

# Detecting the Causes of Ill-Conditioning in Structural Finite Element Models

Ramaseshan Kannan<sup>b,a</sup>, Stephen Hendry<sup>a</sup>, Nicholas J. Higham<sup>b,1</sup>,  
Françoise Tisseur<sup>b,2</sup>

<sup>a</sup>*Arup, 13 Fitzroy Street, London W1T 4BQ, UK. Email: stephen.hendry@arup.com*

<sup>b</sup>*School of Mathematics, The University of Manchester, Manchester, M13 9PL, England.  
Email: nick.higham@manchester.ac.uk, ramaseshan.kannan@maths.manchester.ac.uk,  
francoise.tisseur@manchester.ac.uk*

---

## Abstract

In 2011, version 8.6 of the finite element-based structural analysis package Oasys GSA was released. A new feature in this release was the estimation of the 1-norm condition number  $\kappa_1(K) = \|K\|_1 \|K^{-1}\|_1$  of the stiffness matrix  $K$  of structural models by using a 1-norm estimation algorithm of Higham and Tisseur to estimate  $\|K^{-1}\|_1$ . The condition estimate is reported as part of the information provided to engineers when they carry out linear/static analysis of models and a warning is raised if the condition number is found to be large. The inclusion of this feature prompted queries from users asking how the condition number impacted the analysis and, in cases where the software displayed an ill conditioning warning, how the ill conditioning could be “fixed”. We describe a method that we have developed and implemented in the software that enables engineers to detect sources of ill conditioning in their models and rectify them. We give the theoretical background and illustrate our discussion with real-life examples of structural models to which this tool has been applied and found useful. Typically, condition numbers of stiffness matrices reduce from  $O(10^{16})$  for erroneous models to  $O(10^8)$  or less for the corrected model.

---

<sup>1</sup>The work of this author was supported by European Research Council Advanced Grant MATFUN (267526)

<sup>2</sup>The work of this author was supported by EPSRC grant EP/I005293/1.

## 1. Introduction

Modelling structural response using the finite element (FE) method involves idealizing structural behaviour and dividing the structure into elements. The solution obtained from such a procedure is therefore inherently approximate. The errors are of a variety of types, such as error in the choice of mathematical model (when the physical system being modelled does not obey the assumptions of the mathematical model chosen), discretization errors (arising from representing infinite dimensional operators in finite spaces), and numerical errors (those caused by representing real numbers as finite precision numbers on a computer). Techniques for understanding, analysing and bounding these errors have developed in pace with the method itself and their study is part of any standard text on finite element analysis—see [1] or [2, chap. 18] for example. User errors (errors in the definition of the model within the software from say, erroneous input), on the other hand, have received significantly less attention in the literature. This is partly due to the practitioner using FE analyses being disconnected with the process of developing it or implementing it in software but mainly because such errors can arise arbitrarily, which poses a barrier to understanding and analysing them. Examples of such errors include

- Lack of connectivity: adjacent elements that are supposed to share a common node but are connected to different nodes that are coincident, resulting in one of the elements acquiring insufficient restraints.
- Failure to idealise member end connections correctly, which can occur if a beam is free to rotate about its axis, although in the physical model there is a nominal restraint against rotation.
- Modelling beam elements with large sections and/or very small lengths, often the result of importing FE assemblies from CAD models.

Irrespective of whether the error arose from approximation or from erroneous input data, it can lead to an ill-conditioned problem during its analysis, that is, one for which the stiffness matrix of the model has a large condition number with respect to inversion. In this work, we show how errors that lead to an ill-conditioned problem can be detected, i.e., we present a technique to identify the parts of the model that cause the ill conditioning. Our

method has been implemented in the commercial structural FE analysis package Oasys GSA [3] and we demonstrate its efficacy with examples of how it has been used to identify errors in user-generated ill-conditioned models.

## 2. Condition number estimation

Linear systems of equations arise in many problems in structural analysis. For a structural model with symmetric positive definite (or semidefinite) stiffness matrix  $K \in \mathbb{R}^{n \times n}$ , elastic static analysis for calculating the displacement  $u \in \mathbb{R}^n$  under the action of loads  $f \in \mathbb{R}^n$  yields the system  $Ku = f$ . Other types of analysis that involve solving linear equations with the stiffness matrix include dynamic, buckling, P-Delta, and nonlinear static analysis. Solving a linear system on a computer involves approximating the entries of  $K$  as floating point numbers, which introduces an error  $\Delta K$  (we disregard, without loss of generality, the error introduced in  $f$  due to the same process). Denoting the corresponding change to  $u$  by  $\Delta u$ , the equation we solve is

$$(K + \Delta K)(u + \Delta u) = f.$$

Rearranging, taking norms and dropping the second order term  $\Delta K \Delta u$  gives the inequality, correct to first order,

$$\frac{\|\Delta u\|}{\|u\|} \leq \kappa(K) \frac{\|\Delta K\|}{\|K\|}, \quad (1)$$

where  $\kappa(K) = \|K\| \|K^{-1}\|$  is the condition number (with respect to inversion). The norm  $\|\cdot\|$  can be any subordinate matrix norm, defined in terms of an underlying vector norm by  $\|K\| = \max_{\|x\|=1} \|Kx\|$ .

Condition numbers measure the maximum change of the solution to a problem with respect to small changes in the problem. Inequality (1) tells us that the relative error in  $u$  is bounded by the relative error in  $K$  times its condition number. This bound is attainable to first order for a given  $K$  and  $f$  [4, Thm. 7.2], so the change in the solution caused simply by storing  $K$  on the computer can be large if  $K$  is ill conditioned. Rounding errors in the solution process can be shown to correspond to an increased  $\|\Delta K\|$  in (1) and so are also magnified by as much as  $\kappa(K)$ .

In IEEE 754 double precision binary floating point arithmetic [5], we have a maximum of the equivalent of 16 significant decimal digits of precision available and we therefore have as little as  $16 - \log_{10} \kappa(K)$  digits of accuracy in

the computed solution. When a matrix has a condition number greater than  $10^{16}$ , the solution algorithm can return results with no accuracy at all—such a matrix is numerically singular and linear systems with this matrix should not be solved. Therefore it is essential to compute or estimate the condition number of the stiffness matrix  $K$  to ensure it is well conditioned.

Since  $K$  is symmetric, its 2-norm condition number  $\kappa_2(K)$  is the ratio of its extremal eigenvalues  $\lambda_{\max} = \max_i \lambda_i$  and  $\lambda_{\min} = \min_i \lambda_i$  [4, Chap. 6]:

$$\kappa_2(K) = \left| \frac{\lambda_{\max}}{\lambda_{\min}} \right|.$$

Computing the maximum and minimum eigenvalues is an expensive operation, particularly since  $K$  is large. In practice we need just an estimate of the condition number that is of the correct order of magnitude, and we can choose any convenient norm to work with [4, chap. 15]. LAPACK [6] offers the `xLACON` routine that computes a lower bound for the 1-norm of a matrix, based on the algorithm of Higham [7]. Higham and Tisseur [8] develop a block generalization of this algorithm and demonstrate that it produces estimates accurate to one or more significant digits at a nominal cost of a few matrix–vector multiplications. The algorithm does not need to access the elements of the matrix explicitly, as long as it can access a routine that returns the matrix–vector product. Thus it can be used to estimate the norm of the inverse of a matrix  $K$  as long as one can form the product  $K^{-1}v =: g$ , which is equivalent to solving the linear system  $Kg = v$ .

