

# Optimal singular correlation matrices estimated when the sample size is less than or equal to the number of random variables

M. Vořechovský\*

Institute of Structural Mechanics, Brno University of Technology, Veveří 95, 602 00 Brno, Czech Republic

## ARTICLE INFO

### Article history:

Received 23 October 2011

Received in revised form

7 June 2012

Accepted 20 June 2012

Available online 7 July 2012

## ABSTRACT

This paper presents a number of theoretical and numerical results for two norms of optimal correlation matrices in relation to correlation control in Monte Carlo type sampling and the designs of experiments. The optimal correlation matrices are constructed for cases when the number of simulations (experiments)  $N_{\text{sim}}$  is less than or equal to the stochastic dimension, i.e. the number of random variables (factors)  $N_{\text{var}}$ . In such cases the estimated correlation matrix can not be positive definite and must be singular. However, the correlation matrix may be required to be as close to the unit matrix as possible (optimal). The paper presents a simple mechanical analogy for such optimal singular positive semidefinite correlation matrices. Many examples of optimal correlation matrices are given, both analytically and numerically.

© 2012 Elsevier Ltd. All rights reserved.

## 1. Introduction

Statistical sampling is of interest not only to statisticians, but to a variety of research fields such as engineering, economics, design of experiments, and operational research. Analysts are often faced with time-consuming and expensive sampling (physical or virtual) to achieve statistically significant output statistics. The need for the exploration of multidimensional stochastic domains is the central topic in statistical, sensitivity and reliability analyses.

In Monte Carlo type simulations, there is more commonly a need to reduce the number of samples. Moreover, it sometimes happens that the number of simulations (sample size  $N_{\text{sim}}$ ) is smaller than the number of random variables (dimension  $N_{\text{var}}$ ). Similarly, in the (mostly preliminary) designs of experiments focused on exploration of the space of variables, the number of experiments might be smaller than the number of factors. In probabilistic mechanics there is an increasing trend for incorporating random field modeling in practical simulations. Simulation of random fields is another example of an application where the number of simulations is often smaller than the number of variables. Suppose the analyst requires just a few samples of a random field. This can happen, for example, when the random fields are simulated to represent random properties of a system analyzed by computationally expensive finite element simulations. In such cases, the number of analyzed samples is very low, much lower than the number of random variables needed

for the representation of the fields. The most frequently used methods for the representation of random fields are (i) the spectral representation method (see e.g. [1–4]), (ii) Karhunen–Loève expansion [5–8] and (iii) the simulation based on the sampling theorem [9], see a comparison of the techniques in [10]. In these techniques, a sample representing  $N_{\text{var}}$  independent variables is needed. Even more striking is the situation when the analysis requires samples of many cross-correlated fields (especially with small correlation lengths) [8].

In all the mentioned applications, the common requirement is to regularly cover an  $N_{\text{var}}$ -dimensional space with  $N_{\text{sim}}$  points. The regularity of distribution of the points is often measured through correlation matrix  $\mathbf{A}$  estimated from the sample. The requirement is that matrix  $\mathbf{A}$  is as close as possible to the unit matrix. The requirement of mutual uncorrelatedness can be viewed as the relaxed requirement of independence between variables (factors).

The topic of correlation control in small sample simulation was considered in well-known works on Latin Hypercube sampling from random vectors [11–13]. Correlation control in small sample Monte Carlo type simulation was also the central topic in [14,15], where a combinatorial optimization algorithm was developed for the mutual ordering of a sample to achieve a good match between the target correlation matrix  $\mathbf{T}$  and actual correlation matrix  $\mathbf{A}$  estimated from the sample. In that work, two scalar measures of the difference  $\mathbf{T} - \mathbf{A}$  were defined and subjected to minimization. The proposed simulated annealing algorithm turned out to be extremely effective and we will show that it delivers optimal results especially when  $N_{\text{sim}} \leq N_{\text{var}}$ .

In this paper, we exploit the two scalar norms of the error matrix  $\mathbf{T} - \mathbf{A}$  and denote them as  $\rho_{\text{max}}$  and  $\rho_{\text{rms}}$ . These norms represent a natural choice and are commonly used in the assessment of correlation errors, see e.g. [16–18]. When the

\* Tel.: +420 737319317; fax: +420 5 41240994.

E-mail address: [vorechovsky.m@fce.vutbr.cz](mailto:vorechovsky.m@fce.vutbr.cz).

URL: <http://www.fce.vutbr.cz/STM/vorechovsky.m/>.

suitable error matrix norms are defined, an analyst preparing the sample (numerical or physical experiment) is interested in achieving the lowest possible value for it. In this paper, we determine the lower bounds of these two norms and we also deliver numerical examples of the best possible correlation matrices for many frequent combinations of  $N_{\text{sim}} \leq N_{\text{var}}$ .

The paper is organized as follows. Section 2 introduces the two analyzed norms of correlation error matrices. Section 3 defines three types of optimal correlation matrices and presents some general facts about these matrices. Section 4 presents solutions for the crossover case when  $N_{\text{sim}} = N_{\text{var}}$ . The subsequent sections present results on the optimal correlation matrices for general sample sizes  $N_{\text{sim}}$ .

## 2. Correlation matrix estimation, errors and its norms

Let us consider the design matrix  $\mathbf{X}$  (sampling plan) that is formed by  $N_{\text{sim}}$  columns (simulations) times  $N_{\text{var}}$  rows (variables, factors). In other words, by a ‘sampling plan’, we mean an  $N_{\text{var}} \times N_{\text{sim}}$  matrix containing realizations of a random vector, see Table 1 in [14].

The estimated correlation matrix is a symmetric matrix of order  $N_{\text{var}}$  and can be written as the sum

$$\mathbf{A} = \mathbf{I} + \mathbf{L} + \mathbf{L}^T \quad (1)$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{L}$  is a strictly lower triangular matrix with entries within the range  $(-1, 1)$ . There are  $N_c$  correlations (entries in the  $\mathbf{L}$  matrix) that describe pairwise correlations:

$$N_c = \binom{N_{\text{var}}}{2} = \frac{N_{\text{var}}(N_{\text{var}} - 1)}{2}. \quad (2)$$

The most frequently used correlation coefficients in  $\mathbf{L}$  are the Pearson, Spearman and Kendall correlations. The sampling versions (estimation formulas) for all three of these correlation coefficients are presented and thoroughly studied in [19].

### 2.1. Norms of correlation error

Eqs. (11) and (14) from paper [14] define the norms  $\rho_{\text{rms}}$  and  $\rho_{\text{max}}$  of the error matrix  $\mathbf{E}$ , a matrix obtained as a difference between the target ( $\mathbf{T}$ ) and actual (estimated,  $\mathbf{A}$ ) correlation matrices. Having calculated the difference  $\mathbf{E}$  between the target and actual matrices, it is a natural choice to define some suitable scalar metrics of  $\mathbf{E}$ .

The most frequent requirement is to generate samples with uncorrelated marginals in which case the target correlation matrix is the unit matrix  $\mathbf{T} = \mathbf{I}$ . This case is very general as the desired dependency pattern can be introduced e.g. by transforming the uncorrelated sample through copulas, or using the eigenvectors and eigenvalues of the desired correlation matrix, or using the Cholesky decomposition, etc. Therefore, in this paper we only consider the general case of desired uncorrelatedness.

If  $\mathbf{T}$  is the unit matrix then the error norms can be reformulated directly as norms of the actual (estimated) correlation matrix  $\mathbf{A}$ . Moreover, we ignore weighting of various entries in the correlation matrix introduced in [14,15] (or, we in fact consider unit weights) and the two norms can be written as follows. The *absolute* norm reads:

$$\rho_{\text{max}} = \max_{1 \leq i < j \leq N_{\text{var}}} |A_{i,j}| \quad (3)$$

and the *root mean square* correlation error metrics

$$\rho_{\text{rms}} = \sqrt{\frac{1}{N_c} \sum_{i=1}^{N_{\text{var}}-1} \sum_{j=i+1}^{N_{\text{var}}} A_{i,j}^2}, \quad (4)$$

i.e. the square root of the average of squares of all off-diagonal correlation matrix entries.

The  $\rho_{\text{max}}$  norm is more conservative than  $\rho_{\text{rms}}$  because it considers the extreme deviation from zero error while  $\rho_{\text{rms}}$  is more of averaging type, see [19].

## 3. General comments

When the number of simulations  $N_{\text{sim}}$  is smaller than, or equal to, the number of variables  $N_{\text{var}}$ , the orthogonality of the design matrix (sampling plan)  $\mathbf{X}$ , i.e. the uncorrelatedness of variables, must inevitably be abandoned. However, in the exploration of models by statistical sampling, the correlation of the design matrix and the departure from orthogonality can be of huge importance and may have a catastrophic influence in the detection of the true active factors of the analyzed problem. In the theory of the design of experiments, where there are many random variables, the usual advice is to perform a so-called main-effect design that requires at least  $N_{\text{sim}} = N_{\text{var}} + 1$  simulations for investigating  $N_{\text{var}}$  factors (in an experiment involving  $N_{\text{var}}$  two-level factors, at least  $N_{\text{var}} + 1$  simulations are required to estimate all the main effects). On the other hand, in many applications this requirement may be wasteful or even impossible to perform (an example might be an experiment represented by extremely costly finite element simulation of a nonlinear structure). Therefore, one has to seek *optimal designs* that are as close as possible to the target correlation matrix  $\mathbf{T} = \mathbf{I}$ , i.e. the orthogonality requirement. The two norms  $\rho_{\text{rms}}$  and  $\rho_{\text{max}}$  will be minimized to achieve an optimal design.

Whenever  $N_{\text{sim}} \leq N_{\text{var}}$  the actual correlation matrix  $\mathbf{A}$  is singular (rank deficient, i.e. zero determinant). The matrix rank of the sampling plan  $\mathbf{X}$  is at most  $N_{\text{sim}} - 1$ . The rank of the corresponding estimated correlation matrix  $\mathbf{A} \propto \mathbf{X}\mathbf{X}^T$  must be equal to it and therefore we write:

$$r = \text{rank}(\mathbf{A}) = N_{\text{sim}} - 1 \quad \text{when } N_{\text{sim}} \leq N_{\text{var}}. \quad (5)$$

Every estimated correlation matrix  $\mathbf{A}$  must be symmetric and nonnegative definite (positive semidefinite–PSD), i.e. all its eigenvalues are real and nonnegative. The matrix trace (sum of the diagonal elements)

$$\text{tr}(\mathbf{A}) = N_{\text{var}} \quad (6)$$

is equal to the sum of all its eigenvalues  $\lambda_i$ ,  $i = 1, \dots, N_{\text{var}}$ . When  $N_{\text{sim}} \leq N_{\text{var}}$ , there must be  $r$  non-zero eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$  among which the matrix order  $N_{\text{var}}$  must be distributed. The remaining  $N_{\text{var}} - r + 1$  eigenvalues are zero. The determinant of  $\mathbf{A}$ , computed as the product of all eigenvalues, is zero, too. Also, all leading principal minors (subdeterminants of square submatrices along the leading diagonal) of order greater than  $r$  are zero.

In this paper, we construct correlation matrices  $\mathbf{R}$  [or  $\mathbf{M}$ ] that are optimal in the sense of  $\rho_{\text{rms}}$  [or  $\rho_{\text{max}}$  respectively]. Additionally, we construct matrices  $\mathbf{R}^M$  that minimize  $\rho_{\text{max}}$  among all possible solutions  $\mathbf{R}$ , i.e. among solutions with minimal  $\rho_{\text{rms}}$ . These optimal matrices can be viewed as the best possible matrices to achieve uncorrelatedness, i.e. the target matrices  $\mathbf{T}$  that are as close as possible to  $\mathbf{I}$ . Before doing so, selected properties of symmetric Toeplitz matrices are reviewed as they are useful in the presented derivation.

### 3.1. Notes on symmetric Toeplitz matrices

This subsection summarizes basic facts about symmetric Toeplitz matrices. This type of matrix will be involved in the solution of many of the following problems and therefore we present the common features here in a compact section. Note that this class of matrices can be viewed as shift-invariant autocorrelation matrices of a series system of elements with equal spacing, see e.g. [8].

Toeplitz matrices  $\mathbf{R}$  are matrices with  $R_{i,j} = R_{i-1,j-1}$  ( $i, j = 2, \dots, n$ ). In our case of a symmetric matrix with the unit main diagonal and elements  $R_{i,j} = [r_{|i-j|}]$ , where  $i, j = 1, \dots, N_{\text{var}}$ ;  $r_0 = 1$ , i.e. a matrix in which each descending diagonal from left to right is constant.

Symmetric Toeplitz matrices are centrosymmetric. Centrosymmetric matrices (as defined in [20, p. 142]) of order  $n$  have  $R_{i,j} = R_{n-i+1,n-j+1}$  ( $i, j = 1, \dots, n$ ), i.e. they are symmetric about their center. Centrosymmetric matrices are also called cross-symmetric [21, p. 361]. Symmetric centrosymmetric matrices are sometimes called *bisymmetric* or *doubly symmetric* matrices. Symmetric Toeplitz matrices form a special subclass of the class of doubly symmetric matrices.

