\mathbf{u}}_j^{\mathrm{p}}\right] = 0 \tag{51}$$

Let us consider now the set of n vectors $\left\{\tilde{\mathbf{u}}_1^{\mathrm{p}}, \tilde{\mathbf{u}}_2^{\mathrm{p}}, \ldots, \tilde{\mathbf{u}}_n^{\mathrm{p}}\right\}$. The vectors each have dimension k since they contain only the penalised DOF. Assuming that α_{m} is large, (51) gives two possibilites:

1. $\tilde{\mathbf{u}}_i^{\mathrm{p}} = \mathbf{0}$,
2. $\tilde{\mathbf{u}}_i^{\mathrm{p}}$ and $\tilde{\mathbf{u}}_j^{\mathrm{p}}$ are non-zero and orthogonal.

The first case is possible for at most $n - k$ of the n eigenmodes, since otherwise the n eigenvectors would not be linearly independent. The second case is possible for at most k of the n modes, since $\tilde{\mathbf{u}}_i^{\mathrm{p}}$ is of dimension k. Therefore, following directly from the above, we can say that

1. $\tilde{\mathbf{u}}_i^{\mathrm{p}} = \mathbf{0}$ for $n - k$ of the n eigenvectors,
2. $\tilde{\mathbf{u}}_i^{\mathrm{p}} \neq \mathbf{0}$ for k of the n eigenvectors.

A.2. The case where $\tilde{\mathbf{u}}_i^p = \mathbf{0}$

To find the eigenfrequencies that correspond to these $n - k$ eigenvectors, we may consider the last $n - k$ equations of the system given in (45). With the assumption that $\tilde{\mathbf{u}}_i^p = \mathbf{0}$ for $n - k$ of the eigenmodes, the problem reduces to

$$\left(\mathbf{K}^{\text{ff}} - \tilde{\omega}_i^2 \mathbf{M}^{\text{ff}} \right) \tilde{\mathbf{u}}_i^{\text{f}} = \mathbf{0} \tag{52}$$

an eigenvalue problem with $n - k$ eigensolutions that is equivalent to the reduced problem, (47). Therefore, the $n - k$ eigenvalues whose corresponding eigenvectors have $\tilde{\mathbf{u}}_i^p = \mathbf{0}$ are equivalent to those of the fully constrained system, i.e.,

$$\tilde{\omega}_i = \omega_i^{\text{red}} \tag{53}$$

while the eigenvectors are given by

$$\tilde{\mathbf{u}}_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{u}_i^{\text{red}} \end{bmatrix} \tag{54}$$

These results agree with the intuitive assumption that, when bipenalising a system using large penalties, $n - k$ of the eigensolutions are equivalent to the $n - k$ eigensolutions of the fully constrained system.

A.3. The case where $\tilde{\mathbf{u}}_i^p \neq \mathbf{0}$

For the case of $i = j$, the orthogonality/orthonormality conditions (48) and (49) become

$$\tilde{\mathbf{u}}_i^T \mathbf{K} \tilde{\mathbf{u}}_i + R\alpha_{\text{m}} \|\tilde{\mathbf{u}}_i^p\|^2 = \tilde{\omega}_i^2 \tag{55}$$

$$\tilde{\mathbf{u}}_i^T \mathbf{M} \tilde{\mathbf{u}}_i + \alpha_{\text{m}} \|\tilde{\mathbf{u}}_i^p\|^2 = 1 \tag{56}$$

Note that when considering these conditions we are scaling the eigenvectors $\tilde{\mathbf{u}}_i$ in a specific way. While any scalar multiple of $\tilde{\mathbf{u}}_i$ will satisfy the original eigenvalue problem (44), this is not the case for the conditions above[†].

Now, considering (56) for large α_{m} and $\|\tilde{\mathbf{u}}_i^p\| \neq 0$, we must have small $\|\tilde{\mathbf{u}}_i^p\|$ of the order α_{m}^{-1}. Assuming that $\|\tilde{\mathbf{u}}_i^f\|$ and therefore $\|\tilde{\mathbf{u}}_i\|$ are also small, the first terms of the above equations may be neglected, leaving

$$R\alpha_{\text{m}} \|\tilde{\mathbf{u}}_i^p\|^2 = \tilde{\omega}_i^2 \tag{57}$$

$$\alpha_{\text{m}} \|\tilde{\mathbf{u}}_i^p\|^2 = 1 \tag{58}$$

from which it follows that

$$\tilde{\omega}_i = \sqrt{R} = \sqrt{\frac{\alpha_{\text{s}}}{\alpha_{\text{m}}}} \tag{59}$$