We incorporated a procedure in GSA that uses this algorithm to calculate an estimate of the 1-norm condition number of  $K$ . This method is invoked and the condition number is reported for all analysis types that involve assembling  $K$ . It first computes  $\|K\|_1$  and then computes the estimate of  $\|K^{-1}\|_1$ , making use of the Cholesky factors of  $K$ . Since the analysis already requires the Cholesky factorization of  $K$  to be computed, the condition estimate comes at a nominal cost of a few triangular system solves.

Once the software had been released, users were informed when they had a model with an ill-conditioned stiffness matrix. As a result, the developers were subsequently faced with the question: “*how do we detect where in the FE model the ill conditioning lies?*”. To answer this question we first need to look at the possible causes of ill conditioning.

### 3. Ill conditioning in structural stiffness matrices

Let the element stiffness matrix contributed by each element  $e$  in the domain  $\Omega$  be  $K^{(e)}$ . The matrix  $K^{(e)} \in \mathbb{R}^{n_e \times n_e}$  is symmetric and it is defined in a coordinate system (of displacements) and degrees of freedom (dofs) local to the element. To map the displacements to the global coordinate system we use the coordinate transformation matrix  $T^{(e)}$ ; we also use a set  $m_e$  of  $n_e$  variables that map the locally numbered dofs to the global dof numbering. The stiffness matrix  $K$  is the sum of contributions from all elements in the domain  $\Omega$  transformed to global:

$$K = \sum_{e \in \Omega} G^{(e)}, \quad (2)$$

where

$$G^{(e)}(m_e, m_e) = T^{(e)T} K^{(e)} T^{(e)}.$$

If an element connects dof  $i$  with dof  $j$ , then  $k_{ii}, k_{jj}$ , and  $k_{ij} = k_{ji}$  are all nonzero, assuming an element with nonzero stiffness in directions of  $i$  and  $j$  connects the dofs. Because of its construction  $K$  is symmetric, with very few nonzero entries per row.

The matrix  $K$  becomes ill conditioned when its columns are nearly linearly dependent. This can happen when

- (a) the structure has one or more pairs or tuples of dofs that do not have sufficient connectivity with the rest of the model or
- (b) certain dofs have stiffnesses disproportionate with the rest of the model.

We say a pair  $(i, j)$  of dofs connected to each other has insufficient connectivity when the stiffness contributions of terms  $k_{is}$  and  $k_{rj}$  for  $r, s \in (1, n)$  with  $r, s \notin (i, j)$  are either very small or at roundoff level compared with  $k_{ii}$ ,  $k_{jj}$ , and  $k_{ij}$ . (The definition can be easily expanded for higher tuples of dofs.)

Possibility (a) occurs when elements do not have sufficient connectivity, for example a beam element that is connected to nodes at which there is no torsional restraint. Typically the resultant matrix would be singular since the structure is a mechanism, but it is possible that due to rounding during coordinate transformations, entries in columns  $k_i$  or  $k_j$  acquire small nonzero values. If  $i$  and  $j$  are the dofs corresponding to axial rotation at the two ends of the column member, such a model would result in a matrix with columns  $i$  and  $j > i$  resembling (with  $k_i$  denoting the  $i$ th column of  $K$ )

$$k_i = [0, \dots, 0, a, 0, \dots, 0, -a, 0, \dots]^T$$

and

$$k_j = [0, \dots, \epsilon, -a, \epsilon, \dots, \epsilon, a, 0, \dots]^T,$$

where the entries  $k_{ii} = k_{jj} = a > 0$  and  $k_{ij} = k_{ji} = -a$  and all other entries  $\epsilon$  in  $k_j$  arising from other dofs connected to dof  $j$  are such that  $|\epsilon| \ll a$ . More generally, a tuple  $\mathcal{S} \subseteq \{1, \dots, n\}$  of dofs can arise such that for  $i \in \mathcal{S}$ ,  $k_{ij}$  is nonzero only for  $j \in \mathcal{S}$ .

The situation that would result in (b) is when certain elements that are disproportionately stiff in particular directions connect with more flexible elements in the neighbourhood. This results in a badly scaled matrix, and can be seen, for example, when beam or column members are split in numerous line elements—usually the result of importing a CAD drawing as an FE assembly.

It is impossible to obtain a full list of modelling scenarios that will result in an ill-conditioned stiffness matrix, but in section 5 we present a few examples of real life models we have encountered. For now, we focus on how the properties of the matrix can be exploited to identify the location of the anomalies that cause ill conditioning. By *location of anomalies*, we mean the identification of the errant dofs  $\mathcal{S}$ , and subsequently problematic elements, such that an examination of the model defined in the vicinity of these elements can help the user identify the issue.

Past work has focused on identifying mechanisms in finite element models. Mechanisms render the stiffness matrix singular, so the problem is the same as finding the null space of the matrix, though in the case of floating structures the matrix can also have a nontrivial null space corresponding to its rigid body modes. Farhat and G eradin [9] and Papadrakakis and Fragakis [10] deal with the computation of the null space of stiffness matrices using a combination of algebraic and geometric information specific to the discretization, whereas Shklarski and Toledo [11] use a graph theoretic approach for computing the null space.

Whilst the identification of mechanisms is useful, rectifying the error is a case of fixing the unconstrained dof. Moreover, since a mechanism results in a singular stiffness matrix, an attempt to find its Cholesky factors (during, say, linear static analysis) is likely to break down and hence signal the problem [4, Chap. 10]. An ill-conditioned matrix, however, poses more challenges. Ill conditioning can result from subtler errors that might be hard to detect, but their presence can lead to numerical inaccuracies in results.

Our technique for diagnosing user errors works for both ill conditioning

errors as well as mechanisms. It provides the user with qualitative information about the location of these errors in the structure being modelled. The method uses eigenvectors of the stiffness matrix, so it does not need new algorithms to be deployed in software packages but rather can make use of existing well-known techniques for solving eigenvalue problems.

#### 4. Using eigenvectors to identify cause of ill conditioning

We now describe our method to identify elements and dofs that cause ill conditioning in the stiffness matrix. The key insight is that the eigenvectors corresponding to extremal eigenvalues of  $K$  (which, with a slight abuse of notation, we will refer to as the *smallest/largest eigenvectors*) contain rich information about the dofs that cause ill conditioning. We assume the model does not have rigid body modes, i.e.,  $K$  does not have a nontrivial null space corresponding to rigid body motion. If the ill conditioning of  $K$  is due to the reasons outlined in section 3, the smallest and/or the largest eigenvectors are *numerically sparse*. We call a normalized vector numerically sparse when it has only a small number of components significantly above the roundoff level. In the remainder of this section, we show that when ill conditioning is caused by insufficient connectivity the eigenvectors corresponding to the one or more smallest eigenvalues are numerically sparse, whereas when the ill conditioning is from the presence of elements with disproportionately large stiffnesses the largest eigenvectors exhibit numerical sparsity. If we define the inner product terms

$$v^{(e)} = \frac{1}{2} u_i(m_e)^T u_i(m_e), \quad e \in \Omega \quad (3)$$

and

$$s^{(e)} = \frac{1}{2} u_i(m_e)^T T^{(e)T} K^{(e)} T^{(e)} u_i(m_e), \quad e \in \Omega \quad (4)$$

for a normalized eigenvector  $u_i$  and element  $e$ , then the elements that cause ill conditioning are those that have large relative values of either  $v^{(e)}$  for the smallest eigenvectors or  $s^{(e)}$  for the largest eigenvectors. The element-wise scalars  $v^{(e)}$  and  $s^{(e)}$ , respectively, can be thought of as virtual kinetic and virtual strain energies associated with the modes of displacements defined by the eigenvectors; therefore we refer to them as virtual energies later on in the text.