Let  $\mathbf{J}$  represent the *exchange matrix* (also known as the *conjugate identity*, *reflection*, *permutation* or *flip matrix*) of order  $n$  defined by  $J_{i,j} = \delta_{i,n+1-j}$  for  $1 \leq i, j \leq n$  where  $\delta_{i,j}$  is the Kronecker delta (i.e.,  $\mathbf{J} = \mathbf{J}^{-1}$  is a matrix with ones on the cross-diagonal and zeros elsewhere).  $\mathbf{J}\mathbf{J} = \mathbf{I}$ , the identity matrix. Premultiplication of a matrix  $\mathbf{M}$  by  $\mathbf{J}$  reverses the row order of  $\mathbf{M}$ . Postmultiplication of  $\mathbf{M}$  by  $\mathbf{J}$  reverses the column order of  $\mathbf{M}$ . The properties of centrosymmetry for a matrix can be written as  $\mathbf{M}\mathbf{J} = \mathbf{J}\mathbf{M} \iff \mathbf{M} = \mathbf{J}\mathbf{M}\mathbf{J}$ .

Andrew [22, p. 158] has shown that every eigenspace of centrosymmetric matrices has a basis consisting of vectors in  $S$ , where  $S$  is the set of all vectors which are either symmetric or skewsymmetric. A vector  $\mathbf{x} = (x_1, \dots, x_n)^T$  is termed *symmetric* if  $x_i = x_{n-i+1}$  for  $i = 1, \dots, n$  and *skewsymmetric* if  $x_i = -x_{n-i+1}$ ,  $i = 1, \dots, n$ . Therefore, a centrosymmetric matrix of even (respectively odd) order  $n$  has a basis consisting of one or both of the forms, i.e. symmetric:  $[\mathbf{y}^T, (\mathbf{J}\mathbf{y})^T]^T$  (respectively  $[\mathbf{y}^T, m, (\mathbf{J}\mathbf{y})^T]^T$ ) or skewsymmetric:  $[\mathbf{y}^T, -(\mathbf{J}\mathbf{y})^T]^T$  (respectively  $[\mathbf{y}^T, m, -(\mathbf{J}\mathbf{y})^T]^T$ ). The symmetric eigenvectors satisfy  $\Phi = \mathbf{J}\Phi$  and they are associated with the eigenvalues termed *even*. The skewsymmetric eigenvectors ( $\Phi = -\mathbf{J}\Phi$ ) are associated with *odd* eigenvalues.

From a theorem stated explicitly by Cantoni and Butler [23, Theorem 2] (but clearly implicit in an earlier work of Andrew [22, Theorem 2]), a symmetric centrosymmetric matrix of order  $n$  has  $\lceil n/2 \rceil$  symmetric and  $\lfloor n/2 \rfloor$  skewsymmetric eigenvectors, which can be obtained by solving the eigenvalue problems of two matrices of these orders ( $\lceil z \rceil$  denote the smallest integer  $\geq z$ , and  $\lfloor z \rfloor$  denote the largest integer  $\leq z$ ). It is also known that a repeated eigenvalue of a real symmetric Toeplitz matrix must be both even and odd [24]. A very clear description of the computational procedure for the two eigenproblems is presented in [25].

#### 4. Lower bounds on $\rho_{\text{rms}}$ and $\rho_{\text{max}}$ when $N_{\text{sim}} = N_{\text{var}}$

The correlation error behavior drastically changes when the sample size exceeds the number of variables  $N_{\text{var}}$ . Therefore, it is very important to study the best possible performance for the crossover sample size, i.e. when:

$$N_{\text{sim}} = N_{\text{var}} = N. \quad (7)$$

As will be shown below, in this case the optimal correlation matrices  $\mathbf{R}$  and  $\mathbf{M}$  match and therefore they are also equal to  $\mathbf{R}^M$ . We denote them simply by  $\mathbf{R}$ . Every  $\mathbf{R}$  matrix is singular and has  $r = N - 1$  nonzero eigenvalues.

A realization of a correlation matrix of order  $N_{\text{var}}$  can be viewed as a point in the  $N_c$  dimensional space of all  $N_c$  different correlation coefficients. The volume of the space of all symmetric matrices with off-diagonal entries ranging from  $-1$  to  $+1$  is  $V = 2^{N_c}$ . It is known [26] that the set of all PSD correlation matrices is a solid body in that space occupying a region in the vicinity of the origin (a point corresponding to mutual uncorrelatedness). Fig. 1 illustrates the situation for  $N_{\text{var}} = 3$ . The visualized boundary of the set of positive definite matrices consists of semi positive definite matrices.

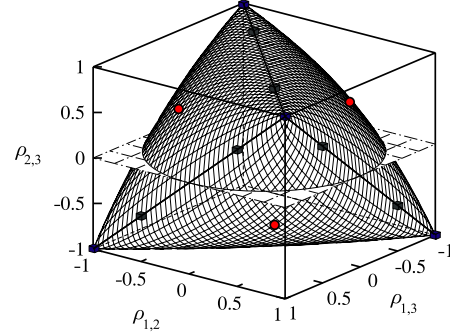


Fig. 1. Boundary of the set of all 3-dimensional positive definite correlation matrices (horizontal planar cuts are ellipses, too). Solid cubes and circles represent attainable correlation matrices when  $N_{\text{sim}} = N_{\text{var}} = 3$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Let us note that the proportion of the PSD correlation matrices in the volume  $V$  is quickly decreasing: it equals one when  $N_{\text{var}} = 2$ ,  $\pi^2/16 \approx 0.617$  when  $N_{\text{var}} = 3$ ,  $0.183$  when  $N_{\text{var}} = 4$ , then for  $N_{\text{var}} = 5, \dots, 10$  we have approximately  $2.2 \cdot 10^{-2}$ ,  $9.5 \cdot 10^{-4}$ ,  $1.3 \cdot 10^{-5}$ ,  $5.5 \cdot 10^{-8}$ ,  $6.4 \cdot 10^{-11}$ , and  $1.9 \cdot 10^{-14}$ .

We seek such a realization  $\mathbf{R}$  of  $\mathbf{A}$  (a point in the  $N_c$ -dimensional space) that is closest to the origin, yet represents a singular matrix and therefore remains on the boundary between positive definite matrices and negative definite (invalid) matrices. A good choice is to try to minimize both  $\rho_{\text{rms}}$  and  $\rho_{\text{max}}$  simultaneously. The latter means that we try to construct a singular matrix with all off-diagonal entries  $A_{i,j}$  so that  $|A_{i,j}| = C$  and we also try to minimize the absolute value  $C$ . Such an ideal matrix  $\mathbf{R}$  yields  $\rho_{\text{rms}} = \rho_{\text{max}} = C$  and will be constructed next using its eigenvalues and eigenvectors.

The spectral representation (principal component analysis) of any correlation matrix  $\mathbf{A}$  (and therefore also  $\mathbf{R}$ ) reads:

$$\mathbf{A} = \Phi \Lambda \Phi^T = \Phi \Lambda \Phi^{-1} \quad (8)$$

where the eigenvectors (columns of  $\Phi$ ) are orthonormal:

$$\Phi \Phi^T = \mathbf{I} = \Phi^T \Phi \quad (9)$$

i.e. each column (eigenvector) and row are normalized to have unit Euclidean length:

$$\sum_{k=1}^N \phi_{i,k}^2 = \sum_{k=1}^N \phi_{k,j}^2 = 1, \quad i, j = 1, \dots, N \quad (10)$$

and satisfy pairwise orthogonality:

$$\sum_{k=1}^N \phi_{i,k} \phi_{j,k} = \sum_{k=1}^N \phi_{k,i} \phi_{k,j} = 0, \quad i \neq j \quad (11)$$

and the diagonal eigenvalue matrix  $\Lambda$  contains eigenvalues  $\lambda_i$ ,  $i = 1 \dots N$ . Let us order the eigenvalues of  $\mathbf{A}$  so that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . The smallest eigenvalue  $\lambda_N = 0$ . Let us now take the remaining  $N - 1$  eigenvalues to be identical (from reasons described below) and call the given  $\mathbf{A}$  matrix  $\mathbf{R}$ :

$$\lambda_i = \frac{N}{N-1}, \quad i = 1, \dots, N-1 \quad (12)$$

and the last eigenvector corresponding to the zero eigenvalue then has the coordinates

$$\phi_{i,N} = \pm \sqrt{1/N}, \quad i = 1, \dots, N. \quad (13)$$

The reason for this is that the diagonal entries of  $\mathbf{R}$  must equal one. From Eq. (8) they can be written as:

$$R_{i,i} = \sum_{k=1}^N \lambda_k \phi_{i,k}^2, \quad i = 1, \dots, N. \quad (14)$$

Using the knowledge of all eigenvalues, we write

$$R_{i,i} = 1 = \lambda \sum_{k=1}^{N-1} \phi_{i,k}^2 = \frac{N}{N-1} \sum_{k=1}^{N-1} \phi_{i,k}^2 \quad (15)$$

and therefore for any  $i$ th component of the eigenvector  $\phi_{i,N}$  it must hold that:

$$\frac{N-1}{N} = \sum_{k=1}^{N-1} \phi_{i,k}^2 = 1 - \phi_{i,N}^2 \quad (16)$$

resulting in:  $\phi_{i,N}^2 = 1/N$ . The unit length of the last eigenvector is satisfied (Eq. (10)).

Expansion of an arbitrary off-diagonal term of  $\mathbf{R}$  based on Eq. (8) yields:

$$R_{i,j} = \sum_{k=1}^N \lambda_k \phi_{i,k} \phi_{j,k} \quad (17)$$

where the last eigenvalue is zero and the rest of them are equal and known (Eq. (12)):

$$R_{i,j} = \lambda \sum_{k=1}^{N-1} \phi_{i,k} \phi_{j,k} = \frac{N}{N-1} \sum_{k=1}^{N-1} \phi_{i,k} \phi_{j,k}. \quad (18)$$

The sum can be computed using Eqs. (11) and (13):

$$\sum_{k=1}^{N-1} \phi_{i,k} \phi_{j,k} + \underbrace{\phi_{i,N} \phi_{j,N}}_{\pm 1/N} = 0. \quad (19)$$

Therefore an arbitrary off-diagonal term in Eq. (18) equals:

$$R_{i,j} = \frac{N}{N-1} \left( \pm \frac{1}{N} \right) = \pm \frac{1}{N-1}. \quad (20)$$

This completes the derivation of lower bounds on  $\rho_{rms}$  and  $\rho_{max}$ :

$$\begin{aligned} \rho_{rms} &\geq \frac{1}{N-1} \\ \rho_{max} &\geq \frac{1}{N-1}. \end{aligned} \quad (21)$$

No algorithm for correlation control can perform better than that. Moreover, these bounds are not always possible to match due to the limited number of attainable correlations, see [19]. These limitations become even more severe with increasing  $N_{var}$  dimension.

In the case of the three random variables, four similar correlation matrices happen to equal the bounds; see the red circles in Fig. 1. There are four such optimal solutions  $\mathbf{R}$  corresponding to the points in the middle of four convex surfaces of the tetrahedron-like boundary shape (the surface has four sharp vertices and contains six straight lines that connect them). However, it can be seen that the number of attainable correlations is very limited and in higher dimensions there are no such points on the surface in the optimal directions dictated by equal correlation magnitudes. In reality, therefore,  $\rho_{rms}$  is closer to the lower bound in Eq. (21) than the conservative measure  $\rho_{max}$ .

Recalling the entries of an optimal matrix  $\mathbf{R}$  that are featured in Eq. (20), one of the possible patterns that matches the lower bound is as follows:

$$R_{i,j} = (-1)^{i-j+1} \frac{1}{N_{var}-1}, \quad i \neq j. \quad (22)$$

Another example of an optimal matrix  $\mathbf{R}$  can be obtained by requiring the same off-diagonal terms of  $\mathbf{R}$ , i.e.  $R_{i,j} = \rho$ . Matrix  $\mathbf{R}$  can be written as

$$\mathbf{R} = \mathbf{I} + \rho (\mathbf{F} - \mathbf{I}) \quad (23)$$

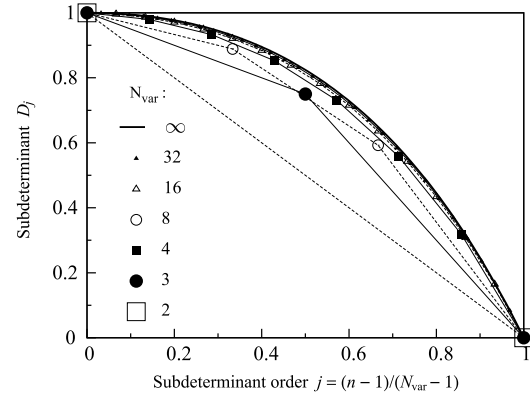


Fig. 2. Convergence of sub-determinants of matrices in (Eqs. (22) and (24)) towards Eq. (28).