A.4. Summary

For an element with k bipenalised DOF and large α_{m},

- $n - k$ of the eigenfrequencies $\tilde{\omega}_i$ are equal to those of the reduced problem ω_i^{red}. The eigenvalue separation property states that $\omega_{\text{max}}^{\text{red}} \leq \omega_{\text{max}}$ and so we can say that none of these $n - k$ eigenfrequencies will exceed ω_{max}.
- $n - k$ of the eigenvectors are given by (54); that is, the penalised DOF are zero and the remaining DOF take the values given by corresponding reduced problem.
- The remaining k eigenfrequencies are equal to $\sqrt{R}$. If we set $R \leq \omega_{\text{max}}^2$, we can be sure that none of these k eigenfrequencies will exceed ω_{max}.

A.5. Numerical example

In order to present these findings visually, a numerical experiment is carried out using a single 4-noded square finite element. Absolute constraints are imposed on an increasing number of DOF until all DOF are constrained. At each stage, the eigenfrequencies of the element are calculated numerically for increasing values of α_{m} with a constant penalty ratio, $R = 16$. The material properties are $E = 1 \text{ N/m}^2$, $\rho = 1 \text{ kg/m}^3$, $h = 1 \text{ m}$, $\nu = 0.3$, and plane stress is assumed. The results are shown in Figure 7.

[†]"In effect, the solution [of the eigenproblem] yields the direction of the eigenvector, and we use the orthonormality condition [...] to fix the magnitudes of the elements in the vector." [1, Section 2.5]

The bottom figure shows the eigenfrequencies of the UP. There are $n = 8$ eigensolutions in all; in this case there are some repeated eigenfrequencies so that $\omega_1 = \omega_2 = \omega_3, \omega_4 = \omega_5$ and $\omega_6 = \omega_7$. As constraints are added, the eigenfrequencies of each RP (indicated by the straight, vertical lines) are clearly bounded by the eigenfrequencies of the less constrained RP below it, as dictated by the eigenvalue separation property. As the penalty parameters are increased, $n - k$ of the BP eigenfrequencies tend to those of the RP; the other k eigenfrequencies tend to $\sqrt{R} = 4$.
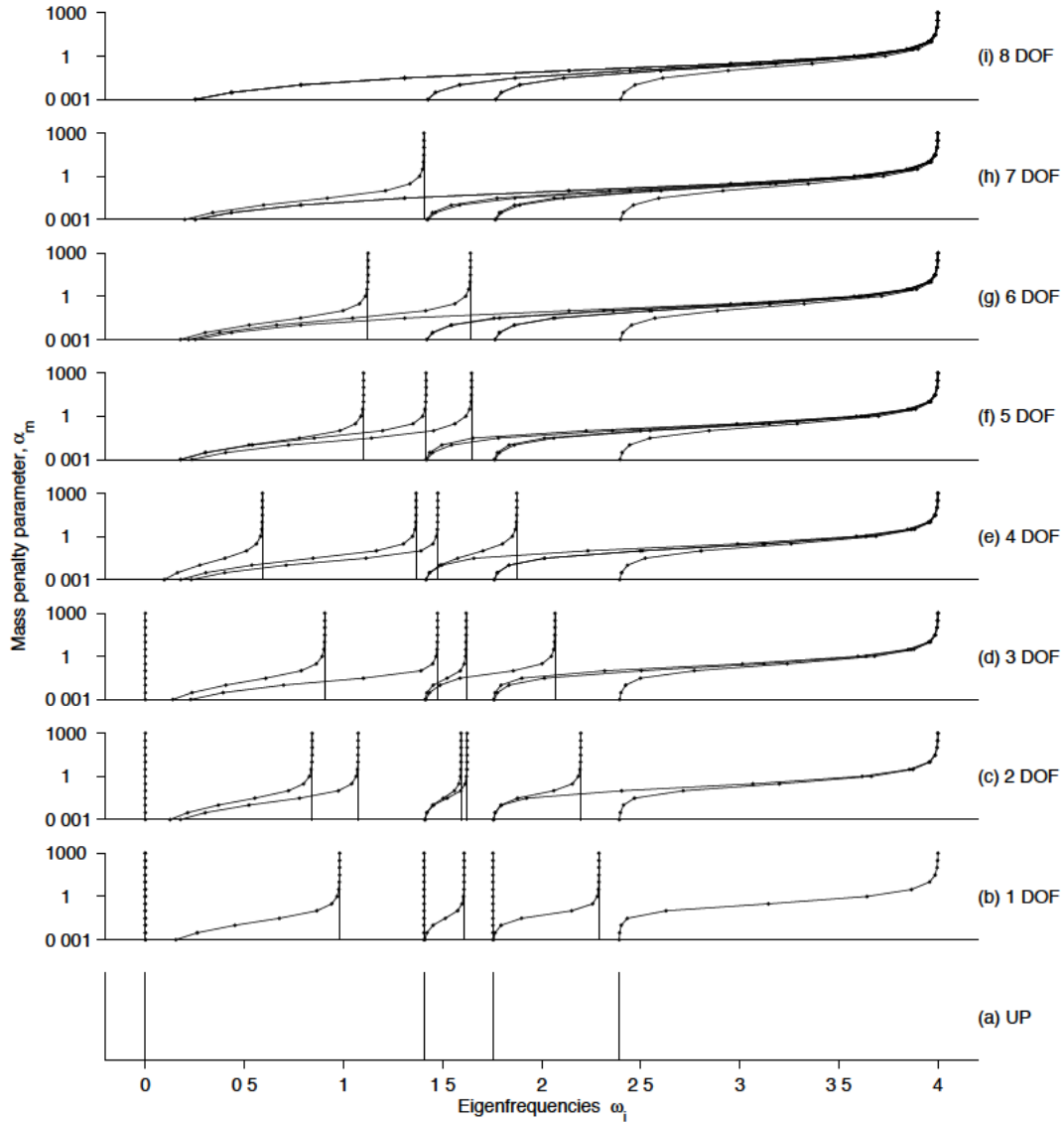