Our method for identifying ill conditioning in a finite element assembly is outlined in Algorithm 1. We call it *model stability analysis*, since it analyses the numerical stability of the underlying model.

---

**Algorithm 1** Algorithm for model stability analysis.

---

This algorithm has the following user-defined parameters:

- $n_s \geq 0$ : the number of smallest eigenpairs;
- $n_\ell \geq 0$ : the number of largest eigenpairs;
- $\tau > 1$ : the condition number threshold for triggering analysis;
- $\text{gf} \geq 1$ : the order of the gap between a cluster of smallest eigenvalues and the next largest eigenvalue.

1. Compute a condition number estimate  $\text{est} \approx \kappa_1(K)$ .
2. If  $\text{est} < \tau$ , exit.
3. Issue ill conditioning warning.
4. Compute the  $n_s$  smallest eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_{n_s}$  and  $n_\ell$  largest eigenvalues  $\lambda_{n-n_\ell+1}, \dots, \lambda_n$  of  $K$  and normalize the associated eigenvectors.
5. With the smallest eigenpairs: determine if a gap exists, i.e., if there is a  $k < n_s$  such that

$$\frac{\lambda_{k-1}}{\lambda_k} > \text{gf} \times \frac{\lambda_k}{\lambda_{k+1}}$$

If no such  $k$  is found go to step 7.

6. For each eigenvector  $u_i$ ,  $i = 1$  to  $k$ , calculate  $v^{(e)}$  for all elements.
  7. With the largest eigenpairs: for each eigenvector  $u_i$ ,  $i = n - n_\ell$  to  $n$ , compute  $s^{(e)}$ .
- 

Once the algorithm finishes executing, the user must find elements with large virtual energies for each eigenpair. Isolating elements with large relative values of  $v^{(e)}$  or  $s^{(e)}$  is based on visual inspection of the values. This is done by graphically colour-contouring the scalars on elements as discs whose radii and colour depend on the relative values of the scalars. Once the elements with large relative energies are identified the user must examine their definition (e.g., support conditions, nodal connectivity or section properties)

for anomalies and fix discrepancies. This should result in  $\kappa(K)$  decreasing, hence the proposed workflow is iterative: execute Algorithm 1, fix anomalies, check if  $\text{est} < \tau$ , and repeat the steps if necessary.

We now explain the reasoning behind our method by relating the ill conditioning to properties of the largest and smallest eigenvectors.

#### 4.1. Smallest eigenpairs

In the case of a few insufficiently connected dofs, the sparsity of eigenvectors stems from their continuity over perturbations. An ill conditioned matrix  $K$  is a small relative distance from a singular matrix  $\widehat{K}$ , as shown by the following result of Gastinel and Kahan [4, Thm. 6.5].

**Theorem 4.1.** *For any  $n \times n$  matrix  $A$  the distance*

$$\text{dist}(A) = \min \left\{ \frac{\|\Delta A\|}{\|A\|} : A + \Delta A \text{ singular} \right\}$$

is given by

$$\text{dist}(A) = \kappa(A)^{-1},$$

where the norm is any subordinate matrix norm.

For an ill conditioned  $K$ , the eigenvectors of  $K$  corresponding to the smallest eigenvalues are the perturbed null vectors of a nearby singular matrix  $\widehat{K}$ , and since the null vectors of  $\widehat{K}$  can reveal unconstrained dofs, as shown by Lemma 4.2 below, the smallest eigenvectors of  $K$  contain the same information as long as the perturbation is small.

We first show that the null vector of a singular matrix with a specific property has a predefined structure. We then demonstrate how a small perturbation to this singular matrix does not change the structure of the null vector. The proof is then generalized for an invariant subspace of a matrix and we show that under mild assumptions the eigenvectors of an ill conditioned matrix reveal dofs that are insufficiently restrained.

We start with the following lemma. All norms used throughout are 2-norms.

**Lemma 4.2.** *Let  $A = [a_1, \dots, a_n] \in \mathbb{R}^{n \times n}$  have columns  $a_1, \dots, a_{n-1}$  linearly independent and  $a_{n-1}$  and  $a_n$  dependent. Then*

$$\text{Null}(A) = \{x \in \mathbb{R}^n : Ax = 0\} = \text{span}\{u\},$$

where  $u$  has the form

$$u = [0, \dots, 0, \alpha, \beta]^T \quad (5)$$

for  $\alpha, \beta \in \mathbb{R}$  with  $\beta \neq 0$ .

*Proof.* By assumption,  $\text{rank}(A) = n - 1$  and hence  $\dim \text{null}(A) = 1$ .

Write  $a_n = \gamma a_{n-1}$  (this is always possible since  $a_{n-1} \neq 0$ ). Then

$$Au = 0 \iff u_1 a_1 + u_2 a_2 + \dots + u_{n-2} a_{n-2} + (u_{n-1} + \gamma u_n) a_{n-1} = 0,$$

and since  $a_1, \dots, a_{n-1}$  are linearly independent,  $u_1 = \dots = u_{n-2} = 0$  and  $u_{n-1} = -\gamma u_n$ . So  $u \neq 0$  if  $u_n \neq 0$  and in that case  $u$  has the required form.  $\square$

The matrix  $A$  in Lemma 4.2 has a simple eigenvalue 0 and the null vector  $u$  is a corresponding eigenvector. Now suppose  $A$  is, additionally, symmetric and consider a perturbed matrix  $\tilde{A} = A + E$  for a symmetric perturbation  $E$ . The next result shows that if  $\tilde{u}$  is the smallest eigenvector of  $\tilde{A}$  then for small perturbations the structure of  $\tilde{u}$  is similar to that of the null vector  $u$ .