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{F}$  is an  $N_{var} \times N_{var}$  matrix of ones. The eigenvalues of such an  $\mathbf{R}$  are  $(1 - \rho) + \rho N_{var}$  (with multiplicity 1) and  $(1 - \rho)$  (with multiplicity  $N_{var} - 1$ ). Setting the former eigenvalue to zero yields  $\rho = R_{i,j}$  for  $i, j = 1, \dots, N_{var}$ :

$$R_{i,j} = -\frac{1}{N_{var}-1}, \quad i \neq j. \quad (24)$$

An interesting fact about both solutions (Eqs. (22) and (24)), is that all submatrices of the order  $n \leq N_{var}$  along the principal diagonal have the following eigenvalues:

$$\lambda_i = \begin{cases} \frac{N_{var}}{N_{var}-1} & i = 1, \dots, n-1 \\ \frac{N_{var}-n}{N_{var}-1} & i = n. \end{cases} \quad (25)$$

That is why the leading principal minors (principal subdeterminants) read:

$$D_n = \prod_{i=1}^n \lambda_i = \left( \frac{N_{var}}{N_{var}-1} \right)^{n-1} \frac{N_{var}-n}{N_{var}-1} \quad (26)$$

which is a decreasing function. To compare it for various dimensions  $N_{var}$  we scale the principal subdeterminant order  $n$  from  $1 \dots N_{var}$  to a nondimensional order  $j = (n - 1)/(N_{var} - 1)$  so that the first leading principal minor has  $j = 0$  and the last one has  $j = 1$ . In other words,  $n = j(N_{var} - 1) + 1$ . The above equation then reads

$$D_j = (1-j) \left( \frac{N_{var}}{N_{var}-1} \right)^{j(N_{var}-1)}. \quad (27)$$

The asymptotic decreasing function of subdeterminants is

$$\lim_{N_{var} \rightarrow \infty} D_j = (1-j) \exp(j). \quad (28)$$

This limiting function is plotted in Fig. 2 together with examples of the leading principal minors for selected dimensions  $N_{var}$ .

### 5. Lower bounds on $\rho_{rms}$ and $\rho_{max}$ when $N_{sim} = 2$

The optimal correlation matrix  $\mathbf{R}$  has only one nonzero eigenvalue  $\lambda_1 = N_{var}$  and corresponding eigenvector with coordinates  $\pm 1/\sqrt{N_{var}}$ . That is why in Eq. (17) any  $A_{i,j} = \pm 1$  which also follows from the sample correlation formulas. The lower bounds on both  $\rho_{rms}$  and  $\rho_{max}$  do not depend on  $N_{var}$  and they equal the unit upper bound:

$$\rho_{rms} = \rho_{max} = 1. \quad (29)$$

**6. Lower bound on  $\rho_{rms}$  for a general case of  $N_{sim} \leq N_{var}$**

The two limiting cases of  $N_{sim} = 2$  and  $N_{sim} = N_{var}$  are solved. We now present the solution to the general problem of  $N_{sim} \leq N_{var}$ . First consider an optimal correlation matrix  $\mathbf{R}$  (optimal in the  $\rho_{rms}$  sense). Using Eq. (4) the minimized norm can be rewritten as:

$$\rho_{rms}(\mathbf{R}) = \sqrt{\frac{\sum_{i=1}^{N_{var}} \sum_{j=1}^{N_{var}} R_{i,j}^2 - N_{var}}{N_{var} (N_{var} - 1)}} \quad (30)$$

(subtract the  $N_{var}$  unit diagonal entries from the sum of squared correlations, then divide it by the number of the off-diagonal entries and then take the square root). It is known that the square of (the optimal) correlation matrix has a trace equal to the sum of the squared eigenvalues. We can truncate the sum after considering the  $r$  nonzero eigenvalues only:

$$\text{tr}(\mathbf{R}^2) = \sum_{i=1}^{N_{var}} \lambda_i^2 = \sum_{i=1}^r \lambda_i^2. \quad (31)$$

Note also that each  $i$ th diagonal entry of  $\mathbf{R}^2$  represents the sum of squared correlations from the  $i$ th row (or column) of  $\mathbf{R}$ . Therefore, we can also write:

$$\text{tr}(\mathbf{R}^2) = \sum_{i=1}^{N_{var}} \sum_{j=1}^{N_{var}} R_{i,j}^2. \quad (32)$$

This sum also appears in Eq. (30). The correspondence between Eqs. (32) and (31) can be used in Eq. (30). Thus, minimization of  $\rho_{rms}$  is equivalent to the minimization of the sum of the  $r$  squared (nonzero) eigenvalues, i.e. eigenvalues whose sum is  $N_{var}$ . This implies that optimality in terms of  $\rho_{rms}$  is achieved for equal nonzero eigenvalues, i.e. the trace  $\text{tr}(\mathbf{R}) = N_{var}$  is distributed uniformly over the  $r$  nonzero eigenvalues:

$$\lambda_i = \begin{cases} \frac{N_{var}}{r} & \text{when } i = 1, \dots, r \\ 0 & \text{when } i = r + 1, \dots, N_{var}. \end{cases} \quad (33)$$

Thus,  $\text{tr}(\mathbf{R}^2) = N_{var}^2/r$  and inserting this result into Eq. (30) finally yields the lower bound for  $\rho_{rms}$ , it being the norm for optimal singular correlation matrices  $\mathbf{R}$ :

$$\begin{aligned} \rho_{rms} \Big|_{N_{var} \geq N_{sim}} &\geq \rho_{rms}(\mathbf{R}) = \sqrt{\frac{N_{var} - r}{(N_{var} - 1)r}} \\ &= \sqrt{\frac{N_{var} - (N_{sim} - 1)}{(N_{var} - 1)(N_{sim} - 1)}}. \end{aligned} \quad (34)$$

It can be checked easily that the above obtained results in Eq. (21) (and Eq. (29)) for the limiting cases  $N_{var} = N_{sim}$  (and  $N_{sim} = 2$  respectively) are recovered. Graphs of this lower bound for selected problem dimensions  $N_{var}$  are plotted in Fig. 3.

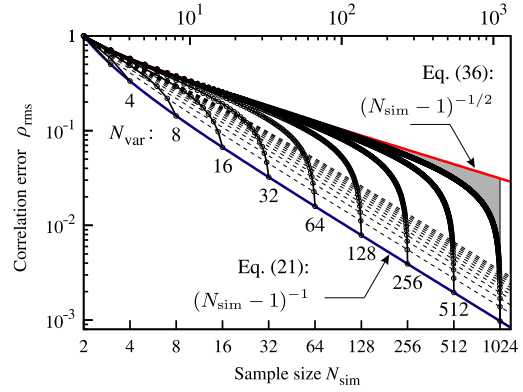
Note that for any given nonnegative difference  $d = N_{var} - N_{sim}$ , the lower bound reads

$$\sqrt{\frac{d + 1}{(N_{sim} + d - 1)(N_{sim} - 1)}} \quad (35)$$

and thus the error remains asymptotically inversely proportional to the sample size, see the dashed lines in Fig. 3.

Note also that the optimal value of  $\rho_{rms}(\mathbf{R})$  in Eq. (34) never exceeds the following bound:

$$\rho_{rms}(\mathbf{R}) \leq \frac{1}{\sqrt{(N_{sim} - 1)}} \quad (36)$$



**Fig. 3.** Lower bound on correlation error  $\rho_{rms}$  (errors for matrices  $\mathbf{R}$ ) given in Eq. (34). The shaded area represents the range between the upper and lower bounds on correlation error when  $N_{sim} = 1024$ .

see Fig. 3. This corresponds to the average performance for the case when sample ordering in  $\mathbf{X}$  is left random, see [19] for details.

It can be shown that the following equality holds between the multiple and power of the optimal  $\mathbf{R}$  matrix:

$$\lambda_1 \mathbf{R} = (\mathbf{R})^2 \quad (37)$$

which, by induction, implies

$$(\lambda_1)^{p-1} \mathbf{R} = (\mathbf{R})^p \quad (38)$$

where  $p$  is an arbitrary natural power. Since the two matrices on either side of the equation match, their traces must also match:

$$\text{tr}((\lambda_1)^{p-1} \mathbf{R}) = \text{tr}(\mathbf{R}^p). \quad (39)$$

The trace on the right hand side reads:  $\text{tr}(\mathbf{R}^p) = \sum_{i=1}^r (\lambda_i)^p$ . Therefore, Eq. (39) can be rewritten as:

$$(\lambda_1)^{p-1} N_{var} = r \cdot (\lambda_1)^p \quad (40)$$

which is true because the eigenvalue has been found to equal  $\lambda_1 = N_{var}/r$ .

It is clear now that the error  $\rho_{rms}$  of any actual correlation matrix  $\mathbf{A}$  resulting from the proposed correlation control algorithm [14,15] lies between the lower bound given in Eq. (34) and the upper bound given in Eq. (36) which corresponds to the average error  $\rho_{rms}$  in random ordering [19], see Fig. 3. What remains unclear is (i) whether the optimal solution can be numerically achieved and (ii) whether such optimal correlation matrix  $\mathbf{R}$  is unique. The first question can be answered by the analysis of attainable values of correlation coefficients which was performed in [19]. The answer to the second question follows. Numerically, the optimal correlation matrix can be obtained by solving its eigenvectors  $\Phi$  given the spectrum of eigenvalues  $\Lambda$  from Eq. (33) and requesting (i)  $\Phi \Phi^T = \Phi^T \Phi = \mathbf{I}$  and (ii) unit main diagonal of  $\mathbf{R} = \Phi \Lambda \Phi^T$ . Indeed, the solution is not unique. For concreteness, we present two equally good (optimal) correlation matrices constructed for  $N_{var} = 4$  variables and  $N_{sim} = 3$  simulations that correspond to the same Euclidean distance from the origin in the space of all  $N_c$  correlations:

$$\begin{pmatrix} \nabla \mathbf{R}_{4,3} \\ \Delta \mathbf{R}_{4,3}^M \end{pmatrix} = \begin{pmatrix} 1 & 0.9 & 0 & -0.4359 \\ \sqrt{2}/2 & 1 & 0.4359 & 0 \\ 0 & \sqrt{2}/2 & 1 & 0.9 \\ -\sqrt{2}/2 & 0 & \sqrt{2}/2 & 1 \end{pmatrix}. \quad (41)$$

The common feature of all optimal correlation matrices  $\mathbf{R}$  is that the sum of squared correlations in any row (column) equals the nonzero eigenvalue  $N_{var}/r$ . After subtracting the diagonal unit, we

have  $(N_{\text{var}} - r) / r$ , which must be distributed over the remaining  $N_{\text{var}} - 1$  entries in that row (column):

$$\sum_{i=1, i \neq j}^{N_{\text{var}}} R_{i,j}^2 = \lambda_1 - 1 = \frac{N_{\text{var}} - r}{r}. \quad (42)$$

So, the average sum of squares in each row is  $(N_{\text{var}} - r) / [r(N_{\text{var}} - 1)]$ , which is exactly what appears in squared Eq. (34). It means that the average squared correlation of any variable with the remaining  $N_{\text{var}} - 1$  variables is identical to the average of all squared correlations in  $\mathbf{R}$ .

There are, however, infinitely many matrices  $\mathbf{R}$  represented by points of intersection of (i), a hypersphere with a radius of  $\rho_{\text{rms}}(\mathbf{R}) \times \sqrt{N_c}$  in the  $N_c$ -dimensional space of all correlations with (ii), the boundary of the set of all positive definite correlation matrices that contains all singular correlation matrices with rank  $r$  (an analogous surface to that in Fig. 1). Are there some solutions that are better than others among these? For example, the two matrices  $\mathbf{R}_{4,3}$  and  $\mathbf{R}_{4,3}^M$  are equally good with respect to  $\rho_{\text{rms}}$ , but  $\mathbf{R}_{4,3}^M$  has a better  $\rho_{\text{max}} = \sqrt{2}/2 \approx 0.707$ . Frequently, the sampling plan  $\mathbf{X}$  cannot be  $\rho_{\text{rms}}$ -optimal and  $\rho_{\text{max}}$ -optimal simultaneously. The problem is that we seek the nearest point to the origin on a part of the surface described earlier, but we generally cannot take the directions given by equal correlation magnitudes as can be done for  $N_{\text{var}} = N_{\text{sim}}$ . A  $\rho_{\text{rms}}$ -optimal design may not be satisfactorily under the  $\rho_{\text{max}}$  criterion, and vice versa. Hence, to construct a satisfactory sampling plan, some trade off between  $\rho_{\text{rms}}$  and  $\rho_{\text{max}}$  should be considered. A good strategy seems to be to find an  $\rho_{\text{rms}}$ -optimal sampling plan with a small  $\rho_{\text{max}}$  (e.g. to prefer  $\mathbf{R}_{4,3}^M$  over  $\mathbf{R}_{4,3}$ ).