Figure 7. The numerically determined eigenfrequencies of a 4-noded quadrilateral element for increasing α_{m}. Each plot shows a different number of bipenalised ($-\bullet-$) and fully constrained ($--$) DOF. The penalty ratio is $R = 16$. Note the y-axes use a log scale.

REFERENCES

1. Bathe KJ. *Finite element procedures*. Prentice Hall: Englewood Cliffs, 1996.
2. Ilanko S. Existence of Natural Frequencies of Systems With Artificial Restraints and Their Convergence in Asymptotic Modelling. *Journal of Sound and Vibration* 2002; **255**(5):883–898, doi:10.1006/jsvi.2001.4191.
3. Paraskevopoulos EA, Panagiotopoulos CG, Manolis GD. Imposition of time-dependent boundary conditions in FEM formulations for elastodynamics: critical assessment of penalty-type methods. *Computational Mechanics* 2010; **45**(2-3):157–166.

4. Belytschko T, Neal MO. Contact-impact by the pinball algorithm with penalty and Lagrangian methods. *International Journal for Numerical Methods in Engineering* 1991; **31**(3):547–572.

5. Ilanko S. Introducing the use of positive and negative inertial functions in asymptotic modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2005; **461**(2060):2545–2562, doi: 10.1098/rspa.2005.1472.

6. Hetherington J, Askes H. Penalty methods for time domain computational dynamics based on positive and negative inertia. *Computers & Structures* 2009; **87**(23-24):1474–1482.

7. Askes H, Caramés-Saddler M, Rodríguez-Ferran A. Bipenalty method for time domain computational dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2010; **466**(2117):1389–1408.

8. Asano N. An approximate hybrid type of virtual work principle for two elastoimpact contact bodies. *Bulletin of the JSME* 1983; **26**(211).

9. Asano N. A penalty function type of virtual work principle for impact contact problems of two bodies. *Bulletin of the JSME* 1986; **29**(257).

10. Asano N. A virtual work principle using penalty function method for impact contact problems of two bodies. *Bulletin of the JSME* 1986; **29**(249).

11. Hughes TJR. *The finite element method*. Dover Publications, 2000.

12. Ling X, Cherukuri HP. Stability analysis of an explicit finite element scheme for plane wave motions in elastic solids. *Computational Mechanics* 2002; **29**(4-5):430–440.

13. Belytschko T, Smolinkski P, Liu WK. Stability of multi-time step partitioned integrators for first-order finite element systems. *Computer Methods in Applied Mechanics and Engineering* 1985; **49**(3):281–297.