**Theorem 4.3.** [12, Thm. 8.1.12] *Suppose  $A$  and  $A + E$  are  $n \times n$  symmetric matrices and that*

$$U = [u_1 \ U_2]$$

*is an orthogonal matrix such that  $u_1 \in \mathbb{R}^n$  is an eigenvector for  $A$ . Partition the matrices  $U^T A U$  and  $U^T E U$  as*

$$U^T A U = \begin{bmatrix} \lambda & 0 \\ 0 & D_2 \end{bmatrix}, \quad U^T E U = \begin{bmatrix} \epsilon & e^T \\ e & E_{22} \end{bmatrix},$$

*where  $D_2$  and  $E_{22}$  are  $(n - 1) \times (n - 1)$ . If  $d = \min_{\mu \in \lambda(D_2)} |\lambda - \mu| > 0$  and  $\|E\|_2 \leq d/5$ , then there exists  $p \in \mathbb{R}^{n-1}$  satisfying*

$$\|p\|_2 \leq \frac{4}{d} \|e\|_2$$

*such that  $\tilde{u}_1 = (u_1 + U_2 p) / \sqrt{1 + p^T p}$  is a unit 2-norm eigenvector for  $A + E$ . Moreover,*

$$\text{dist}(\text{span}\{u_1\}, \text{span}\{\tilde{u}_1\}) = \sqrt{1 - (u_1^T \tilde{u}_1)^2} \leq \frac{4}{d} \|e\|_2.$$

We apply the result above to a symmetric positive semidefinite  $K$  with a simple eigenvalue 0 that is perturbed by a symmetric  $E$  such that  $\tilde{K} = K + E$  is positive definite. If  $(0, u)$  is the smallest eigenpair of  $K$ , with  $\lambda_2$  being the next largest eigenvalue, then the perturbed eigenvector  $\tilde{u}$  differs from the original null vector by a distance proportional to the product of the norm of the perturbation and the reciprocal of  $d = \lambda_2$ . In other words, if the null vector  $u$  has the structure defined in (5), its perturbation  $\tilde{u}$  has the form

$$\tilde{u}^T = [\delta_1, \dots, \delta_{n-1}, \alpha + \delta_{n-1}, \beta + \delta_n],$$

with  $\delta_i$  of order  $\|e\|_2/\lambda_2$ . Therefore, the smallest eigenvector of an ill conditioned stiffness matrix will have large components corresponding to the nearly dependent columns. It is easy to show that elements that share these dofs have large values of the inner product/virtual energy  $v^{(e)}$  as defined in (3). Let  $\mathcal{D}$  be the set of dependent dofs and let element  $e$  have a mapping  $m_e$  that has a larger intersection with  $\mathcal{D}$  than the mapping set of any other element in the assembly, i.e.,

$$|\mathcal{D} \cap m_e| > |\mathcal{D} \cap m_i| \quad \forall i \in \Omega, \quad i \neq e,$$

where  $|\cdot|$  is the cardinality of the set. Then,  $\tilde{u}(m_e)^T \tilde{u}(m_e) > \tilde{u}(m_i)^T \tilde{u}(m_i)$  for  $i \in \Omega, i \neq e$ .

We applied the bound in Theorem 4.3 to the special case of  $K$  with two columns dependent, resulting in a simple eigenvalue 0. More generally, there could be  $p \geq 2$  columns that are dependent, corresponding to  $p$  dofs that are badly restrained. If  $K$  is scaled such that the largest eigenvalue is 1, then it has a cluster of eigenvalues close to 0 and these eigenvalues have eigenvectors with few large components. For eigenvalues that are clustered, the eigenvectors are sensitive to perturbations but the invariant subspace corresponding to the cluster is less sensitive. Our result is therefore easily generalized for a cluster of eigenvalues and the invariant subspace of the associated eigenvectors: the clustered eigenvalues of an ill conditioned symmetric matrix are a perturbation of repeated zero eigenvalues of a nearby singular matrix and the eigenvectors form a basis for the subspace that is in the neighbourhood of the null space. The following theorem, which is a special case of [12, Thm. 8.1.10], states this result. Here, we need the Frobenius norm,  $\|A\|_F = (\sum_{i,j=1}^n |a_{ij}|^2)^{1/2}$ .

**Theorem 4.4.** *Let a symmetric positive semidefinite matrix  $A \in \mathbb{R}^{n \times n}$  have the eigendecomposition*

$$A = U^T \Lambda U,$$

where  $\Lambda = \text{diag}(\lambda_i)$  with  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  and  $U$  is a matrix of eigenvectors. Assume the spectrum of  $A$  is such that the first  $r$  eigenvalues in  $\Lambda$  are 0 and  $\lambda_{r+1} > 0$ , and partition  $U = [U_1 \ U_2]$ , so that the columns of  $U_1 \in \mathbb{R}^{n \times r}$  span the null space of  $A$ . Let  $E \in \mathbb{R}^{n \times n}$  be a symmetric matrix and partition  $U^T E U$  conformably with  $U$  as

$$U^T E U = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}.$$

If

$$\|E\|_F \leq \frac{\lambda_{r+1}}{5},$$

then there exists a matrix  $P \in \mathbb{R}^{(n-r) \times r}$  with

$$\|P\|_F \leq 4 \frac{\|E_{21}\|_F}{\lambda_{r+1}}$$

such that the columns of  $\tilde{U}_1 = (U_1 + U_2 P)(I + P^T P)^{-\frac{1}{2}}$  form an orthonormal basis for a subspace invariant for  $A + E$ .

Theorem 4.4 suggests that where the ill conditioning is associated with a cluster of small eigenvalues we need to examine all the eigenvectors associated with the cluster to gain an understanding of the causes of ill conditioning. The identification of the gap is important and so is its size—a large gap ensures that the perturbed eigenvectors are “close” to the null vectors, and hence contain useful information about the errors.

#### 4.2. Largest eigenpairs

When  $K$  is ill conditioned because a few elements possess large stiffnesses in comparison with other elements in the model, the largest eigenvectors are numerically sparse and we can use the element-wise virtual strain energy  $s^{(e)}$  from (4) to identify such elements. The relationship between  $\lambda_{\max}(K)$  and  $\max_e \{\lambda_{\max}(K^{(e)})\}$  is already well known from the results of Wathen [13] and Fried [14]. Here, we show the connection between the largest eigenvector of  $K$  and the largest eigenvector of  $K^{(e)}$  and use the relationship to demonstrate

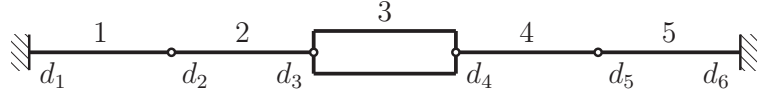


Figure 1: A sub-assembly. Element 3 is the stiffest, with a large difference in stiffness.

that the stiffest element has the largest virtual strain energy. We start with a localized sub-assembly of elements and then extend the argument for the entire model.

Consider an FE sub-assembly as in Figure 1, with a stiffness matrix  $S$ . We assume, without loss of generality, that each element in the sub-assembly has stiffness only in one direction at each end, so there is only one dof at nodes  $d_1, \dots, d_6$ , and that no other coupling exists between these nodes and the rest of the model. Let the transformed stiffness matrix of element  $r$  in the direction of the dofs at the nodes be  $k_r \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ . Since the ill conditioning scenario we are interested in is one where a few elements have much larger stiffnesses than most other neighbouring elements (that in turn have stiffnesses comparable with each other), let  $k = k_1 = k_2 = k_4 = k_5$  and let element 3 be the stiffest, with  $k_3 = \mu k$  for  $\mu \gg 1$ . Then the stiffness matrix  $S$  of the sub-assembly is

$$S = k \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & \mu + 1 & -\mu & 0 \\ 0 & -\mu & \mu + 1 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}. \quad (6)$$

We note here that the simplification of using elements with only one dof at each node is only to keep the dimension of  $S$  low and it does not affect the discussion: we can use any element as long as it has at least one dof about which it has a large relative stiffness. We are interested in the largest eigenvector and the corresponding energies of elements when the assembly deforms in that mode.