The task is now to find  $\mathbf{R}$ -matrices minimizing  $\rho_{\text{max}}$ . These matrices will be denoted as  $\mathbf{R}^M$  from here on.

Let us illustrate the problem for the above mentioned case of  $N_{\text{sim}} = 3$  and  $N_{\text{var}} = 4$ . Generally, the correlation matrix has six entries in the lower triangle:

$$\mathbf{R} = \begin{pmatrix} 1 & a & b & -c \\ a & 1 & c & b \\ b & c & 1 & a \\ c & e & f & 1 \end{pmatrix}. \quad (43)$$

Imposing the equality condition of sums of squared correlations in any row and any column, one finds that the absolute values  $|a|$  and  $|f|$  must match. Similarly,  $|b|$  and  $|e|$  must match. Finally,  $|c|$  and  $|d|$  also must match. The matrix simplifies into only several possible patterns (combinations of signs), one of which is shown in the upper triangle. The vector size  $1 + a^2 + b^2 + c^2 = N_{\text{var}}/r = 2$  must be fulfilled. Next, we request that the third leading principal minor be zero:  $1 + 2abc - (a^2 + b^2 + c^2) = 0$ . Since the sum in the parentheses equals one, at least one of  $a$ ,  $b$  or  $c$  must be zero. One of the possible solutions is the set  $a = 0$ ,  $b = \pm\sqrt{1 - c^2}$  (a circle in  $b, c$ ). There are another two circles corresponding to  $b = 0$  and  $c = 0$ . All of these solutions lead to the zero determinant of the whole matrix. Then, further minimization of  $\rho_{\text{max}}$  implies the condition  $c^2 = 1 - c^2$ , i.e.  $c = \pm\sqrt{2}/2$ . An equivalent solution has been found in  $\mathbf{R}_{4,3}^M$ . Note that many solutions exist equivalent to  $\mathbf{R}_{4,3}^M$ , corresponding to those points on the circles where the angle between variables with nonzero correlations is  $\pi/4$ .

The task of the numerical search of optimal  $\mathbf{R}^M$  matrices for the general dimensions  $N_{\text{var}}$  and  $N_{\text{sim}}$  can be viewed as an inverse eigenvalue problem. It concerns the reconstruction of a matrix from prescribed spectral data (information about eigenvalues or eigenvectors). The objective is to construct a matrix that maintains a certain specific structure as well as that given spectral property. One can employ optimization algorithms to attack the problem.

We have developed a simple iterative technique to find a matrix with the desired eigenvalues (diagonal matrix  $\mathbf{\Lambda}$ ). Suppose we

have a symmetric matrix  $\mathbf{R}_{(i)}$  with ones on the main diagonal in the  $i$ th iteration. Then we find its eigenvector matrix  $\Phi_{(i)}$  and the diagonal eigenvalue matrix  $\mathbf{\Lambda}_{(i)}$  so that  $\mathbf{R}_{(i)} = \Phi_{(i)}\mathbf{\Lambda}_{(i)}\Phi_{(i)}^T$ . Then, we calculate a new iteration by (i) replacing the actual eigenvalues with the target ones:  $\mathbf{R}_{(i+1)} = \Phi_{(i)}\mathbf{\Lambda}_{(i)}^T$  and (ii) setting the diagonal entries of  $\mathbf{R}_{(i+1)}$  to ones. The iterations are repeated until the eigenvalues of  $\mathbf{R}_{(i)}$  match the requested  $\mathbf{\Lambda}$ . By perturbing the initial symmetric matrix  $\mathbf{R}_{(0)}$  one can find an optimal solution  $\mathbf{R}^M$  that has the prescribed eigenvalues.

We have developed another technique to construct the optimal matrices. The technique is based on analogy with a simple mechanical system. This is described next.

## 7. Mechanical analogy for $\mathbf{R}$ , $\mathbf{R}^M$ and $M$ matrices when $N_{\text{sim}} \leq N_{\text{var}}$

The correlation matrices  $\mathbf{R}$  that attain the lower bound on  $\rho_{\text{rms}}$  in a general case of  $N_{\text{sim}} \leq N_{\text{var}}$  can be shown to be nicely analogous with a simple mechanical system. This can be explained in a particularly illustrative way.

The entries of the correlation matrix can be viewed as the cosines of angles between random variables. Zero correlation corresponds to a right angle (uncorrelatedness) while the extreme correlation of  $\pm 1$  corresponds to angles of zero or  $180^\circ$  (zero or  $\pi$  radians).

Now imagine a system of  $N_{\text{var}}$  rigid bars all connected by a single hinge located in the origin of the coordinate system. These bars represent random variables. All pairs of these bars (directions) are also connected by perfectly elastic nonlinear rotational springs, all with the same constitutive law.

Minimization of  $\rho_{\text{rms}}$  corresponds to the minimization of a sum of squared cosines (correlations). Similarly, the total stored energy of the system is obtained as a sum of the energies accumulated in all springs. Therefore, each rotational spring connecting variables  $i$  and  $j$  contributes to the sum of the stored elastic energy by  $E_{i,j}$ :

$$E_{i,j} = \cos^2(\alpha_{i,j}) \quad (44)$$

where  $\alpha_{i,j}$  is the angle between this pair. The number of pairs (rotational springs) is  $N_c$ , see Eq. (2).

Once the number of bars exceeds the dimension of the space ( $r$ ), the total elastic energy (corresponding to  $\rho_{\text{rms}}$ ) cannot equal zero. The springs are stretched and the system automatically reaches a state of extremal energy. The constitutive law (moment  $M_{i,j}$  in each rotational spring acting in the plane of the pair of springs) can be derived by differentiating the energy function with respect to the corresponding angle:

$$M_{i,j} = \frac{\partial E_{i,j}}{\partial \alpha_{i,j}} = -2 \cos(\alpha_{i,j}) \sin(\alpha_{i,j}) = \sin\left[2\left(\alpha_{i,j} - \frac{\pi}{2}\right)\right]. \quad (45)$$

This equation is visualized in Fig. 4. The elastic energy stored in each such spring according to Eq. (44) is proportional to the shaded area. It increases from the zero difference between  $\alpha_{i,j}$  and the right angle  $\pi/2$  (the only case of zero energy). As can be seen in Fig. 4, the springs are highly nonlinear. The current stiffness of the springs depends on the angle and it can be calculated as

$$k_{i,j} = \frac{\partial M_{i,j}}{\partial \alpha_{i,j}} = -2 \cos(2\alpha_{i,j}) = 2 - 4 \cos^2(\alpha_{i,j}). \quad (46)$$

The dimension of the space where all these bars are interacting equals  $r = N_{\text{sim}} - 1$ . For example, the situation of  $N_{\text{sim}} = 4$  corresponds to three-dimensional space in which one can place a system with a maximum of three bars to obtain zero energy (the bars are aligned with the three orthogonal directions). Having more bars

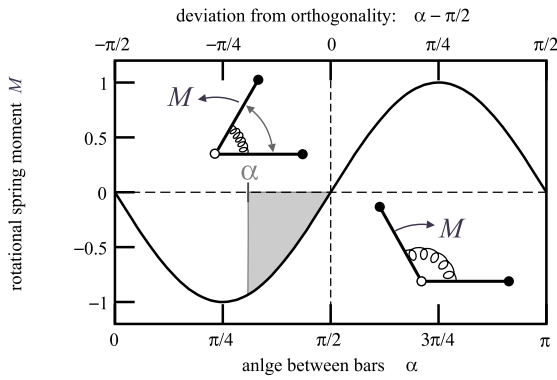


Fig. 4. Rotational spring for mechanical analogy with optimal matrices  $\mathbf{R}$ .

causes the springs to be stretched. The minimal energy that can be attained is dictated by the result in Eq. (34) and equals:

$$\Pi_{N_{\text{var}} \geq N_{\text{sim}}} = N_c \frac{N_{\text{var}} - (N_{\text{sim}} - 1)}{(N_{\text{var}} - 1)(N_{\text{sim}} - 1)} = \frac{N_{\text{var}}}{2} \frac{N_{\text{var}} - r}{r}. \quad (47)$$

Is the configuration of bars in the system that attains this minimal energy unique? No. It can be shown that there is an energetic plateau and the extremal energy can be obtained by infinitely many configurations. These configurations correspond to correlation matrices with the same spectrum of eigenvalues—such as the two examples in Eq. (41). In this example, once the configuration corresponding to cosines in  $\mathbf{R}_{4,3}$  is found, one can start pushing the pair of bars with the greatest absolute cosine of their angle towards each other to achieve  $\mathbf{R}_{4,3}^M$  while remaining at the energetic minimum (i.e., keeping  $\rho_{\text{rms}}$ ).

We have performed numerical tests with a dynamical simulation program implementing the described system. It should be noted that the system has an obvious unstable point with extremal energy that corresponds to all angles being equal to zero. A small temporary disturbance from such an unstable state leads to rearrangement into a desired configuration with minimal  $\rho_{\text{rms}}$ . In order to avoid being trapped in an unstable configuration, it is advisable to begin the simulation with a random configuration. Once the system finds a state with minimal  $\rho_{\text{rms}}$  (an  $\mathbf{R}$  matrix), one can perturb it to reach  $\mathbf{R}^M$ . In the numerical simulation program this can be achieved by temporarily slightly increasing the stiffness  $k_{i,j}$  of the spring connecting the bars with some angle  $\alpha_{i,j}$  that has the greatest absolute difference from the right angle. This adjustment makes the (damped) system slide over the energetic plateau towards the desired state (usually exhibiting some special symmetries). Fig. 5(b) illustrates the difference between a configuration corresponding to  $\mathbf{R}$  matrix optimal in the sense of  $\rho_{\text{rms}}$  (before the additional adjustment, red lines) and the configuration  $\mathbf{R}^M$  with the same  $\rho_{\text{rms}}$  and further minimized  $\rho_{\text{max}}$ . The number of variables is  $N_{\text{var}} = 5$  and the sample size is  $N_{\text{sim}} = 3$  (top) and 4 (bottom). The numerical values of the correlations corresponding to the  $\mathbf{R}$  matrices in Fig. 5(b) are:

$$\begin{pmatrix} \nabla \mathbf{R}_{5,3} \\ \nabla \mathbf{R}_{5,4} \end{pmatrix} = \begin{pmatrix} 1 & 0.7118 & 0.2149 & -0.3512 & -0.9076 \\ -0.361 & 1 & 0.8389 & 0.4076 & -0.3512 \\ -0.3383 & 0.2998 & 1 & 0.8389 & 0.2149 \\ 0.3696 & 0.6666 & -0.2906 & 1 & 0.7118 \\ 0.5342 & -0.0461 & 0.6148 & 0.0358 & 1 \end{pmatrix}. \quad (48)$$

The two corresponding matrices  $\mathbf{R}^M$  are formulated in Eq. (75).

We note that the mechanical system can easily be implemented for solving a spatial dimension  $r$  that is greater than three. Placement of the endpoints of the bars is, in fact, placing points on the “surface” of a unit hypersphere, or  $(r - 1)$ -sphere, it being

the boundary of an  $r$ -dimensional ball with unit radius, where  $r$  is the arbitrary natural dimension. The Euclidean coordinates of all  $N_{\text{var}}$  endpoints in  $r$ -dimensional space can be arranged in an  $(N_{\text{var}} \times r)$  matrix denoted  $\mathbf{B}$ . The corresponding correlation matrix is then obtained simply as:

$$\mathbf{R} = \mathbf{B}\mathbf{B}^T. \quad (49)$$

Each row  $i$  of the  $\mathbf{B}$  matrix (a point in Cartesian coordinates) can be expressed using the corresponding  $(r - 1)$  hyperspherical coordinates (angles)  $\varphi_{i,1}, \dots, \varphi_{i,r-1}$  as:

$$B_{i,j} = \begin{cases} \cos(\varphi_{i,j}) \cdot \prod_{k=1}^{j-1} \sin(\varphi_{i,k}) & \text{for } j = 1, \dots, r-1 \\ \prod_{k=1}^{j-1} \sin(\varphi_{i,k}) & \text{for } j = r. \end{cases} \quad (50)$$

Note that the last angular coordinate  $\varphi_{i,r-1}$  of any variable  $i$  ranges over  $(0; 2\pi)$  and the remaining angles range over  $(0; \pi)$ . The correlation matrix can thus be stored by using the  $N_{\text{var}} \times (r - 1)$  matrix of angular coordinates  $\varphi_{i,j}$ . Whenever  $N_{\text{sim}} < \frac{1}{2}(N_{\text{var}} + 3)$ , this is a more compact storage format than saving  $N_c$  correlation coefficients.

