

# Chapter 10

## Damage Detection Using Large-Scale Covariance Matrix

Luciana Balsamo, Raimondo Betti, and Homayoon Beigi

**Abstract** Statistical pattern recognition based structural damage detection is often developed exploiting the methods of outlier analysis. In this context, damage occurrence is assessed by analyzing whether a set of features extracted from the response of the system under unknown conditions departs from the population of features extracted from the response of the healthy system. The metric dominantly used for this purpose is the Mahalanobis Squared Distance (MSD). Evaluation of MSD of a point from a population requires the use of the inverse of the population's covariance matrix. It is known that when the feature dimensions are comparable or larger than the number of observations, the covariance matrix is ill-conditioned and numerically problematic to invert in the former case, while singular and not even invertible in the latter. To overcome this difficulty, three alternatives to the canonical damage detection procedure are investigated: data compression through Discrete Cosine Transform, use of pseudo-inverse of the covariance matrix, and use of shrinkage estimate of the covariance matrix. The performance of the three methods is compared using the experimental data recorded on a four story steel frame excited at the base by means of the shaking table available at the Carleton Laboratory at Columbia University.

**Keywords** Damage detection • Large-scale covariance matrix • Discrete cosine transform • Pseudo inverse • Shrinkage covariance matrix

### 10.1 Introduction

Statistical pattern recognition based structural damage detection is the task of assessing damage occurrence using information extracted from the structural response. It is developed by first learning the patterns drawn by such information when extracted from the system under healthy conditions, and by then comparing the learnt patterns with the patterns drawn by the same information extracted from the response of the system under unknown conditions: if the new patterns depart from the learnt ones more than a prescribed threshold, the structure is declared damaged. The information extracted from the structural response are known as *damage sensitive features* (dsf). Damage sensitive features need to be sensitive to structural changes due to damage, while remain insensitive to structural changes due to external effects, like environmental or operational conditions. The process of learning the patterns drawn by the damage sensitive features extracted from the response of the healthy system is known as *training*, while the process of comparing the trained features with those extracted from the structure under unknown conditions is named *testing*. One approach to measure the departure of the two populations of features is to evaluate the squared Mahalanobis distance of the testing features from the trained ones [1, 2].

Let us denote as  $\mathbf{x}$  a  $p$ -dimensional point representing the testing feature vector, and by  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  the mean vector and covariance matrix of the trained feature population, the Mahalanobis squared distance of  $\mathbf{x}$  from the trained model is defined as:

$$D(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (10.1)$$

---

L. Balsamo (✉) • R. Betti  
Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY 10027, USA  
e-mail: lb2591@columbia.edu

H. Beigi  
Recognition Technologies, Inc. White Plains, NY 10601, USA

In the real applications, the values of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are unknown, so that an estimate of such statistics is required. Usually, the unbiased sample counterparts of such statistics are used. Let us assume that the population of training features is collected into a matrix  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \in \mathbb{R}^{p \times n}$ : each column of  $\mathbf{Y}$  represents an observation of a  $p$ -dimensional feature vector. The sample mean of the training data set is given by:

$$\hat{\boldsymbol{\mu}} = \mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i, \quad (10.2)$$

while the unbiased sample covariance matrix of the training set is evaluated according to

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{m})(\mathbf{y}_i - \mathbf{m})^T. \quad (10.3)$$

As evidenced by Eq. (10.1), the evaluation of the Mahalanobis squared distance requires the computation of the inverse of the covariance matrix. Nonetheless, if one is presented with a data set of features of dimension  $p$  comparable or larger than the number of observations  $n$ , it is known that the estimate of the covariance matrix using its unbiased sample counterpart will not be neither accurate, nor reliable in the first case, and not even invertible in the second.

In this paper, three possible alternatives to the canonical damage detection approach are investigated. The first alternative focuses on decreasing the dimension of the feature vectors by employing a technique based on Discrete Cosine Transform (Sect. 10.2.1). The second approach overcomes the problem of covariance inversion by employing its pseudo-inverse (Sect. 10.2.2). The third option proposes the use of the shrinkage covariance matrix in place of the sample one (Sect. 10.2.3). In Sect. 10.3, the details of the feature extraction procedure and of the damage detection algorithm employed to obtain the results, presented in Sect. 10.4, are described. Damage detection is attempted using the acceleration time histories recorded on a four story steel frame excited by means of the shaking table available at the Carleton Laboratory of Columbia University. To mimic operational conditions variability, two undamaged conditions are considered. Damage is simulated by replacing some columns with elements of decreased cross section.