Using the MATLAB Symbolic Math Toolbox the largest eigenvector of  $S$  is found to be (unnormalized)

$$u = [-1, \gamma, -\gamma, 1]^T, \text{ where } \gamma = \mu + (\sqrt{4\mu^2 - 4\mu + 5} - 1)/2.$$

Table 1 provides expressions for  $s^{(e)}$  when the sub-assembly is in a displacement mode given by  $u$ . Clearly,  $s^{(3)} > s^{(e)}$  for  $e = 1, 2, 4, 5$  for all  $\mu > 1$ .

Table 1:  $s^{(e)}$  for elements  $e$  in the sub-assembly.

| Element $e$ | 1 | 2              | 3           | 4              | 5 |
|-------------|---|----------------|-------------|----------------|---|
| $s^{(e)}$   | 1 | $1 + \gamma^2$ | $4\gamma^2$ | $1 + \gamma^2$ | 1 |

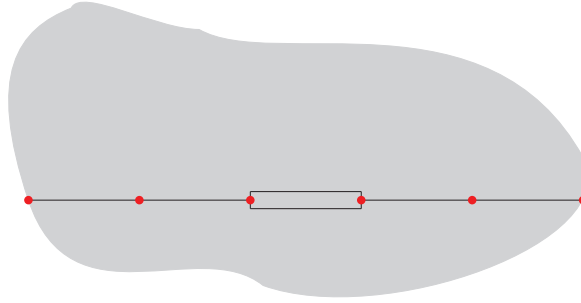


Figure 2: An arbitrary FE model.

Hence the stiffest element has the largest virtual energy. We also observe that the components of the eigenvector vary as polynomials of the ratio of stiffnesses, therefore the larger the variation in stiffness magnitudes, the larger the difference in the order of the components of the vector, making it more numerically sparse.

To generalize this observation, we show that when such an assembly is part of a larger structure, the largest eigenvector of the larger model has a form similar to that of the largest eigenvector of the sub-assembly.

Assume the structure in (1) is embedded in an arbitrary FE model as in Figure 2, with stiffnesses at nodes  $d_3$  and  $d_4$  larger than those at any other dof. The fixities at the ends of elements 1 and 5 in Figure 1 become shared nodes between the sub-assembly and the rest of the model. We represent the stiffness matrix of the entire structure as  $K \in \mathbb{R}^{n \times n}$ . Let  $b$  be the maximum number of nonzeros in any row of  $K$ . The quantity  $b$  represents the element connectivity in the structure.

We can order  $K$  as

$$K = \begin{bmatrix} M & F \\ F^T & S \end{bmatrix}, \quad (7)$$

where  $S$  is the same as in (6) and represents the sub-assembly and  $M \in \mathbb{R}^{(n-4) \times (n-4)}$  contains dofs from the rest of the model. The stiffness terms

for shared dofs are in  $F \in \mathbb{R}^{(n-4) \times 4}$ , which is of low rank and small norm compared with  $S$  (and can be written out explicitly). Then, under mild assumptions, the largest eigenvector of the block-diagonal matrix

$$\hat{K} = \begin{bmatrix} M & 0 \\ 0 & S \end{bmatrix} \quad (8)$$

is of the form  $\hat{v}^T := [0, \dots, 0, u] \in \mathbb{R}^n$ , where  $u$  is the largest eigenvector of  $S$ . To show this we need to show that  $\|M\|_2 < \|S\|_2$ . If  $K$  is scaled using its construction in (2) such that

$$\max_{\substack{e \in \Omega \\ e \neq 2}} \|K^{(e)}\| = 1,$$

we have

$$\|K^{(3)}\| \gg 1,$$

since element 3 is the stiffest. Then informally,  $\max_{i,j} |m_{ij}| = O(1)$  and  $\max_{i,j} |s_{ij}| = O(\mu)$ . Using [4, Prob. 6.14] and the definition of  $b$  and the symmetry of  $M$ , we have

$$\begin{aligned} \|M\|_2 &\leq b^{1/2} \max_j \|M(:, j)\|_2 \leq b^{1/2} \cdot b^{1/2} \max_{i,j} |m_{ij}| \\ &= b \max_{i,j} |m_{ij}| = O(b) < O(\mu) = \max_{i,j} |s_{ij}| \leq \|S\|_2, \end{aligned}$$

assuming  $b \ll \mu$ . This assumption is reasonable since the number of nonzeros is typically of a smaller order compared with the stiffness of the element causing the ill conditioning and conforms with our observations. We also note that  $b$  is akin to the factor  $p_{\max}$  in [14, Lemma 2] or  $d_{\max}$  in [13, Equation 3.7]. The matrix

$$K = \hat{K} + \begin{bmatrix} 0 & F \\ F^T & 0 \end{bmatrix},$$

is a small perturbation of  $\hat{K}$  (of norm  $\epsilon$ , say). Therefore, using Theorem 4.3, the largest eigenvector  $v$  of  $K$  has a small distance (proportional to  $\epsilon$ ) from  $\hat{v}$  and so retains the structure. This implies  $v$  is numerically sparse, leading to large relative values of virtual strain energies for elements that cause ill conditioning of the stiffness matrix.

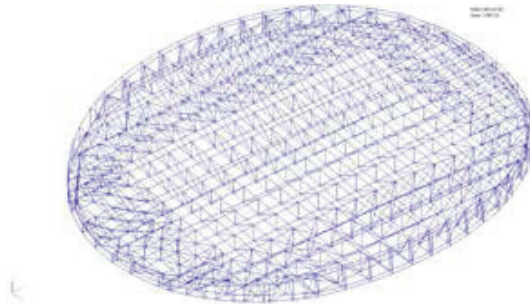


Figure 3: 3D view of the arena roof-truss model. The lines in blue represent truss elements.

## 5. Examples

The method described in section 4 has been implemented as an analysis option in the software package Oasys GSA [3]. Eigenvalues and eigenvectors are computed using subspace iteration with deflation of converged vectors. The following examples illustrate how the feature has been used on structural models created by engineers on real-life projects. Since the illustrations make use of models created in GSA, we use software-specific definitions of common entities and concepts encountered in FE modelling and define new terms when we introduce them. In all but the last example, examining element virtual energies for the small eigenvalues revealed the causes of the ill conditioning.

### 5.1. Roof truss for performing arts arena

Figure 3 shows the roof truss model of a performing arts arena that was analysed in GSA. The roof is modelled as ‘beam’ and ‘bar’ elements, which are both one dimensional. Beam elements have 6 dofs at each end, 3 each for translation and rotation, whereas bars have two dofs corresponding to axial extensions only. The main trusses run along the shorter diameter of the oval, with longitudinal bracing connecting each truss at the top and bottom chord to provide lateral stability. The model is supported on pins and rollers at joints close to the circumference.

On an initial version of the model, the condition number was estimated to be  $O(10^{17})$ , meaning that the stiffness matrix was numerically singular. Model stability analysis was executed and the smallest eigenvalues of the matrix are  $\lambda_1 = 0.0$ ,  $\lambda_2 = 1.142 \times 10^{-13}$ , and  $\lambda_3 = 1.396 \times 10^5$ .