To illustrate this point, we present the matrices for the optimal correlation matrix  $\mathbf{R}_{3,3}^M = \mathbf{M}_{3,3}$  (three variables each represented by three simulations). This optimal matrix is visualized in two different figures, namely by a red circle in Fig. 1 and in a different space also in Fig. 5, top left. The matrix (one vector) of three angular coordinates  $\varphi_{3,3} = (0, \frac{\pi}{3}, \frac{2\pi}{3})^T$ , from which the corresponding matrices can be constructed as:

$$\mathbf{B}_{3,3} = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}, \quad \mathbf{R}_{3,3}^M = \begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}. \quad (51)$$

Matrix  $\mathbf{B}$  can be viewed as a product of the eigenvector matrix and the diagonal matrix of the square roots of the eigenvalues (recall the spectral representation in Eq. (8)):

$$\mathbf{R} = \mathbf{B}\mathbf{B}^T = \underbrace{\Phi \Lambda}_{\mathbf{B}}^{1/2} (\Lambda^{1/2} \Phi^T). \quad (52)$$

The eigenvectors corresponding to the zero eigenvalues can be ignored. In the case of  $\mathbf{R}$  matrices, the nonzero eigenvalues are identical and therefore the  $\mathbf{B}$  matrix becomes just the  $\Phi$  matrix multiplied by a scalar:

$$\mathbf{B} = \sqrt{\lambda_1} \Phi \quad (53)$$

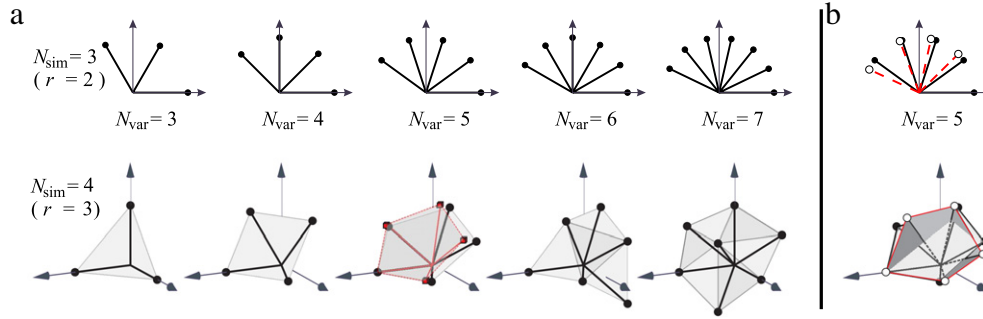
so that the rows of  $\mathbf{B}$  (rigid bars) are scaled to unit length.

In the above studied numerical example in Eq. (51), the two eigenvectors of correlation matrix  $\mathbf{R}_{3,3}^M$  were selected as

$$\Phi_{3,3} = \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \quad (54)$$

Visualizations of  $\mathbf{R}^M$  matrices in Fig. 5 has been selected so that the first row of  $\mathbf{B}$  matrix is a vector aligned with  $x$  axis, i.e. the coordinates of the first row vector read  $(1, 0, \dots, 0)$ .

Note also that when the  $\rho_{\text{max}}$  criterion is the object of minimization (to obtain  $\mathbf{M}$ ), the energy function in the mechanical



**Fig. 5.** Visualization of optimal singular correlation matrices. (a)  $\mathbf{R}^M$  (solid circles). All of these solutions are also  $\mathbf{M}$  matrices except for  $N_{\text{var}} = 5$  and  $N_{\text{sim}} = 4$  (visualized with solid boxes). (b) Examples of optimal correlation matrices  $\mathbf{R}$  (empty circles) used in Eq. (48) with those from (a). Top row:  $N_{\text{sim}} = 3$  (dimension  $r = 2$ ). Bottom row:  $N_{\text{sim}} = 4$  (dimension  $r = 3$ ).

system must be redefined accordingly. What suffices is to make the energy equal to the cosine of angle  $\alpha_{i,j}$  that has the greatest absolute deviation from the right angle. As the configuration of the damped system evolves in time, different angles may alternately be featured in the energy function. When a proper time step and damping is used, the system stabilizes at the best possible configuration. More details on the mechanical model are left for a subsequent paper.

Our studies on the optimal  $\mathbf{R}^M$  matrices suggest that such matrices form special patterns. To illustrate this, the two following subsections present general solutions to the problem for the two limiting cases, i.e. (i), when  $N_{\text{sim}} = 3$ , and (ii), when  $N_{\text{sim}} = N_{\text{var}} - 1$ .

### 8. Optimal matrices $\mathbf{R}^M$ when $N_{\text{sim}} = 3$

We have found that one particular pattern of the optimal  $\mathbf{R}^M$  matrices can be written as *symmetric Toeplitz* matrix  $\mathbf{R} = (R_{i,j})$ . In particular, it takes the following form:

$$R_{i,j}^M = \left(\frac{2}{r}\right) \cos\left(\pi \frac{i-j}{N_{\text{var}}}\right) \quad (55)$$

where the multiplier  $\left(\frac{2}{r}\right) = 1$ . Such an optimal correlation matrix has  $\rho_{\text{rms}}$  according to Eq. (34) and the absolute extreme correlation:

$$\rho_{N_{\text{sim}}=3}^{\text{max}}(\mathbf{R}^M) = \cos\left(\frac{\pi}{N_{\text{var}}}\right). \quad (56)$$

This extreme correlation appears right next to the main diagonal and negatively also in the corners of the secondary diagonal of  $\mathbf{R}^M$ , see e.g. the example  $\mathbf{R}_{4,3}^M$  in Eq. (41). It is easy to check that the sum of squared off-diagonal correlations in any row (column) equals  $(N_{\text{var}} - r)/r$ , as required in Eq. (42):

$$\sum_{j=\text{const}, i \neq j}^{N_{\text{var}}} (R_{i,j}^M)^2 = \sum_{i=1}^{N_{\text{var}}-1} \cos^2\left(\frac{i\pi}{N_{\text{var}}}\right) = \frac{N_{\text{var}} - 2}{2}. \quad (57)$$

In the above equation, we exploited the fact that

$$\kappa = \begin{cases} N_{\text{var}} \text{ odd} : 2 \sum_{i=1}^{\frac{N_{\text{var}}-1}{2}} \cos^2\left(\frac{i\pi}{N_{\text{var}}}\right) \\ N_{\text{var}} \text{ even} : 2 \sum_{i=1}^{\frac{N_{\text{var}}}{2}-1} \cos^2\left(\frac{i\pi}{N_{\text{var}}}\right) \end{cases} = \frac{N_{\text{var}} - 2}{2}. \quad (58)$$

Why are all the correlations in Eq. (55) cosines depending on  $N_{\text{var}}$ ? The mechanical model illustrates the answer. When  $N_{\text{sim}} = 3$ , one finds a way to distribute  $N_{\text{var}}$  directions in two dimensions. The solution corresponds to a regular “fan” of bars where all the pairs of consecutive bars ( $|i - j| = 1$ ) contain the angle of  $\pi/N_{\text{var}}$  (see

Eq. (56)). This angle corresponds to all the entries of the correlation matrix  $\mathbf{R}^M$  right next to the principal diagonal. The maximal angle is  $\pi(N_{\text{var}} - 1)/N_{\text{var}}$ . The cosine of this value corresponds to the correlation between the first and the last variables (the correlations in the corners on the secondary diagonal). The top row of Fig. 5 illustrates the solution for various  $N_{\text{var}}$ .

This paragraph characterizes the solution  $\mathbf{R}^M$ . The two orthonormal eigenvectors corresponding to the repeated nonzero eigenvalue of the symmetric Toeplitz matrix  $\mathbf{R}^M$  can be taken as the symmetric and skew symmetric eigenvectors (recall the results from Section 3.1). The individual coordinates  $i$  of the eigenvectors then read:

$$\phi_i^{\text{sym}} = \sqrt{\frac{2}{N_{\text{var}}}} \sin\left(\pi \frac{i - \frac{1}{2}}{N_{\text{var}}}\right) \quad (59)$$

$$\phi_i^{\text{skew}} = \sqrt{\frac{2}{N_{\text{var}}}} \cos\left(\pi \frac{i - \frac{1}{2}}{N_{\text{var}}}\right), \quad i = 1, \dots, N_{\text{var}}.$$

The eigen-properties are now explicitly given. We now examine the determinants of the suggested pattern of  $\mathbf{R}^M$ . The leading principal minor  $D_1 = 1$  and the second leading principal minor  $D_2 = \sin^2(\pi/N_{\text{var}})$ . All other subdeterminants along the leading diagonal equal zero.

Let us now study the ratio between  $\rho_{\text{max}}$  and  $\rho_{\text{rms}}$  for the optimal matrices  $\mathbf{R}^M$ :

$$\begin{aligned} \nu(N_{\text{var}}) &= \frac{\rho_{\text{max}}}{\rho_{\text{rms}}} = \frac{\cos\left(\frac{\pi}{N_{\text{var}}}\right)}{\sqrt{\frac{N_{\text{var}}-2}{2(N_{\text{var}}-1)}}} \\ &= \sqrt{2} \sqrt{\frac{N_{\text{var}}-1}{N_{\text{var}}-2}} \cos\left(\frac{\pi}{N_{\text{var}}}\right). \end{aligned} \quad (60)$$

The ratio  $\nu$  initially grows with  $N_{\text{var}}$  to attain its maximum of  $\approx 1.436$  when  $N_{\text{var}} = 17$  and then it decreases and converges towards its limit:

$$\lim_{N_{\text{var}} \rightarrow \infty} \nu = \sqrt{2} \approx 1.414 \quad (61)$$

i.e. the ratio of errors never exceeds 1.436 and reaches up to approx. 41% for large  $N_{\text{var}}$ .

### 9. Optimal matrices $\mathbf{R}^M$ when $N_{\text{sim}} = N_{\text{var}} - 1$

We have found that the optimal  $\mathbf{R}^M$  can again form a *symmetric Toeplitz* matrix with ones on the diagonal and the following off-diagonal entries:

$$R_{i,j}^M = (-1)^{i-j+1} \left(\frac{2}{r}\right) \cos\left(\pi \frac{i-j}{N_{\text{var}}}\right). \quad (62)$$



This pattern is very similar to Eq. (55). The only difference is that the signs on each row/column alternate. The multiplier  $\left(\frac{2}{r}\right) = \frac{2}{N_{\text{var}}-2}$ . Again, each row (column) features a pair of nonzero off-diagonal entries that are equal in their absolute values. Such an optimal correlation matrix has  $\rho_{\text{rms}}$  according to Eq. (34) and the absolute extreme correlation:

$$\rho_{\text{max}}^{N_{\text{sim}}=N_{\text{var}}-1}(\mathbf{R}^M) = \frac{2}{N_{\text{var}}-2} \cos\left(\frac{\pi}{N_{\text{var}}}\right). \quad (63)$$

Again, it is easy to check the validity of Eq. (42), i.e. that the sum of squared off-diagonal correlations in any row (column) equals  $(N_{\text{var}} - r)/r$ . Using Eqs. (57) and (58) we can write:

$$\sum_{i=1, i \neq j}^{N_{\text{var}}} (R_{i,j}^M)^2 = \kappa \left(\frac{2}{N_{\text{var}}-2}\right)^2 = \frac{2}{N_{\text{var}}-2}. \quad (64)$$

This happens to be the inverse value of the same sum in Eq. (57). The duality between the case of  $N_{\text{sim}} = 3$  and  $N_{\text{sim}} = N_{\text{var}} - 1$  will be shown (Section 10) to be a more general feature of  $\mathbf{R}^M$ .