## 10.2 How to Handle Large-Scale Data Sets

### 10.2.1 Data Compression Using Discrete Cosine Transform

The objective of Principal Component Analysis (PCA) is that of projecting the original data set into a space whose basis are parallel to the principal components of the data set itself. A detailed treatment on the PCA technique is given in [3]. The principle behind PCA is that if one starts with a data set constituted of correlated features, it should be possible to decrease the dimensionality of such features by disregarding the dimensions associated with higher degree of correlation, and retaining only the ones associated with larger variance.

A popular approach to perform PCA is that of engaging into the Karhunen-Loève Transformation (KLT). Let us assume that a set of features is collected in a matrix  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\} \in \mathbb{R}^{p \times n}$ , whose  $i^{\text{th}}$  column represents the  $i^{\text{th}}$  observation of the  $p$ -dimensional feature vector  $\mathbf{y}_i$ . The first operation required to perform KLT is transforming the original data set  $\mathbf{Y}$  into one of zero mean,  $\hat{\mathbf{Y}}$ . Subsequently, the covariance matrix of  $\hat{\mathbf{Y}}$  may be estimated, typically through its unbiased sample estimate,  $\hat{\mathbf{S}}$ . The next operation requires the evaluation of the eigenvalues and eigenvectors of  $\hat{\mathbf{S}}$ , i.e. the definition of the matrix of eigenvectors,  $\mathbf{V}$ , and that of eigenvalues,  $\boldsymbol{\Lambda}$ , such that:

$$\mathbf{V}^{-1} \hat{\mathbf{S}} \mathbf{V} = \boldsymbol{\Lambda}, \quad (10.4)$$

where  $\boldsymbol{\Lambda}$ , is a diagonal matrix, whose main diagonal elements are the eigenvalues of  $\hat{\mathbf{S}}$  sorted in descending order,  $\mathbf{V}$  is the corresponding matrix of eigenvectors, whose columns are the eigenvectors of  $\hat{\mathbf{S}}$  arranged such that the first column represents the eigenvector associated with the largest eigenvalue of  $\hat{\mathbf{S}}$ , while the last column is the eigenvector associated with the smallest eigenvalue of  $\hat{\mathbf{S}}$ . By pre-multiplying the data points in  $\hat{\mathbf{Y}}$  by  $\mathbf{V}^T$ , the data set  $\hat{\mathbf{Y}}$  is rotated into a space whose principal axes are aligned with the eigenvectors of  $\hat{\mathbf{S}}$ : the first axis is associated to the direction along which  $\hat{\mathbf{Y}}$  has the largest variance, while the last dimension is associated with the direction along which  $\hat{\mathbf{Y}}$  has the least variance. Then, by picking only the first  $d$ ,  $d \leq p$ , elements of the rotated data set, it is possible to reduce the dimensions of the features in  $\hat{\mathbf{Y}}$  to a smaller dimension, without losing much information.

KLT is optimal in decorrelating the features into the transformed domain, in compacting the most information using only few coefficients and in minimizing the mean-square error (MSE) between the reconstructed and original feature vector. An important drawback of KLT is that its basis vectors are data dependent, since the basis functions of KLT are the eigenvectors of the covariance matrix of the features population. However, the objective of this work is to investigate the case where the estimation of the covariance matrix is unreliable, due to scarcity of observations. To resolve this vicious cycle, the Discrete Cosine Transform comes into play [4]. The Discrete Cosine Transform of a data sequence  $y[t], t = 0, \dots, N-1$ , is given by:

$$g_y[k] = a_k \sum_{t=0}^{N-1} y[t] \cos \left[ \frac{(2t+1)k\pi}{2N} \right], k = 0, \dots, N-1, \quad (10.5)$$

where  $a_k$  is equal to  $\frac{\sqrt{2}}{N}$  for  $k = 0$ , and to  $\frac{2}{N}$  otherwise, while  $g_y(k)$  is the  $k$ th DCT coefficient. DCT shares with KLT the same characteristics of data decorrelation, energy compaction, and minimum MSE between reconstructed and original signal, but its basis vectors are data independent. In [4], it is shown how the basis vectors of the DCT provide a good approximation of the eigenvectors of a class of Toeplitz matrices, which are often used to model the data covariance matrix of some weakly stationary processes. For said reasons, one can avail of DCT to compact data dimensions. In this paper, after having evaluated the  $p$ -dimensional feature vectors, a  $d$ -point DCT is applied to each feature instance, where  $d < p$ .