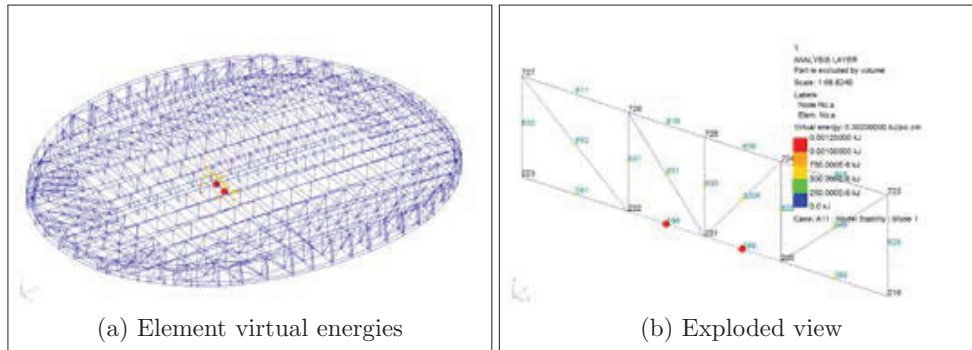


Figure 4: Roof truss model: virtual energies for first eigenvector.

Figure 4a shows the contour plot for element-wise virtual kinetic energies associated with the first eigenvector. The contour shows the largest values for exactly 6 amongst 850 elements in the model.

Figure 4b shows a closer view of three of those nodes along with the elements connected to them. Elements 589 and 590 are beams, whereas all other elements connecting at node 221 are bars, which is a modelling oversight. As a result, the dof corresponding to rotation about the X-axis at node 200 is unrestrained, which leads to a mechanism at the node. Restraining rotation about the X-axis at nodes 220 and 222 fixes the beams 589 and 590 from spinning about their axis. After the restraints are applied, the condition number estimate reported drops to  $O(10^5)$ .

### 5.2. Steel connection detail

The model in Figure 5 is a steel connection detail. The flange and the web of the connection are modelled using parabolic (eight noded, quadrilateral) plate elements [15, sec. 12.11.7]. Adjacent webs are connected with bolts, modelled as beams.

Initially, the stiffness matrix for the static analysis had a condition number estimate of  $O(10^{20})$ . Model stability analysis returned a distribution of eigenvalues tabulated in Table 2. We notice a cluster of eigenvalues of order  $10^{-11}$  and a large gap between the 40th and 41st eigenvalues. When element virtual kinetic energies ( $v^{(e)}$  in (3)) associated with the first 40 eigenvectors are contoured on the model, a handful of elements are seen to have large relative values in each mode. Figures 6 and 7 illustrate four of these plots.

The matrix is numerically singular and it arises from the choice of the

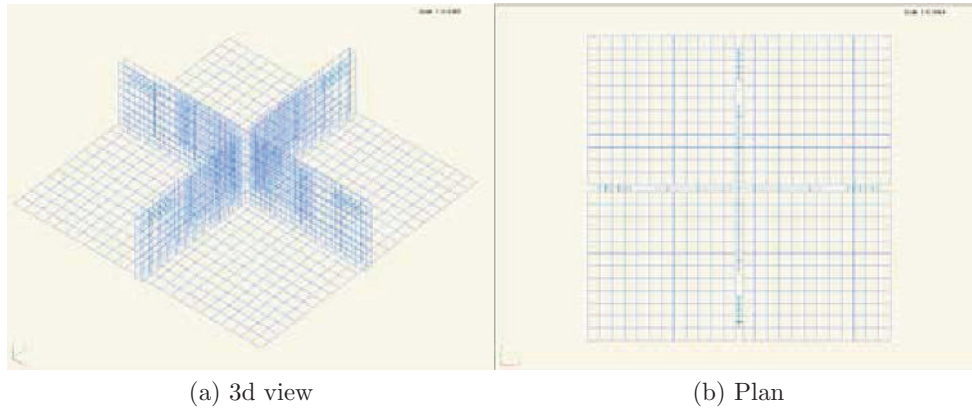


Figure 5: Connection detail. Blue quadrilaterals are plate elements.

Table 2: Eigenvalues of the stiffness matrix from the steel connection model.

|             |                |     |                |                |             |     |
|-------------|----------------|-----|----------------|----------------|-------------|-----|
| $\lambda_1$ | $\lambda_2$    | ... | $\lambda_6$    | $\lambda_7$    | $\lambda_8$ | ... |
| 2.80E-12    | 2.30E-12       | ... | 1.70E-11       | 5.40E-11       | 5.50E-11    | ... |
| ...         | $\lambda_{23}$ | ... | $\lambda_{40}$ | $\lambda_{41}$ |             |     |
|             | 7.28E-11       | ... | 7.70E-11       | 1.99E+3        |             |     |

element type for modelling the web and their connection to the beams. Eight noded plate elements have only five dofs per node, three for displacements and two for bending: rotation about the normal axis (drilling dof) is not supported. When such an element is connected to a beam, which has six dofs, oriented perpendicular to it, the connecting node acquires an active dof in the rotational direction about its axis. Since the plate does not have stiffness in this direction, the dof is unconstrained and this renders the matrix singular. The 41st eigenvalue is not small compared with the first 40, so the element virtual energies for a corresponding eigenvector are more evenly distributed (mode 41, Figure 7).

Restraining the beams against rotation about their X-axis brings the condition number estimate of the stiffness matrix down to  $O(10^8)$ .

### 5.3. Façade panels for a building

Whereas the previous examples involved singular stiffness matrices, this example deals with an ill conditioned matrix arising from erroneous nodal connectivity. Figure 8 shows a portion of a larger model of façade cladding