Computation of eigenvectors of  $\mathbf{R}^M$  can, in this case, be greatly simplified by using the following. We will demonstrate it for even  $N_{\text{var}}$ ; the extension to the odd dimension is simple. Firstly, it is helpful to partition the matrix into four symmetric square blocks each of which has the order  $N_{\text{var}}/2$ :

$$\mathbf{R}^M = \begin{pmatrix} \mathbf{P} & \mathbf{JCJ} \\ \mathbf{C} & \mathbf{JPJ} \end{pmatrix} \quad (65)$$

and where  $\mathbf{J}$  is the flip matrix defined in Section 3.1. As shown in Lemma 3 of [23], the above matrix and

$$\begin{pmatrix} \mathbf{P} - \mathbf{JC} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} + \mathbf{JC} \end{pmatrix} = \begin{pmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \quad (66)$$

are orthogonally similar. In the above equation, entries of the square symmetric matrices  $\mathbf{K}$  and  $\mathbf{S}$  read  $(i, j = 1, \dots, N_{\text{var}}/2)$ :

$$Q_{i,j} = P_{i,j} - P_{N_{\text{var}}-i+1,j} \quad (67)$$

$$S_{i,j} = P_{i,j} + P_{N_{\text{var}}-i+1,j}.$$

Theorem 2 in [23] shows that the  $N_{\text{var}}/2$  skew symmetric eigenvectors  $\phi_j$  of  $\mathbf{R}^M$  can be determined by solving the equation

$$\underbrace{(\mathbf{P} - \mathbf{JC})}_{\mathbf{K}} \mathbf{u}_j = \lambda_j \mathbf{u}_j \quad (68)$$

in which we are only interested in  $\frac{1}{2}(N_{\text{var}} - 2)$  orthonormal eigenvectors  $\mathbf{u}_j$  corresponding to the nonzero eigenvalues  $\lambda_j = N_{\text{var}}/(N_{\text{var}} - 2), j = 1, 2, \dots, \frac{1}{2}(N_{\text{var}} - 2)$ . From this solution, all the skew symmetric eigenvectors of  $\mathbf{R}^M$  can be constructed by adding negative reflected coordinates  $\mathbf{u}_j$  and normalizing the resulting vectors:

$$\phi_j^{\text{skew}} = \frac{1}{\sqrt{2}} (\mathbf{u}_j, -\mathbf{J}\mathbf{u}_j)^T. \quad (69)$$

The symmetric eigenvectors can be constructed similarly from the solution of the following half-sized problem:

$$\underbrace{(\mathbf{P} + \mathbf{JC})}_{\mathbf{S}} \mathbf{v}_j = \lambda_j \mathbf{v}_j. \quad (70)$$

Here, we again use the  $\frac{1}{2}(N_{\text{var}} - 2)$  eigenvectors  $\mathbf{v}_j$  with nonzero eigenvalues  $\lambda_j = N_{\text{var}}/(N_{\text{var}} - 2)$  to construct the symmetric eigenvectors as:  $\phi_j^{\text{sym}} = 1/\sqrt{2} (\mathbf{v}_j, \mathbf{J}\mathbf{v}_j)^T$ . In our case, however, the following equality holds for the two half-sized submatrices:

$$\mathbf{S} = (\mathbf{P} + \mathbf{JC}) = \mathbf{J}(\mathbf{P} - \mathbf{JC})\mathbf{J} = \mathbf{JKJ} \quad (71)$$

meaning that the matrix in Eq. (70) is, in fact, the matrix from the problem in Eq. (68) flipped along the secondary diagonal. All the

orthonormal eigenvectors  $\mathbf{v}_j$  can be taken as  $\mathbf{J}\mathbf{u}_j, j = 1, 2, \dots, \frac{1}{2}(N_{\text{var}} - 2)$ . Therefore, the symmetric eigenvectors of  $\mathbf{R}^M$  read:

$$\phi_j^{\text{sym}} = \frac{1}{\sqrt{2}} (\mathbf{J}\mathbf{u}_j, \mathbf{u}_j)^T. \quad (72)$$

To conclude, the set of  $\frac{1}{2}(N_{\text{var}} - 2)$  orthonormal eigenvectors of matrix  $\mathbf{R}^M$  that correspond to the repeated eigenvalue  $\lambda = N_{\text{var}}/(N_{\text{var}} - 2)$  can be obtained as a collection of the skew symmetric vectors in Eq. (69) and of the symmetric eigenvectors in Eq. (72). Therefore, it suffices to find eigenvectors of  $\mathbf{K}$  in Eq. (68) from which the eigenvectors of  $\mathbf{R}^M$  can be easily constructed. The eigenvectors  $\mathbf{u}_j$  of  $\mathbf{K}$  that correspond to the repeated eigenvalue can be written, for example, as:

$$u_{i,j} = \begin{cases} 0 & i > j \\ d_j & i = j \\ c_j \cdot (-1)^{i-j+1} \sin\left(\pi \frac{i-\frac{1}{2}}{N_{\text{var}}}\right) & i < j. \end{cases} \quad (73)$$

In other words, they form a lower triangular matrix with the diagonal elements  $d_j$ . The components of the eigenvector corresponding to the zero eigenvalue read:

$$u_{i,N_{\text{var}}/2} = \sqrt{\frac{2}{N_{\text{var}}}} (-1)^i \sin\left(\pi \frac{i-\frac{1}{2}}{N_{\text{var}}}\right). \quad (74)$$

The pairs  $d_j$  and  $c_j$  are found from the conditions of orthonormality;  $d_j \in (\sqrt{2}, 1)$  and  $c_j \in (0, \sqrt{2})$ .

To give an example that directly compares the matrices for  $N_{\text{sim}} = 3$  and  $N_{\text{sim}} = N_{\text{var}} - 1$ , we present two centrosymmetric Toeplitz matrices  $\mathbf{R}^M$  for  $N_{\text{var}} = 5$ . In particular, the lower triangle presents the situation when  $N_{\text{sim}} = 4$  and the upper triangle  $N_{\text{sim}} = 3$  (studied in the previous section). The constants are:  $a = \cos(\pi/5) = \frac{1}{4}(1 + \sqrt{5})$ ,  $b = \cos(2\pi/5)$ ,  $c = a \cdot 2/3$ ,  $d = b \cdot 2/3$ :

$$\begin{pmatrix} \nabla \mathbf{R}_{3,3}^M \\ \mathbf{R}_{5,4}^M \end{pmatrix} = \begin{pmatrix} 1 & a & b & -b & -a \\ c & 1 & \ddots & \ddots & -b \\ -d & \ddots & 1 & \ddots & b \\ -d & \ddots & \ddots & 1 & a \\ c & -d & -d & c & 1 \end{pmatrix}. \quad (75)$$

A noticeable fact is that the sorted absolute correlations for  $N_{\text{sim}} = 3$  and  $N_{\text{sim}} = N_{\text{var}} - 1$  are identical up to a constant ratio of 2/3. This fact is going to be discussed next in Section 10.

Again, the ratio between  $\rho_{\text{max}}$  and  $\rho_{\text{rms}}$  for the optimal matrices  $\mathbf{R}^M$  reads:

$$\frac{\rho_{\text{max}}}{\rho_{\text{rms}}} = \frac{\frac{2}{N_{\text{var}}-2} \cos\left(\frac{\pi}{N_{\text{var}}}\right)}{\sqrt{\frac{2}{(N_{\text{var}}-1)(N_{\text{var}}-2)}}} = \nu(N_{\text{var}}). \quad (76)$$

The ratio is strictly identical to that in Eq. (60) obtained for  $N_{\text{sim}} = 3$ . Therefore, the limit given in Eq. (61) for large  $N_{\text{var}}$  holds.

### 10. Optimal matrices $\mathbf{R}$ , $\mathbf{R}^M$ and $\mathbf{M}$ for arbitrary $N_{\text{sim}} < N_{\text{var}}$

The optimal matrices  $\mathbf{R}^M$  for arbitrary sample sizes  $N_{\text{sim}} \in (3; N_{\text{var}} - 1)$  do not seem to follow any general simple pattern.

We have found, however, that, for a given dimension  $N_{\text{var}}$ , there is a special relationship between pairs of solutions  $\mathbf{R}^M$  for sample sizes

- $N_{\text{sim}} = n$  and a somewhat associated sample size
- $N_{\text{sim}} = N_{\text{var}} - n + 2$ .

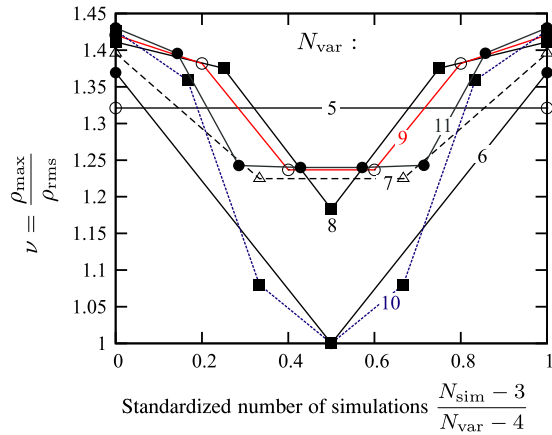


Fig. 6. Ratio  $\rho_{\max}/\rho_{\text{rms}}$  for the optimal correlation matrices  $\mathbf{R}^M$ .

Table 1

Values of  $\rho_{\max}$  for the optimal matrices  $\mathbf{M}$  (numerator) and  $\mathbf{R}^M$  (denominator). If no fraction is used, the two values of  $\rho_{\max}$  match.

$N_{\text{sim}}$	$N_{\text{var}}$	3	4	5	6	7	8	9	10
2	1	1	1	1	1	1	1	1	1
3	0.5	0.70711	0.80901	0.86603	0.90097	0.92388	0.93969	0.95106	
4		0.33333	$\frac{0.44721}{0.53934}$	0.44721	0.57735	$\frac{0.64759}{0.67112}$	$\frac{0.66936}{0.69094}$	$\frac{0.68615}{0.69195}$	
5			0.25	$\frac{0.33333}{0.43301}$	$\frac{0.39039}{0.43301}$	$\frac{0.41426}{0.44721}$	$\frac{0.43426}{0.48889}$	$\frac{0.43426}{0.44096}$	
6				0.2	$\frac{0.28621}{0.36039}$	$\frac{0.33333}{0.40267}$	$\frac{0.33333}{0.39111}$	0.33333	
7					0.16667	$\frac{0.24094}{0.30796}$	$\frac{0.27825}{0.34547}$	$\frac{0.28076}{0.29397}$	
8						0.14286	$\frac{0.2}{0.26848}$	$\frac{0.23607}{0.29655}$	
9							0.125	$\frac{0.18274}{0.23776}$	
10								0.11111	

To be specific, the smaller of the two sample sizes must satisfy:  $2 \leq n < \frac{1}{2}N_{\text{var}} + 1$ . The previously obtained solutions  $\mathbf{R}^M$  for the pair  $N_{\text{sim}} = 2$  and  $N_{\text{sim}} = N_{\text{var}}$  constitute boundary solutions of a certain type between which the pairs for the above sample sizes are found, symmetrically distributed with regard to the average sample size. Another studied pair is when  $N_{\text{sim}} = 3$  and  $N_{\text{sim}} = N_{\text{var}} - 1$ .

Let us now consider the extreme absolute correlations  $\rho_{\max}(\mathbf{R}^M)$  for that pair of solutions. Assume the maximum correlation  $\rho_{\max}(\mathbf{R}^M)$  for the smaller sample  $N_{\text{sim}} = n$  is known. Then, the maximum absolute correlation for the complementary problem with the sample  $N_{\text{sim}} = N_{\text{var}} - n + 2$  is just a multiple  $\eta$  of the other one. The ratio between the two norms happens to be the ratio between the corresponding eigenvalues:

$$\eta(N_{\text{var}}, n) = \frac{\rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=N_{\text{var}}-n+2}}{\rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n}} = \frac{\lambda_1(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=N_{\text{var}}-n+2}}{\lambda_1(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n}} = \frac{n-1}{N_{\text{var}}-(n-1)} \quad (77)$$

Note that the inverse value of this ratio was featured above in Eq. (42).

The ratio  $\eta$  does not only hold for the extreme absolute correlations. We have found that it also holds for all absolute correlations in the two matrices  $\mathbf{R}^M$ . Specifically, if one takes a sorted list of all absolute correlations (angular deviations from orthogonality) in  $\mathbf{R}^M$  for a certain dimension  $N_{\text{var}}$  and sample size  $N_{\text{sim}} = n$ , the analogical list for the same dimension and  $N_{\text{sim}} = N_{\text{var}} - n + 2$  is just the  $\eta$ -multiple of the previous list. This is exemplified in Eq. (75) and there are more examples in Appendix A,

where solutions  $\mathbf{R}^M$  for small dimensions are presented, see Eqs. (86), (88), (92) and (93).

In other words, if one knows the distribution of  $N_{\text{var}}$  rigid bars from the mechanical analogy in  $(n-1)$ -dimensional space, the same number of bars in the space of dimension  $(N_{\text{var}} - n + 1)$  have, between all pairs, the same deviations from right angles multiplied by  $\eta$ .

This gives us a tool to construct the matrix  $\bar{\mathbf{R}}^M = \bar{R}_{i,j}$  for  $N_{\text{sim}} = N_{\text{var}} - n + 2$  when the matrix  $\mathbf{R}^M = R_{i,j}$  for the smaller sample size  $N_{\text{sim}} = n$  is known. How? The absolute values of entries  $R_{i,j}$  can be just replaced by  $\eta$ -multiples to obtain the absolute values of entries  $\bar{R}_{i,j}$ . It is easy to check that such a construction from known  $\mathbf{R}^M$  (which fulfils the identity in Eq. (42)) fulfils the same identity as well.

Let us now focus on the ratio  $\nu$  between  $\rho_{\max}$  and  $\rho_{\text{rms}}$ :

$$\nu(N_{\text{var}}, N_{\text{sim}}) = \frac{\rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}}}{\rho_{\text{rms}}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}}} \quad (78)$$

We have found that, for given  $N_{\text{var}}$ , this ratio attains its maximum for  $N_{\text{sim}} = 3$  and the same maximum also for  $N_{\text{sim}} = N_{\text{var}} - 1$ . These two maxima were studied in Eqs. (60) and (76). Numerical results suggest that the ratio  $\nu$  symmetrically decreases as the sample size  $N_{\text{sim}}$  approaches the midpoint of the studied interval  $N_{\text{sim}} \in (3, N_{\text{var}} - 1)$ , i.e. for  $N_{\text{sim}} = \frac{1}{2}N_{\text{var}} + 1$ . This is illustrated in Fig. 6. We remark that for the two boundary sample sizes  $N_{\text{sim}} = 2$  and  $N_{\text{sim}} = N_{\text{var}}$ , the ratio equals 1, because the two norms match, see Eqs. (29) and (21).