### 10.2.2 Pseudo-Inverse of the Covariance Matrix

In this section, the details of the evaluation of the Monroe-Penrose pseudo-inverse,  $\mathbf{S}^\dagger$ , of a square matrix  $\mathbf{S}$  of order  $p$  are briefly recalled. Pseudo-inverse computation starts by evaluating the singular value decomposition of the matrix to be inverted:

$$\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{S}, \quad (10.6)$$

where  $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{p \times p}$  are unitary matrices containing the left and right singular vectors of  $\mathbf{S}$ , respectively, while  $\mathbf{D}$  is a diagonal matrix, whose main diagonal elements,  $d_{ii}, i = 1, \dots, p$ , are the singular values associated with  $\mathbf{S}$ , sorted in descending order. The number of non-zero singular values is equal to the rank of the matrix. If  $\mathbf{S}$  is ill-conditioned, some of its singular values are very close to zero. The rows and columns of  $\mathbf{D}$  associated with the  $r$  singular values lower than a prescribed tolerance,  $\tau$ , can be then deleted from  $\mathbf{D}$ , resulting in a new diagonal matrix  $\hat{\mathbf{D}} \in \mathbb{R}^{(p-r) \times (p-r)}$  containing only non-zero singular values. In MatLab the tolerance value is set equal to  $\epsilon \cdot p \cdot \max(d_{ii})$ , where  $\epsilon$  is the distance from 1 to the next largest double precision number, that is  $\epsilon = 2^{(-52)}$  [5], and  $\max(d_{ii})$  is the largest singular value. Subsequently, the last  $r$  columns of  $\mathbf{V}$  and  $\mathbf{U}$  are discarded, resulting in the matrices  $\hat{\mathbf{V}}, \hat{\mathbf{U}} \in \mathbb{R}^{p \times (p-r)}$ . Finally, the pseudo-inverse of  $\mathbf{S}$  can be defined as

$$\mathbf{S}^\dagger = \hat{\mathbf{V}}\hat{\mathbf{D}}^{-1}\hat{\mathbf{U}}^T. \quad (10.7)$$

### 10.2.3 Shrinkage Covariance Matrix

Unbiased sample covariance matrix is the covariance matrix estimate most often employed for modeling the correlation properties of a features data set. However, as pointed out by Stein in [6], such estimate is reliable only as long as the number of observations is large enough, i.e. only as long as the number of observations,  $n$ , is at least one order of magnitude larger than the dimension of the single observation vector,  $p$ . If such condition is not satisfied, the sample estimate of the covariance matrix produces very poor results. In absence of enough observations, a popular approach is that of considering only the diagonal elements of the covariance matrix. As well described in [7], while the unbiased sample covariance matrix has zero bias, i.e. its expected value is prescribed to be equal to the value of the population covariance matrix,  $\mathbf{\Sigma}$ , the diagonal covariance matrix has minimum variance, but large bias. The shrinkage estimation of the covariance matrix allows for constructing a covariance matrix that shares the advantages of both models. Denoting as  $\mathbf{S}$  the unbiased sample covariance matrix, and as  $\mathbf{T}$  the diagonal covariance matrix, the shrinkage estimate of the covariance matrix is given by:

$$\mathbf{S}^* = \lambda \mathbf{T} + (1 - \lambda) \mathbf{S}, \quad (10.8)$$

where  $\lambda$  is known as the *shrinkage coefficient*. Ledoit and Wolf [8] were the first to propose an analytical formula for the estimation of the shrinkage coefficient:

$$\lambda = \frac{\sum_{i=1}^p \sum_{j=1}^p [\text{Var}(S_{ij}) - \text{Cov}(S_{ij}, T_{ij})]}{\sum_{i=1}^p \sum_{j=1}^p (S_{ij} - T_{ij})^2}. \quad (10.9)$$

Adoption of Eq. (10.9) to estimate the shrinkage coefficient leads to the minimization of the mean square error between  $\mathbf{S}^*$  and  $\mathbf{\Sigma}$ . Since  $\mathbf{S}^*$  is given by the addition of an invertible matrix with a square matrix, the shrinkage estimate of the covariance matrix is invertible and can be then used in place of the unbiased sample estimate of  $\mathbf{\Sigma}$ .

### 10.3 Feature Extraction and Damage Detection Algorithm

In this section, the procedure to extract the damage sensitive feature, as well as the steps involved in the damage detection algorithm are described. For a more detailed treatment of the feature extraction procedure the reader is referred to [9].