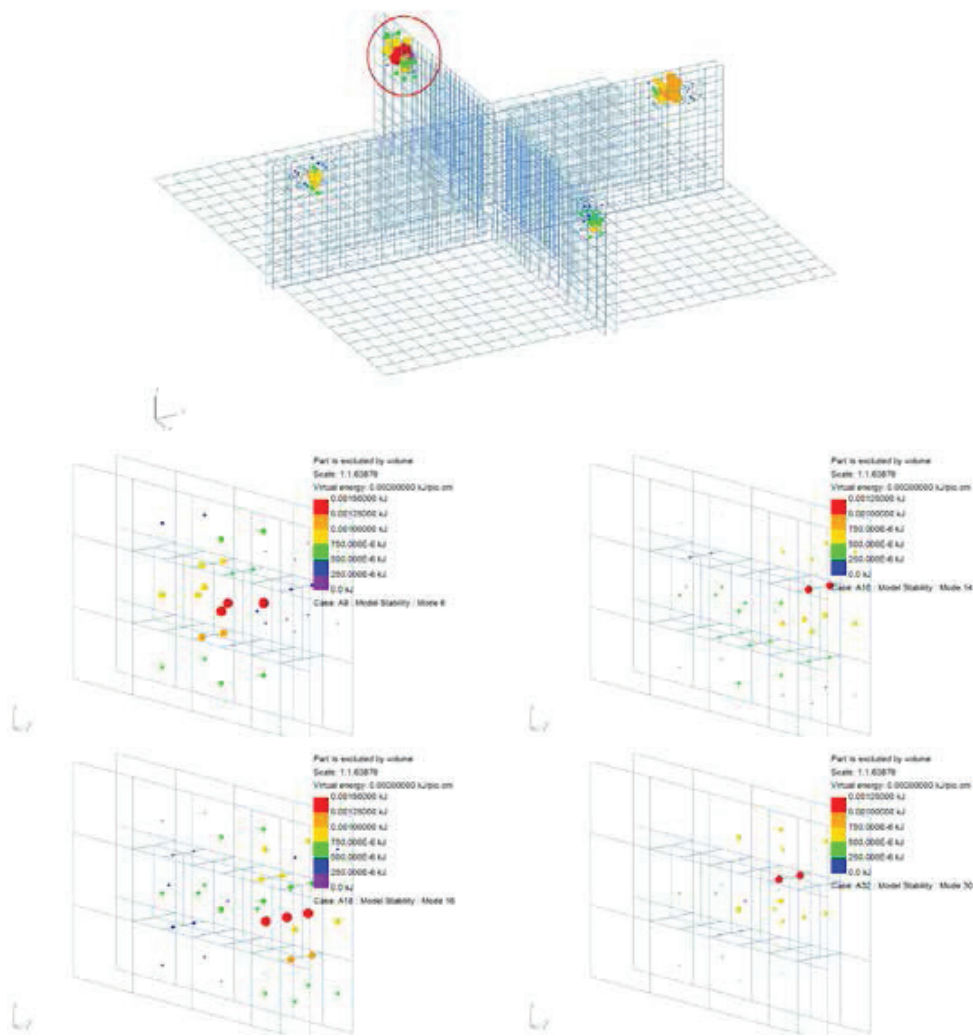


Figure 6: Contour plots for virtual energies for the connection model. Clockwise from top: energies associated with eigenvector 1, eigenvector 14, eigenvector 30, eigenvector 16 and eigenvector 6. Figures for eigenpairs 6, 14, 16 and 30 are magnified views of the encircled part in eigenvector 1.

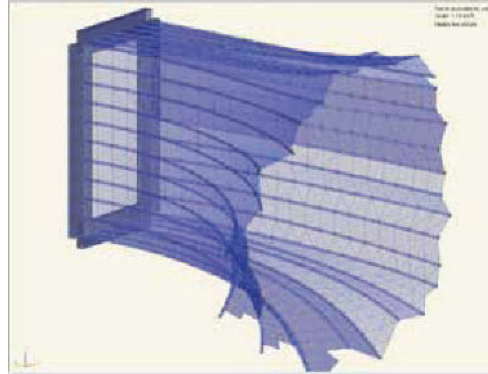


Figure 8: A small portion of the façade model.

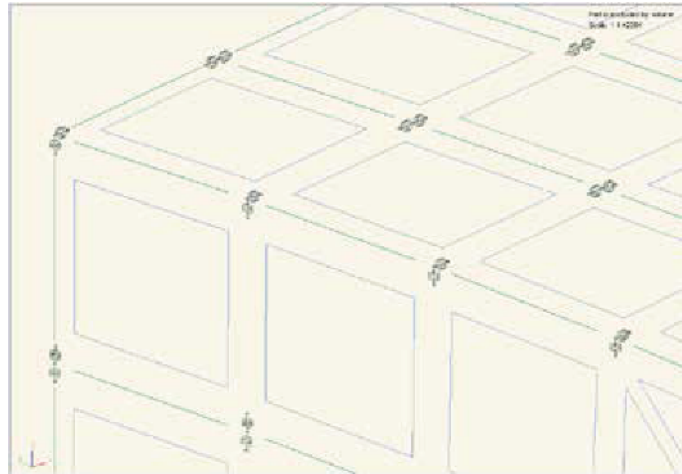
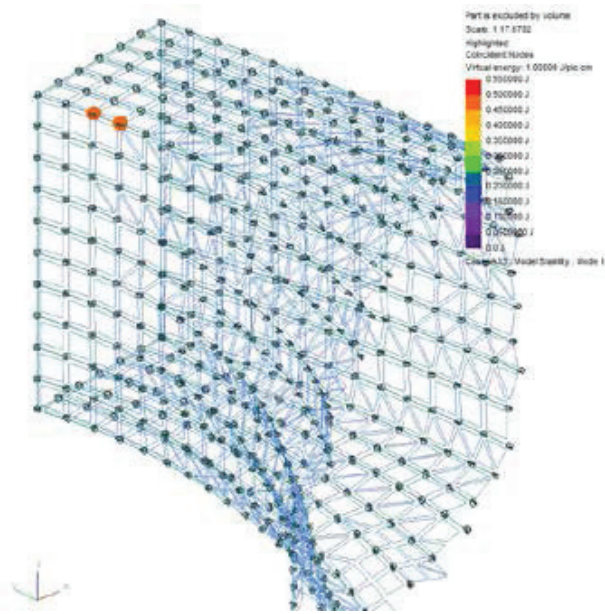
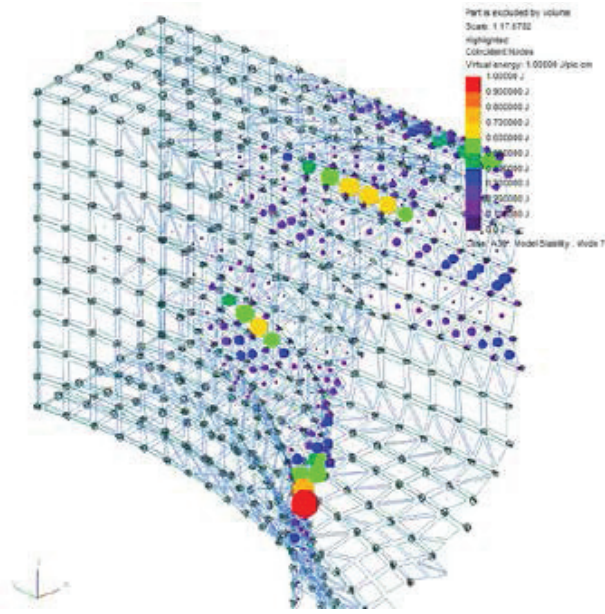


Figure 9: Close-up view of the element connectivity for façade model. Blue quadrilaterals are elements modelling the façade panels, green lines represent beams and springs are drawn as coils. Gaps between the elements are a graphic view setting and are only for visual clarity.



(a) Eigenpair 1



(b) Eigenpair 7

Figure 10: Element virtual energies for the first and seventh eigenpairs from the façade model.

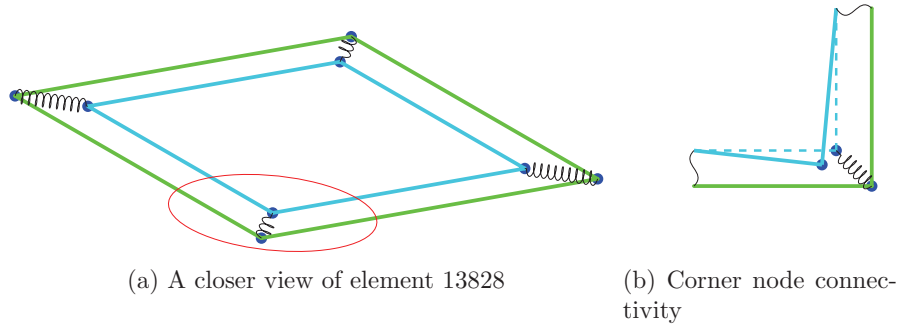


Figure 11: Close-up view of the element with large energy in Figure 10a.