We now prove that the ratio  $\nu$  is identical for any pair with  $N_{\text{sim}} = n$  and  $N_{\text{sim}} = N_{\text{var}} - n + 2$ . Using Eqs. (77) and (34) yields:

$$\begin{aligned} \nu(N_{\text{var}}, N_{\text{sim}} = N_{\text{var}} - n + 2) &= \frac{\eta(N_{\text{var}}, n) \rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n}}{\rho_{\text{rms}}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=N_{\text{var}}-n+2}} \\ &= \eta(N_{\text{var}}, n) \rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n} \sqrt{\frac{(N_{\text{var}} - 1)(N_{\text{sim}} - 1)}{N_{\text{var}} - (N_{\text{sim}} - 1)}} \\ &= \rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n} \sqrt{\frac{(N_{\text{var}} - 1)(n - 1)}{N_{\text{var}} - (n - 1)}} = \frac{\rho_{\max}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n}}{\rho_{\text{rms}}(\mathbf{R}^M)_{N_{\text{var}}, N_{\text{sim}}=n}} \\ &= \nu(N_{\text{var}}, N_{\text{sim}} = n) \end{aligned} \quad (79)$$

The numerical results (numerators in Table 1) obtained for  $\mathbf{M}$  matrices suggest that neither decreasing the number of simulations  $N_{\text{sim}}$  nor increasing the dimension  $N_{\text{var}}$  can decrease the error  $\rho_{\max}$ :

$$\rho_{\max}(\mathbf{M})_{N_{\text{var}}-1, N_{\text{sim}}} \leq \rho_{\max}(\mathbf{M})_{N_{\text{var}}, N_{\text{sim}}} \leq \rho_{\max}(\mathbf{M})_{N_{\text{var}}, N_{\text{sim}}-1} \quad (80)$$

We conclude this section by formulating the final conjectured inequality: for a given stochastic dimension  $N_{\text{var}}$  and sample size  $N_{\text{sim}} < N_{\text{var}}$ , the extreme absolute correlations in the optimal matrices  $\mathbf{M}$  and  $\mathbf{R}^M$  are found within the following bounds:

$$\rho_{\text{rms}}(\mathbf{R}^M) \leq \rho_{\max}(\mathbf{M}) \leq \rho_{\max}(\mathbf{R}^M) \leq 1.436 \rho_{\text{rms}}(\mathbf{R}^M) \quad (81)$$

where the term  $\rho_{\text{rms}}(\mathbf{R}^M)$  is explicitly given in Eq. (34).

## 11. Conclusions

This paper presents a number of theoretical and numerical results concerning two norms of optimal correlation matrices in relation to correlation control in Monte Carlo type sampling and

designs of experiments. The optimal singular correlation matrices are constructed for the cases when the number of simulations  $N_{sim}$  is less than or equal to the number of random variables  $N_{var}$ . The need to use sparse data (sample) in higher dimensions is a general trend in the field of probabilistic engineering practise. The main results are:

- An explicit formula for the root mean square error  $\rho_{rms}$  of the optimal correlation matrices  $\mathbf{R}$  that minimize this error.
- Tight bounds for the error  $\rho_{max}$  of the optimal matrices  $\mathbf{R}^M$  selected from all possible  $\mathbf{R}$  matrices to subsequently minimize the absolute error  $\rho_{max}$ .
- Bounds for the error  $\rho_{max}$  of the optimal matrices  $\mathbf{M}$ , i.e. matrices that are optimal in terms of this norm.
- A mechanical analogy between the correlation matrices and a model of bars. The associated dynamical model can be used to find the optimal matrices  $\mathbf{R}$ ,  $\mathbf{R}^M$  and  $\mathbf{M}$ .
- A number of numerical examples of optimal matrices  $\mathbf{R}$ ,  $\mathbf{R}^M$  and  $\mathbf{M}$  constructed for  $N_{var} \in \{2; 9\}$ .
- Formulas for all entries of the optimal matrices  $\mathbf{R}^M$  for general problem dimension  $N_{var}$ , where the sample size  $N_{sim} = 3$ ,  $N_{sim} = N_{var} - 1$  and  $N_{sim} = N_{var}$ .

An open problem to be solved is to check and possibly prove the conjectured inequalities in Eqs. (81) and (80). Moreover, derivation of the lower bound on the error  $\rho_{max}$  of the optimal matrices  $\mathbf{R}^M$  and  $\mathbf{M}$  (similar to the bound in derived Eq. (34)) would be helpful.

**Acknowledgments**

This outcome has been achieved with the financial support of the Czech Science Foundation under project No. P105/11/1385, and from the Czech The Ministry of Education, Youth and Sports under project No. FAST-S-12-5. The support is gratefully acknowledged.

The author thanks his friend, the mathematician Tomáš Hanžl, for his valuable discussions and suggestions. A colleague, Dr. Petr Frantík, is thanked for the computer implementation of the mechanical model presented in this paper.

**Appendix. Numerical examples of optimal singular correlation matrices  $\mathbf{R}^M$  and  $\mathbf{M}$**

Patterns of the optimal correlation matrices  $\mathbf{R}^M$  for the case of:

- $N_{sim} = 3$  are given by Eq. (55);
- $N_{sim} = N_{var} - 1$  are given by Eq. (62);
- $N_{sim} = N_{var}$  are given e.g. by Eq. (22) or by Eq. (24).

Selected solutions for the  $\mathbf{M}$  matrices were already given analytically. In particular, these solutions were given for the cases when  $\mathbf{M} = \mathbf{R}^M$ , which happens for  $N_{sim} = 3$  and  $N_{sim} = N_{var}$ .

The remainder of this Appendix shows examples of  $\mathbf{R}^M$  and  $\mathbf{M}$  for the remaining cases. The dimensions of these correlation matrices range from five to nine variables.

**A.1. Solutions for  $N_{var} = 5$**

$$\mathbf{M}_{5,4} = \begin{pmatrix} 1 & a & -a & -a & a \\ & 1 & a & -a & -a \\ & & 1 & a & -a \\ & & & 1 & a \\ & & & & 1 \end{pmatrix} \tag{82}$$

which is a centrosymmetric Toeplitz matrix with the eigenvalues  $\{2, 2, 1, 0, 0\}$ . The constant  $a = 1/\sqrt{5} \approx 0.44721$ .

**A.2. Solutions for  $N_{var} = 6$**

$$\mathbf{M}_{6,4} = \mathbf{R}_{6,4}^M \tag{83}$$

$$\begin{pmatrix} \nabla \mathbf{R}_{6,4}^M \\ \Delta \mathbf{M}_{6,5} \end{pmatrix} = \begin{pmatrix} 1 & a & a & -a & -a & -a \\ b & 1 & a & a & -a & a \\ -b & b & 1 & a & a & -a \\ b & -b & b & 1 & a & a \\ b & -b & -b & b & 1 & -a \\ -b & b & b & -b & b & 1 \end{pmatrix}. \tag{84}$$

The constants  $a = 1/\sqrt{5} \approx 0.44721$  and  $b = 1/3 \approx 0.33333$ .

**A.3. Solutions for  $N_{var} = 7$**

$$\mathbf{M}_{7,4} = \mathbf{R}_{7,4}^M \tag{85}$$

$$\begin{pmatrix} \nabla \mathbf{R}_{7,4}^M \\ \Delta \mathbf{R}_{7,5}^M \end{pmatrix} = \left[ \begin{array}{ccc|cccc} 1 & 0 & 0 & c & -c & c & c \\ 0 & 1 & 0 & c & c & -c & c \\ 0 & 0 & 1 & c & c & c & -c \\ \hline d & d & d & 1 & b & b & b \\ d & -d & -d & e & 1 & -b & -b \\ d & d & -d & -e & -e & 1 & -b \\ -d & d & -d & e & e & -e & 1 \end{array} \right],$$

$$\begin{aligned} b &= 1/3 \approx 0.33333 & (6\times) \\ c &= 1/\sqrt{3} \approx 0.57735 & (12\times) \\ d &= \sqrt{3}/4 \approx 0.43301 & (12\times) \\ e &= 1/4 = 0.25 & (6\times). \end{aligned} \tag{86}$$

Note that  $\mathbf{R}_{7,4}^M$  and  $\mathbf{R}_{7,5}^M$  constitute a pair of matrices related in the sense of Eq. (77). In this case,  $n = 4$  and the ratio  $\eta(7, 4) = 3/4$ . One can check that  $d/c = e/b = \eta(7, 4)$ . Note also that the three zero correlations correspond to three mutually orthogonal variables, see Fig. 5(a).

$$\begin{pmatrix} \nabla \mathbf{M}_{7,5} \\ \Delta \mathbf{M}_{7,6} \end{pmatrix} = \begin{pmatrix} 1 & -f & f & -f & f & f & f \\ g & 1 & f & f & 0 & f & 0 \\ g & -g & 1 & -f & -f & f & f \\ -g & g & g & 1 & f & -f & f \\ -g & g & g & g & 1 & f & 0 \\ g & g & g & g & -g & 1 & -f \\ -g & -g & g & -g & g & g & 1 \end{pmatrix},$$

$$\begin{aligned} f &\approx 0.39039 & (18\times) \\ g &\approx 0.28621 & (21\times). \end{aligned} \tag{87}$$

**A.4. Solutions for  $N_{var} = 8$**

$$\begin{pmatrix} \nabla \mathbf{R}_{8,4}^M \\ \Delta \frac{1}{\eta} \mathbf{R}_{8,6}^M \end{pmatrix} = \begin{pmatrix} 1 & r_1 & r_8 & r_1 & r_{10} & -r_{11} & -r_1 & r_5 \\ -r_1 & 1 & r_8 & r_{10} & r_1 & -r_1 & -r_{11} & r_5 \\ r_8 & r_8 & 1 & r_1 & r_1 & r_7 & r_7 & -r_2 \\ r_1 & r_{10} & -r_1 & 1 & r_{12} & r_1 & -r_4 & -r_9 \\ -r_{10} & -r_1 & r_1 & r_{12} & 1 & -r_4 & r_1 & -r_9 \\ r_{11} & r_1 & r_7 & r_1 & r_4 & 1 & -r_6 & -r_3 \\ -r_1 & -r_{11} & -r_7 & r_4 & r_1 & -r_6 & 1 & -r_3 \\ r_5 & r_5 & r_2 & r_9 & -r_9 & -r_3 & r_3 & 1 \end{pmatrix},$$

$$\begin{aligned} r_1 &\approx 0.67112 & (9\times) & r_7 &\approx 0.30392 & (2\times) \\ r_2 &\approx 0.65310 & (1\times) & r_8 &\approx 0.27802 & (2\times) \\ r_3 &\approx 0.56668 & (2\times) & r_9 &\approx 0.26572 & (2\times) \\ \text{where } r_4 &\approx 0.48481 & (2\times) & r_{10} &\approx 0.09898 & (2\times) \\ r_5 &\approx 0.47784 & (2\times) & r_{11} &\approx 0.00665 & (2\times) \\ r_6 &\approx 0.34247 & (1\times) & r_{12} &\approx 0.00469 & (1\times) \\ \eta &= 0.6. \end{aligned} \tag{88}$$

The pair of solutions  $\mathbf{R}_{8,4}^M$  and  $\mathbf{R}_{8,6}^M$  is another example of matrices related through  $\eta$  in the sense of Eq. (77).