The damage sensitive feature used in this work is a modified version of the Mel-Frequency Cepstral Coefficients. The use of such feature is novel in the field of civil engineering, but it is customary in the field of speaker and speech recognition [10]. The extraction of such features begins by segmenting the time history in pieces, called *frames*, short enough to be considered stationary. Non rectangular windows are then applied to each frame, in order to get rid of undesirable effects at the onset and offset of the frame spectra. The power spectrum of each frame is then evaluated and warped into a frequency scale given by:

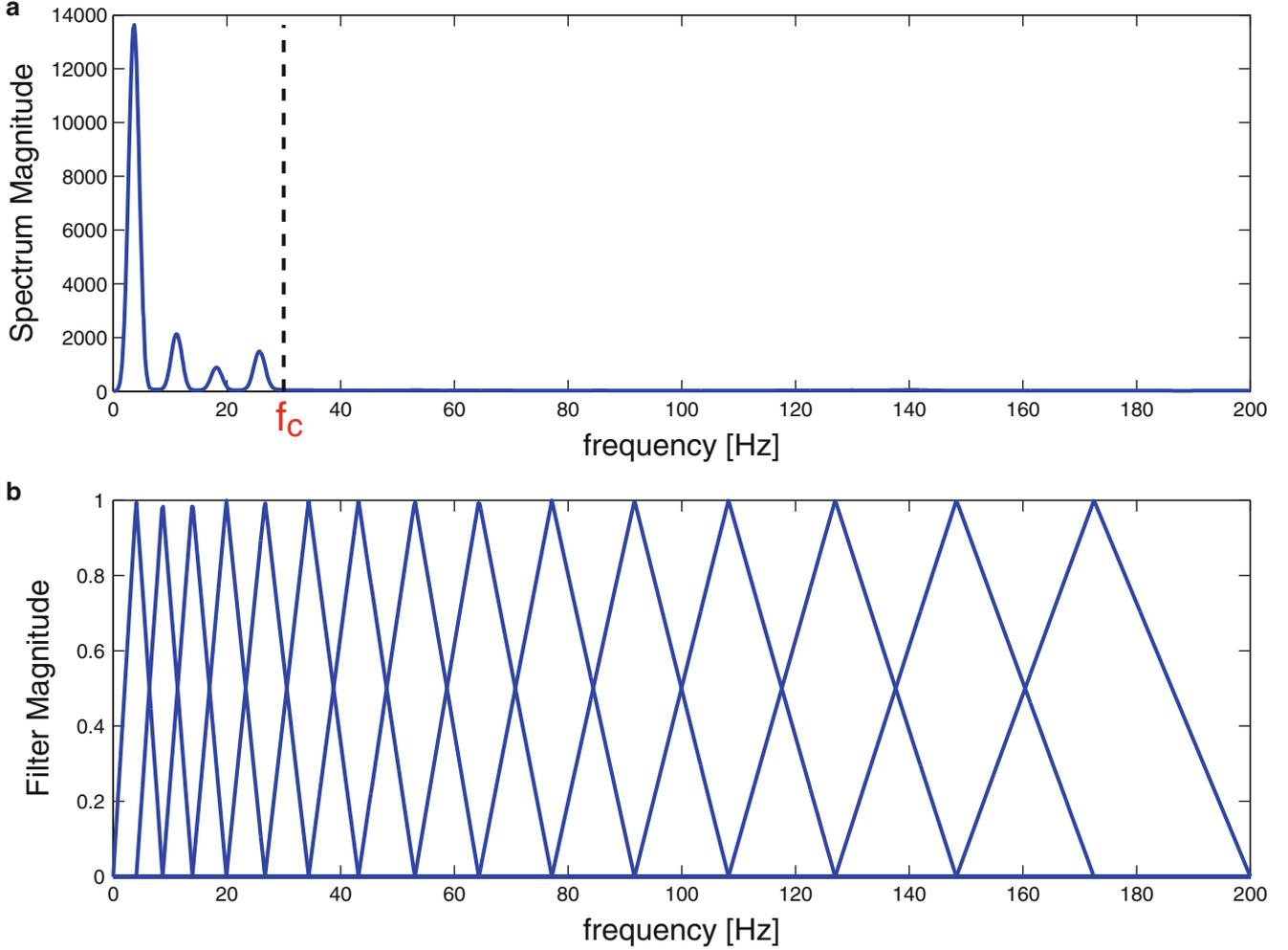
$$f_{warping} = f_c \log_2 \left( 1 + \frac{f}{f_c} \right) \quad (10.10)$$