Table 4: Eigenvalues of stiffness matrix (of size  $n$ ) from the Tall Building model.

|             |             |     |                |     |                |                |     |                 |             |
|-------------|-------------|-----|----------------|-----|----------------|----------------|-----|-----------------|-------------|
| $\lambda_1$ | $\lambda_2$ | ... | $\lambda_{25}$ | ... | $\lambda_{49}$ | $\lambda_{50}$ | ... | $\lambda_{n-1}$ | $\lambda_n$ |
| 1.22E3      | 1.70E3      |     | 9.73E3         |     | 1.24E4         | 1.25E4         |     | 2.02E17         | 2.02E17     |

connects the spring. The same error is repeated on the other node along the same edge of the element. As a result, the element is supported on only two corners on springs, which makes this part of the model highly flexible compared to surrounding parts, but not a mechanism.

Connecting the plane element to the two springs at the corners reduces the estimated condition number of the model to  $O(10^8)$ .

#### 5.4. Tall Building with concrete core

In this example, the ill conditioning was from the presence of elements that were very stiff in comparison with other elements. The GSA model in Figure 12 is of a tall building with a concrete core and has about 46000 elements using over a 1000 beam and slab sections with different properties. The model reported an initial condition number of order  $10^{15}$ .

Model stability analysis for up to the 50 smallest eigenvalues did not show a gap in the spectrum; a few of these are listed in Table 4. We therefore examine element virtual strain energies for the largest modes.

These highlight a handful of elements, as shown in Figure 13. An examination of the section properties of the elements reveals that certain members are modelled as strings of short beam elements (Figure 14), resulting in these

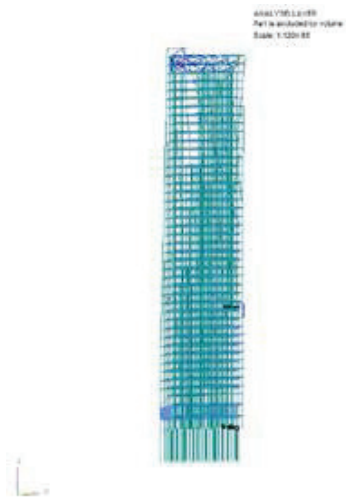


Figure 12: Tall building with stiff core.

elements acquiring large stiffness values. Rectifying the model in this case involves replacing the string of elements by single beam elements wherever they occur, which results in the condition number decreasing by several orders of magnitude.

Unlike in the case of smallest eigenvectors, the number of largest eigenvectors to be examined is arbitrary. This is because identification of elements with “large” stiffness is relative to the stiffness of the other elements in the model. Developing a criterion for selecting the number of largest eigenpairs is left to future work. At present we recommend iterating by correcting anomalies, re-analysing the model to find the new condition number and repeating the process until the condition number falls below the threshold  $\tau$  used in the method in section 4.

## 6. Conclusion

We have demonstrated a method for detecting a class of ill conditioned problems in structural analysis. These problems arise when there are dofs that contribute stiffnesses that are disproportional in magnitude and when there are tuples of dofs that have insufficient connectivity with the rest of the FE assembly, both of which are scenarios that commonly arise as user errors during model generation. We exploit the sparsity structure of the

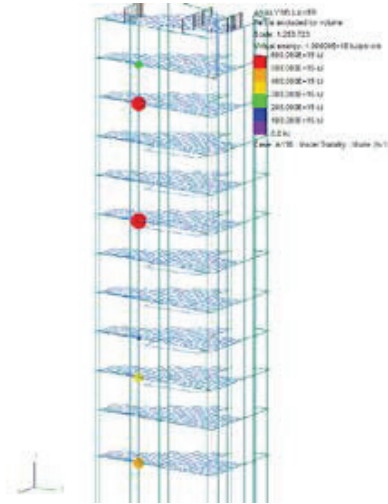


Figure 13: Contour of element virtual energies for the largest eigenvector of the stiffness matrix of tall building model.

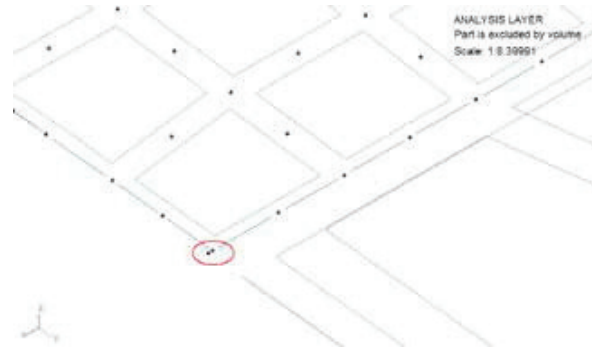


Figure 14: Magnified view of the element with high energy. The encircled beam has large stiffness.

smallest and largest eigenpairs and use element virtual energies to highlight elements that cause the ill conditioning. This method has been implemented in a commercial FE analysis package and we have shown through examples of real-life models that it works in practice. Further work will focus on developing a criteria for determining the number of largest eigenpairs used for investigating ill conditioning from large element stiffnesses.

### Acknowledgement

The authors wish to thank Professor David Silvester for his comments and suggestions.

### References

- [1] Klaus-Jürgen Bathe. *Finite Element Procedures*. Prentice-Hall International, Englewood Cliffs, NJ, USA, 1996.
- [2] Robert D. Cook, David S. Malkus, and Michael E. Plesha. *Concepts and Applications of Finite Element Analysis*. Wiley, New York, third edition, 1989.
- [3] Oasys Software. Oasys GSA suite. <http://www.oasys-software.com/gsa>, 2011. Retrieved on January 20, 2012.
- [4] Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2002.
- [5] IEEE standard for floating-point arithmetic. *IEEE Std 754-2008*, pages 1–58, August 2008.
- [6] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999.
- [7] Nicholas J. Higham. FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation (Algorithm 674). *ACM Trans. Math. Software*, 14(4):381–396, December 1988.

- [8] Nicholas J. Higham and Françoise Tisseur. A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra. *SIAM J. Matrix Anal. Appl.*, 21(4):1185–1201, 2000.
- [9] Charbel Farhat and Michel G eradin. On the general solution by a direct method of a large-scale singular system of linear equations: application to the analysis of floating structures. *International Journal for Numerical Methods in Engineering*, 41(4):675–696, 1998.
- [10] M. Papadrakakis and Y. Fragakis. An integrated geometric–algebraic method for solving semi-definite problems in structural mechanics. *Computer Methods in Applied Mechanics and Engineering*, 190(4950):6513 – 6532, 2001.
- [11] Gil Shklarski and Sivan Toledo. Computing the null space of finite element problems. *Computer Methods in Applied Mechanics and Engineering*, 198(3740):3084–3095, 2009.
- [12] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, fourth edition, 2013.
- [13] A.J. Wathen. An analysis of some element-by-element techniques. *Computer Methods in Applied Mechanics and Engineering*, 74(3):271–287, 1989.
- [14] I. Fried. Bounds on the extremal eigenvalues of the finite element stiffness and mass matrices and their spectral condition number. *Journal of Sound and Vibration*, 22(4):407–418, 1972.
- [15] Oasys Software. *GSA Version 8.6 reference manual*. Arup, 13 Fitzroy Street London W1T 4BQ, 2012. Available from [http://www.oasys-software.com/media/Manuals/Latest\\_Manuals/gsa8.6\\_manual.pdf](http://www.oasys-software.com/media/Manuals/Latest_Manuals/gsa8.6_manual.pdf).