where  $f_{warping}$  is the value of the frequency in the warped scale associated to the value  $f$  of the linear frequency scale, while  $f_c$  is a user-defined cutoff frequency representing the boundary of the major power content in the system spectrum. The relation between  $f_{warping}$  and  $f$  is linear up to  $f_c$ , and becomes logarithmic after such value. Frequency warping is obtained by grouping together the spectrum values into  $M$  critical bands, and weighting each cluster by a triangular filter. The series of  $M$  triangular filters has centers equally spaced on the  $f_{warping}$  scale. In this work, the value of  $M$  is set equal to the entire part of  $3 \log(f_s)$ , as suggested in [11], where  $f_s$  is the value of the sampling rate in Hz. In particular, since all time histories used in this work are sampled at 400 Hz, 15 filters have been used to define the feature vectors. Figure 10.1 represents the spectrum obtained averaging all time histories recorded for the undamaged scenario 1 described in Sect. 10.4 from which the selection of  $f_c$  is made available, and the corresponding triangular filters. Finally, the logarithm of each warped spectrum is evaluated and the real part of an  $M$ -point inverse Fourier Transform is computed via inverse Discrete Cosine Transform. The first coefficient is then discarded, as it has been demonstrated that such coefficient is very sensitive to external factors effects, as well as input effects. Let us denote as  $\ell$  the number of frames in which a time history is segmented, at the end of this process,  $\ell$   $(M - 1)$ -dimensional vectors are extracted: the average of such vectors is evaluated leading to a single vector consisting of  $(M - 1)$  elements. Let us now assume that  $s$  sensors are available: in this work, a data set is defined as the ensemble of  $s$  time histories recorded from each available sensor during a single measurement campaign. From each of the  $s$  response time histories, an  $(M - 1)$ -point long feature vector is extracted and stacked with the other  $(s - 1)$  vectors, so that each data set is represented by a feature vector  $\mathbf{y} \in \mathbb{R}^{s \cdot (M-1) \times 1}$ . From now on, for the sake of notation brevity and consistency with the previous treatments, the dimension of a feature vector extracted from a data set will be denoted as  $p$ , i.e.  $p$  is equal to  $s \cdot (M - 1)$ .

Let us refer to  $n_{tr}$  as the number of data sets available for training. At the end of the feature extraction procedure, if the technique described in Sect. 10.2.1 is engaged, each of the  $n_{tr}$  feature vectors is transformed into a  $d$  elements vector by means of DCT, where  $d$  is either equal to  $\frac{3p}{4}$  if  $p \leq n_{tr}$ , or to  $\frac{3n_{tr}}{4}$  otherwise.

For the methods described in Sects. 10.2.1 and 10.2.2, the training model is constructed evaluating the sample mean  $\mathbf{m}_{tr}$ :

$$\mathbf{m}_{tr} = \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \mathbf{y}^{(i)} \quad (10.11)$$



**Fig. 10.1** Warping procedure: (a) cutoff frequency selection, (b) triangular filters

and the unbiased sample covariance matrix  $\mathbf{S}_{tr}$ :

$$\mathbf{S}_{tr} = \frac{1}{n_{tr} - 1} \sum_{i=1}^{n_{tr}} (\mathbf{y}^{(i)} - \mathbf{m}_{tr}) (\mathbf{y}^{(i)} - \mathbf{m}_{tr})^T \quad (10.12)$$

where  $\mathbf{y}^{(i)}$  represents the realization of the feature vector extracted from the  $i^{\text{th}}$  training data set. If the technique described in Sect. 10.2.3 is engaged, the shrinkage estimate of the covariance matrix,  $\mathbf{S}^*$ , is evaluated in place of  $\mathbf{S}$ .

The threshold can then be defined. Since the paucity of observations makes the assumption of normally distributed features difficult to be satisfied, a resampling technique is employed to set the boundary between damaged and undamaged. As mentioned, the departure of the testing feature from the training model is analyzed by means of Mahalanobis Squared Distance. The MSD of the  $i^{\text{th}}$  training feature vector from a population, obtained from the training population from which the  $i^{\text{th}}$  realization is left out, is computed. The resulting  $n_{tr}$  values are sorted in ascending order, and the value exceeded by only the 5% of instances is picked as threshold value,  $\gamma$ .

Finally, let us denote as  $n_{te}$  the number of data sets available for testing. From each available data set, a  $p$ -dimensional feature vector is extracted. If feature dimension reduction is required, a  $d$ -point DCT of each testing feature vector realization is performed, where  $d$  has the same values considered for training. If more than one data set is available, the mean of the available testing vectors,  $\mathbf{m}_{te}$ , is computed. If  $n_{te}$  is equal to 1, as the case for short-time applications, and for the examples considered in this work,  $\mathbf{m}_{te}$  is simply the single  $p$ -dimensional vector extracted from the only available data set. The structure can be then declared damaged if

$$D(\mathbf{m}_{te}) = (\mathbf{m}_{te} - \mathbf{m}_{tr})^T \mathbf{S}_{tr}^{-1} (\mathbf{m}_{te} - \mathbf{m}_{tr}) > \gamma. \quad (10.13)$$

where  $\mathbf{S}_{lr}^{-1}$  is replaced either by the pseudo-inverse of  $\mathbf{S}_{lr}$ ,  $\mathbf{S}_{lr}^\dagger$ , in case the second method is employed, or by the inverse of the shrinkage estimate of the covariance matrix,  $(\mathbf{S}_{lr}^*)^{-1}$ , if the third approach is considered.

## 10.4 Results

The structure used to compare the performance of the three proposed approaches is a four-story A36 steel frame with an inter-story of 533 mm and floor plate dimensions of 610×457×12.7 mm. The floors are braced diagonally in only one direction, hereafter denoted as strong direction. The structure was excited along the weak direction of bending by means of a medium-scale uniaxial hydraulic shake table. All structural connections are bolted. In addition to said reference configuration, denoted as **U1** in the following, an additional undamaged scenario (**U2**) was considered in order to simulate operational variability. The second undamaged condition is simulated by adding two masses at the third floor between columns A and B, on both edges C and D. The first damage scenario (**D1**) was simulated by replacing the column elements on side A of the third inter-story with elements with a cross-section reduced to  $\frac{3}{4}$  of the original, while the second damage scenario (**D2**) was modeled by replacing all the column elements of the third inter-story by elements of reduced cross-section.

The structure was instrumented with 8 piezo-electric accelerometers located as displayed in Fig. 10.2, measuring accelerations along the weak direction. The sensor setup is such that also torsional effects may be captured. In fact, preliminary studies have shown that the structure is not of ideal shear-type kind, but torsional modes in the strong direction of bending may be identified, even when the load is applied along the weak direction. Four ground acceleration time histories recorded during El Centro (1940), Hachinohe (1983), Northridge (1994) and Kobe (1995) earthquakes, in addition to the acceleration time history obtained from the design spectrum of EC8 were applied as inputs. To ensure that the structure was excited by the proper range of the time histories power spectra, a time scale of  $\frac{1}{\sqrt{3}}$  was introduced and, to prevent yielding and additional unexpected damage, the input time histories were properly scaled in magnitude. Inputs and outputs were sampled at 400 Hz.

The training set is constituted by 50 data sets: for each of the two undamaged scenarios, 5 experiments for each of the five inputs are considered. The testing set is constituted by 40 data sets, 10 for the undamaged scenarios (5 for each condition), and 15 for each of the two damaged conditions. Each testing data set is used individually, so that 40 tests are performed. Five different sensor setups are simulated, using only the response of some instruments. In the following, the five sensor setups are denoted as **S1**, **S2**, etc.: **S1**) all sensors, **S2**) sensors 1, 2, 3, 4, **S3**) sensors 1, 3, 5, 7, **S4**) sensors 1, 4, 7, and **S5**) mid-span floor accelerations. The data set of the last sensor setup is obtained by adding the acceleration response time history recorded by the sensor on edge C with the one recorded by the sensor on edge D and dividing by 2, at each floor: for example, the first mid-span acceleration is evaluated by averaging the time histories recorded by sensors 1 and 8.

The results are presented in Table 10.1. In Table 10.1, the error for the individual condition identification, as well as the overall Type I and Type II errors, are shown. Type I error occurs when a healthy condition is labeled as damaged, while Type II error occurs when a damaged condition is classified as undamaged. The three approaches presented in Sects. 10.2.1, 10.2.2 and 10.2.3 are labeled as *Method 1*, *Method 2* and *Method 3*, respectively. In Table 10.1, under each sensor setup label, the reduced dimension  $d$  of the feature vector, its original dimension  $p$  and the value of the shrinkage coefficient  $\lambda$  are also presented.

While the first method results in better outcomes for what concerns Type I error, methods 2 and 3 perform generally better than the first in detecting damage. Firstly, the high Type I error rate observed for some sensor setups for Method 2 is worth of a comment. For each undamaged scenario, only 5 tests are performed: it is sufficient that only one instance be incorrectly classified as damaged, for that state to be characterized by 20% error. Therefore, even though Method 2 gives certainly the worst results in terms of Type I error, probably the error rate would decrease if more test data sets were available. On the other hand, the high Type II error rate observed for the first method for the first three sensor setups is due to a too high threshold value, as shown from Fig. 10.3. The explanation of this behavior is clarified analyzing the results of Method 4, where, in an attempt to combine the advantages of the three approaches, the three techniques are used together. This combination is pursued by extracting the damage sensitive features as described in Sect. 10.3, applying the DCT-based feature dimension reduction, by then modeling the training covariance matrix via shrinkage estimation, and, finally, by evaluating the squared Mahalanobis distance using the pseudo-inverse of the shrunk covariance matrix. As expected, the results improve dramatically and clarify that the model of the covariance matrix used in the first method is not reliable enough to be inverted and delivering accurate squared Mahalanobis distance. In fact, the inaccuracy of the sample covariance matrix estimate of the features used in the first method is directly related to the value of the threshold, as its inverse is evaluated to define such value. Indeed, that a well conditioned covariance matrix estimate can deliver accurate results is confirmed

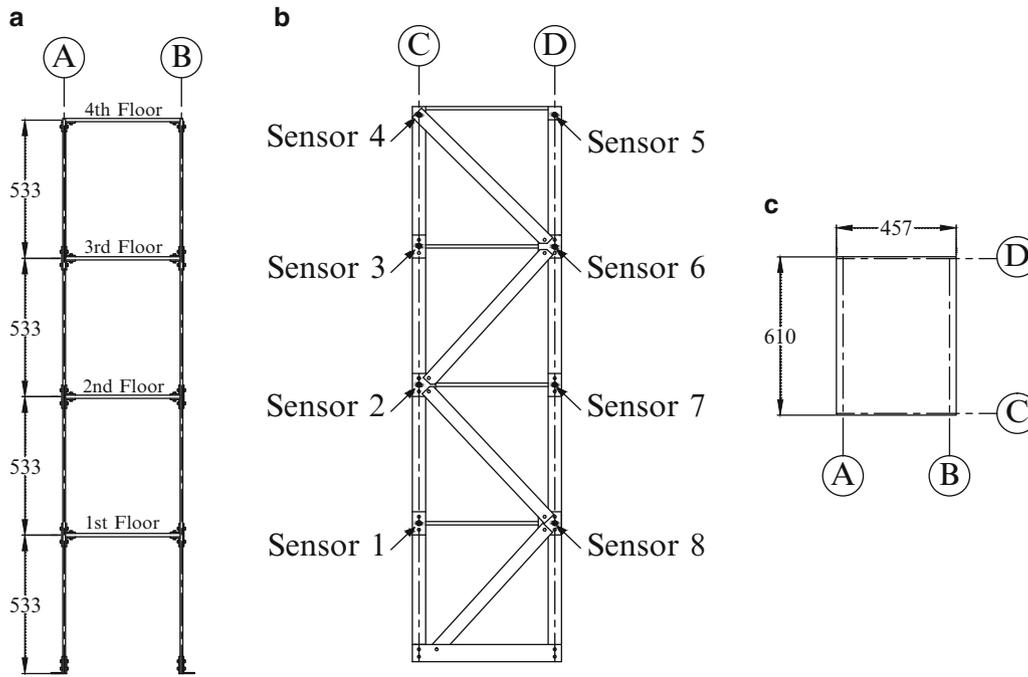
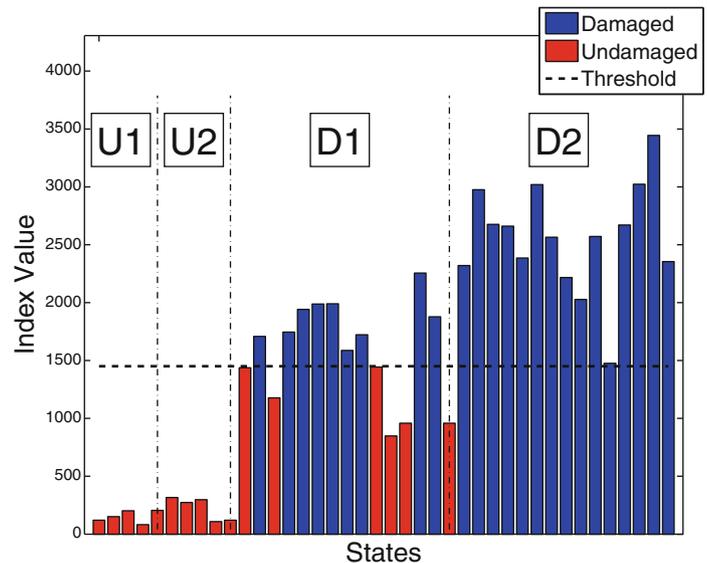


Fig. 10.2 Sensors location

Table 10.1 Results

Sensor Setup	State	Method 1	Method 2	Method 3	Method 4
<b>S1:</b> $d = 40$ $p = 120$ $\lambda = 0.0164$	U1	0.00%	0.00%	0.00%	0.00%
	U2	0.00%	0.00%	0.00%	0.00%
	Type 1	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
	D1	40.00%	0.00%	0.00%	0.00%
	D2	0.00%	0.00%	0.00%	0.00%
	Type II	<b>20.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
<b>S2:</b> $d = 40$ $p = 60$ $\lambda = 0.0163$	U1	0.00%	0.00%	0.00%	0.00%
	U2	0.00%	0.00%	0.00%	0.00%
	Type 1	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
	D1	40.00%	0.00%	0.00%	0.00%
	D2	0.00%	0.00%	0.00%	0.00%
	Type II	<b>20.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
<b>S3:</b> $d = 40$ $p = 60$ $\lambda = 0.0165$	U1	0.00%	20.00%	0.00%	0.00%
	U2	0.00%	20.00%	0.00%	0.00%
	Type 1	<b>0.00%</b>	<b>20.00%</b>	<b>0.00%</b>	<b>0.00%</b>
	D1	13.33%	0.00%	0.00%	0.00%
	D2	0.00%	0.00%	0.00%	0.00%
	Type II	<b>6.67%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
<b>S4:</b> $d = 45$ $p = 36$ $\lambda = 0.0169$	U1	0.00%	0.00%	0.00%	0.00%
	U2	0.00%	20.00%	0.00%	0.00%
	Type 1	<b>0.00%</b>	<b>10.00%</b>	<b>0.00%</b>	<b>0.00%</b>
	D1	0.00%	0.00%	0.00%	0.00%
	D2	0.00%	0.00%	0.00%	0.00%
	Type II	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>
<b>S5:</b> $d = 40$ $p = 60$ $\lambda = 0.0168$	U1	0.00%	0.00%	20.00%	0.00%
	U2	0.00%	20.00%	0.00%	0.00%
	Type 1	<b>0.00%</b>	<b>10.00%</b>	<b>10.00%</b>	<b>0.00%</b>
	D1	0.00%	0.00%	0.00%	0.00%
	D2	0.00%	0.00%	0.00%	0.00%
	Type II	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>	<b>0.00%</b>

**Fig. 10.3** Results for Method 1, Sensor Setup 1



also by analyzing the results of Method 3, where the feature dimensions are not reduced, but the shrinkage estimate of the covariance matrix is used. It is also noteworthy that the shrinkage coefficient is very small (the values of  $\lambda$  oscillate between 0.0163 and 0.0169), so that the correction imposed to the sample covariance matrix is minimal.

## 10.5 Conclusions

Three alternatives have been proposed to deal with the problem of paucity of data to solve the structural damage detection assignment, namely reduction of feature dimensionality via discrete cosine transform, inversion of the unbiased sample covariance matrix via pseudo-inverse, and use of the shrinkage estimate in place of the more common sample estimate of the covariance matrix. Used by itself, each approach has its own advantages and disadvantages. Mainly, it has been observed that while reducing the feature dimension to the principal components allows to decrease the Type I error rate, to the detriment of Type II error, operating on the stability of the inverse of the covariance matrix estimate aids decreasing false acceptance error, leading, however, to an increase of false alarm error rate. Nonetheless, it has been shown that the advantages of all methods may be combined by using all alternatives together.

**Acknowledgements** The authors would like to acknowledge the other members of the Shake Table Facility Group, namely, the manager of the Carleton Laboratory Adrian Brügger, Prof. Andrew Smyth, Dr. Manolis Chatzis and Suparno Mukhopadhyay, for the collaboration during the experimental phase of the work. The authors would also like to acknowledge the support from the Center for Advanced Infrastructure and Transportation (CAIT), Rutgers, the State University of New Jersey, Award No. Rutgers S1760924.

## References

1. Worden K, Manson G, Fieller NRJ (2000) Damage detection using outlier analysis. *J Sound Vib* 229(3):647–667
2. Cross EJ, Manson G, Worden K, Pierce SG (2012) Features for damage detection with insensitivity to environmental and operational variations. *Proc Roy Soc A Math Phys Eng Sci* 468(2148):4098–4122
3. Jolliffe I (2002) *Principal component analysis*, 2nd edn. Springer, New York
4. Ahmed N, Natarajan T, Rao KR (1974) Discrete cosine transform. *IEEE Trans Comput* 23:P90–93
5. MATLAB 7.9 (R2009b) The MathWorks, Inc., Natick, MA, United States, 2009
6. Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proceedings of the third Berkeley symposium on mathematics, statistics and probability*, vol 1. University of California Press, Berkeley, p 197–206
7. Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* 4(1):1–30
8. Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J Empir Finance* 10:603–621

9. Balsamo L, Betti R, Beigi H (2013) Structural damage detection using speaker recognition techniques. In: Proceedings of the 11th international conference on structural safety and reliability (ICOSSAR), Columbia University, New York, NY, USA, June 16–20, 2013
10. Beigi H (2011) Fundamentals of speaker recognition. Springer, New York
11. Fraile R, Sáenz-Lechón N, Godino-Llorente J, Osma-Ruiz V, Gómez-Vilda P (2008) Use of mel-frequency cepstral coefficients for automatic pathology detection on sustained vowel phonations: mathematical and statistical justification. In: Proc. 4th international symposium on image/video communications over fixed and mobile networks, Bilbao, Brazil, July, 2008