$$\begin{pmatrix} \nabla R_{9,4}^M \\ \Delta \frac{1}{\eta} R_{9,7}^M \end{pmatrix} = \begin{pmatrix} 1 & -r_1 & -r_1 & -r_4 & r_3 & r_5 & -r_1 & r_3 & -r_8 \\ -r_1 & 1 & r_1 & -r_6 & -r_1 & -r_7 & r_6 & r_4 & r_1 \\ r_1 & r_1 & 1 & r_1 & r_8 & r_2 & r_7 & -r_5 & r_2 \\ r_4 & -r_6 & -r_1 & 1 & r_1 & r_1 & r_6 & -r_1 & -r_7 \\ r_3 & r_1 & r_8 & r_1 & 1 & r_1 & -r_4 & -r_3 & -r_5 \\ r_5 & r_7 & r_2 & r_1 & -r_1 & 1 & -r_1 & r_8 & r_2 \\ -r_1 & -r_6 & r_7 & r_6 & r_4 & r_1 & 1 & -r_1 & -r_1 \\ r_3 & -r_4 & -r_5 & -r_1 & r_3 & -r_8 & r_1 & 1 & r_1 \\ r_8 & r_1 & -r_2 & r_7 & -r_5 & r_2 & -r_1 & r_1 & 1 \end{pmatrix}, \text{ where } \begin{matrix} r_1 \approx 0.69094 & (15\times) \\ r_2 \approx 0.49748 & (3\times) \\ r_3 \approx 0.45317 & (3\times) \\ r_4 \approx 0.29216 & (3\times) \\ r_5 \approx 0.26765 & (3\times) \\ r_6 \approx 0.04431 & (3\times) \\ r_7 \approx 0.03348 & (3\times) \\ r_8 \approx 0.00897 & (3\times) \\ \eta = 0.5 \end{matrix} \quad (92)$$

$$\begin{pmatrix} \nabla R_{9,5}^M \\ \Delta \frac{1}{\eta} R_{9,6}^M \end{pmatrix} = \begin{pmatrix} 1 & -r_1 & -r_1 & r_1 & r_2 & r_4 & -r_4 & -r_5 & -r_1 \\ -r_1 & 1 & r_3 & r_2 & -r_1 & -r_5 & r_1 & -r_2 & -r_4 \\ -r_1 & -r_3 & 1 & -r_3 & r_1 & -r_1 & r_1 & r_3 & -r_3 \\ r_1 & -r_2 & r_3 & 1 & -r_5 & r_1 & r_1 & -r_4 & -r_2 \\ -r_2 & -r_1 & r_1 & -r_5 & 1 & -r_4 & r_4 & r_1 & -r_1 \\ r_4 & r_5 & r_1 & -r_1 & -r_4 & 1 & r_2 & r_1 & r_1 \\ r_4 & r_1 & r_1 & r_1 & -r_4 & r_2 & 1 & r_1 & -r_5 \\ r_5 & -r_2 & r_3 & -r_4 & -r_1 & r_1 & -r_1 & 1 & r_2 \\ -r_1 & r_4 & r_3 & r_2 & -r_1 & -r_1 & -r_5 & r_2 & 1 \end{pmatrix}, \text{ where } \begin{matrix} r_1 \approx 0.48887 & (9\times) \\ r_2 \approx 0.45574 & (1\times) \\ r_3 \approx 0.27120 & (2\times) \\ r_4 \approx 0.20573 & (2\times) \\ r_5 \approx 0.04137 & (2\times) \\ \eta = 0.8 \end{matrix} \quad (93)$$

$$\begin{pmatrix} \nabla M_{9,4} \\ \Delta M_{9,5} \end{pmatrix} = \begin{pmatrix} 1 & -m_3 & m_1 & m_1 & m_1 & m_3 & m_1 & m_1 & -m_3 \\ -n_1 & 1 & m_1 & -m_1 & m_3 & m_1 & -m_1 & m_3 & m_2 \\ -n_1 & n_1 & 1 & -m_3 & m_1 & m_1 & m_3 & m_2 & m_3 \\ n_1 & -n_2 & n_1 & 1 & m_3 & -m_1 & m_1 & m_1 & m_3 \\ -n_1 & n_1 & n_1 & n_1 & 1 & m_1 & m_1 & -m_3 & -m_1 \\ n_1 & n_1 & -n_2 & n_1 & n_1 & 1 & -m_3 & -m_2 & -m_2 \\ b & n_1 & -n_1 & -n_1 & -n_1 & n_1 & 1 & m_3 & -m_1 \\ n_1 & n_1 & n_1 & n_1 & -n_2 & n_1 & n_1 & 1 & m_1 \\ -b & n_1 & -n_2 & -n_1 & n_1 & n_1 & b & -n_1 & 1 \end{pmatrix}, \text{ where } \begin{matrix} m_1 \approx 0.66936 & (19\times) \\ m_2 \approx 0.40421 & (4\times) \\ m_3 \approx 0.07357 & (13\times) \\ b = 1/3 \approx 0.33333 & (3\times) \\ n_1 \approx 0.43426 & (30\times) \\ n_2 \approx 0.13148 & (3\times) \end{matrix} \quad (94)$$

$$\begin{pmatrix} \nabla M_{9,6} \\ \Delta M_{9,7} \end{pmatrix} = \begin{pmatrix} 1 & b & b & -b & b & -b & b & b & b \\ -m_1 & 1 & b & b & -b & -b & b & -b & b \\ m_1 & -m_1 & 1 & b & -b & b & -b & b & b \\ -m_2 & m_1 & m_1 & 1 & -b & b & b & b & -b \\ m_3 & m_1 & m_1 & m_1 & 1 & b & b & b & b \\ m_1 & m_1 & -m_1 & m_1 & m_1 & 1 & b & b & b \\ m_1 & m_1 & m_1 & m_1 & -m_1 & -m_2 & 1 & b & -b \\ m_1 & m_1 & m_1 & -m_1 & -m_4 & m_3 & m_1 & 1 & -b \\ -m_1 & -m_1 & m_1 & m_1 & -m_1 & m_1 & -m_1 & m_1 & 1 \end{pmatrix}, \text{ where } \begin{matrix} b = 1/3 \approx 0.33333 & (36\times) \\ m_1 \approx 0.27825 & (31\times) \\ m_2 \approx 0.22673 & (2\times) \\ m_3 \approx 0.19257 & (2\times) \\ m_4 \approx 0.16051 & (1\times) \end{matrix} \quad (95)$$

$$(M_{9,8}) = \begin{pmatrix} 1 & u & u & u & u & -u & u & -u & u \\ & 1 & u & u & u & u & -u & u & -u \\ & & 1 & -u & -u & u & u & u & u \\ & & & 1 & -u & u & u & u & u \\ & & & & 1 & u & u & u & u \\ & & & & & 1 & u & -u & u \\ & & & & & & 1 & u & -u \\ & & & & & & & 1 & u \\ & & & & & & & & 1 \end{pmatrix}, \text{ where } u = 0.2 \quad (36\times). \quad (96)$$

Box I.

$$(\nabla M_{8,4}) = \begin{pmatrix} 1 & m_3 & -m_4 & m_1 & -m_1 & -m_1 & -m_3 & m_1 \\ & 1 & -m_1 & m_4 & m_3 & m_1 & m_1 & m_1 \\ & & 1 & -m_1 & -m_1 & -m_2 & m_4 & 0 \\ & & & 1 & m_4 & -m_1 & 0 & 0 \\ & & & & 1 & m_1 & -m_3 & -m_1 \\ & & & & & 1 & m_1 & 0 \\ & & & & & & 1 & m_1 \\ & & & & & & & 1 \end{pmatrix},$$

where  $\begin{matrix} m_1 \approx 0.647594 & (15\times) \\ m_2 \approx 0.421881 & (2\times) \\ m_3 \approx 0.176203 & (4\times) \\ m_4 \approx 0.108793 & (4\times) \end{matrix}$

(89)

$$\begin{pmatrix} \nabla R_{8,5}^M \\ \Delta M_{8,5} \end{pmatrix} = \begin{pmatrix} 1 & 0 & a & 0 & a & -a & a & a \\ h & 1 & 0 & a & a & a & -a & a \\ k & -k & 1 & 0 & -a & a & a & a \\ -k & k & h & 1 & a & a & a & -a \\ k & k & k & k & 1 & -a & 0 & 0 \\ k & k & k & k & h & 1 & 0 & 0 \\ -k & -k & k & k & k & -k & 1 & -a \\ -k & -k & k & k & -k & k & h & 1 \end{pmatrix},$$

where  $\begin{matrix} a = 1/\sqrt{5} \approx 0.44721 & (20\times) \\ h = 3 - \sqrt{8} = k^2 \approx 0.17157 & (4\times) \\ k = \sqrt{2} - 1 \approx 0.41421 & (24\times) \end{matrix}$

(90)

$$\begin{pmatrix} \nabla \mathbf{M}_{8,6} \\ \triangle \mathbf{M}_{8,7} \end{pmatrix} = \begin{pmatrix} 1 & b & b & b & b & b & b & -b \\ s & 1 & -b & b & b & b & b & b \\ t & t & 1 & b & b & b & b & b \\ -t & t & t & 1 & -b & b & b & b \\ t & t & -t & t & 1 & b & b & b \\ t & -t & t & t & t & 1 & -b & b \\ t & -t & -t & t & -t & -t & 1 & b \\ t & t & -t & -t & -t & t & s & 1 \end{pmatrix}, \quad (91)$$

$$b = 1/3 \approx 0.33333 \quad (28 \times)$$

$$\text{where } s \approx 0.10404 \quad (2 \times)$$

$$t \approx 0.24094 \quad (26 \times).$$

#### A.5. Solutions for $N_{\text{var}} = 9$

Eqs. (92)–(96) are given in Box I.

#### References

- [1] Shinozuka M. Simulation of multivariate and multidimensional random processes. *The Journal of the Acoustical Society of America* 1971;49(1 Part 2): 357–68.
- [2] Shinozuka M, Deodatis G. Simulation of stochastic processes by spectral representation. *Applied Mechanics Reviews* 1991;44(4):191–204.
- [3] Spanos PD, Ghanem RG. Stochastic finite element expansion for random media. *ASCE Journal of Engineering Mechanics* 1989;115(5):1035–53.
- [4] Kala Z. Stability problems of steel structures in the presence of stochastic and fuzzy uncertainty. *Thin-Walled Structures* 2007;45(10–11):861–5.
- [5] Sakamoto S, Ghanem R. Polynomial chaos decomposition for the simulation of non-Gaussian nonstationary stochastic processes. *Journal of Engineering Mechanics* 2002;128(2):190–201.
- [6] Puig B, Poirion F, Soize C. Non-Gaussian simulation using Hermite polynomial expansion: convergences and algorithms. *Probabilistic Engineering Mechanics* 2002;17(3):253–64.
- [7] Phoon K, Huang H, Quek S. Simulation of strongly non-Gaussian processes using Karhunen–Loève expansion. *Probabilistic Engineering Mechanics* 2005; 20(2):188–98.
- [8] Vořechovský M. Simulation of simply cross correlated random fields by series expansion methods. *Structural Safety* 2008;30(4):337–63.
- [9] Grigoriu M. Simulation of stationary process via a sampling theorem. *Journal of Sound and Vibration* 1993;166(2):301–13.
- [10] Grigoriu M. Evaluation of Karhunen–Loève, spectral, and sampling representations for stochastic processes. *Journal of Engineering Mechanics* 2006;132(2): 179–89.
- [11] Iman RC, Conover WJ. Small sample sensitivity analysis techniques for computer models with an application to risk assessment. *Communications in Statistics: Theory and Methods* 1980;A9(17):1749–842.
- [12] Iman RC, Conover WJ. A distribution free approach to inducing rank correlation among input variables. *Communications in Statistics B* 1982;11:311–34.
- [13] Iman RL, Shortencarier MJ. A FORTRAN 77 program and user's guide for the generation of Latin hypercube and random samples for use with computer models. Technical report NUREG/CR-3957 report. US Nuclear Regulatory Commission. SAND83-2365. 1984.
- [14] Vořechovský M, Novák D. Correlation control in small sample Monte Carlo type simulations I: a simulated annealing approach. *Probabilistic Engineering Mechanics* 2009;24(3):452–62.
- [15] Vořechovský M, Novák D. Statistical correlation in stratified sampling. In: Der Kiureghian A, Madanat S, Pestana JM, editors. ICASP 9, international conference on applications of statistics and probability in civil engineering, held in San Francisco, USA. Rotterdam (Netherlands): Mill Press; 2003. p. 119–24.
- [16] Owen AB. Controlling correlations in Latin hypercube samples. *Journal of the American Statistical Association (Theory and Methods)* 1994;89(428): 1517–22.
- [17] Wang Y-T, Lam F, Barrett JD. Simulation of correlated modulus of elasticity and compressive strength of lumber with gain factor. *Probabilistic Engineering Mechanics* 1995;10(2):63–71.
- [18] Huntington DE, Lyrintzis CS. Improvements to and limitations of Latin hypercube sampling. *Probabilistic Engineering Mechanics* 1998;13(4):245–53.
- [19] Vořechovský M. Correlation control in small sample Monte Carlo type simulations II: analysis of estimation formulas, random correlation and perfect uncorrelatedness. *Probabilistic Engineering Mechanics* 2012;29:105–20.
- [20] Aitken AC. Determinants and matrices. University mathematical texts, Edinburgh: Oliver and Boyd; 1939.
- [21] Graybill FA. Introduction to matrices with applications in statistics. Belmont (CA): Wadsworth Publishing Co.; 1969.
- [22] Andrew AL. Eigenvectors of certain matrices. *Linear Algebra and its Applications* 1973;7(2):151–62.
- [23] Cantoni A, Butler P. Eigenvalues and eigenvectors of symmetric centrosymmetric matrices. *Linear Algebra and its Applications* 1976;13(3):275–88.
- [24] Delsarte P, Genin Y. Spectral properties of finite Toeplitz matrices. In: Fuhrmann PA, editor. Mathematical theory of networks and systems, proceedings of the MTNS-83 international symposium, held in Beer Sheva, Israel. Lecture notes in control and information sciences, vol. 58. Berlin (New York): Springer-Verlag; 1983. p. 194–213. 1984.
- [25] Muthiyalu N, Usha S. Eigenvalues of centrosymmetric matrices. *Computing (Springer)* 1992;48(2):213–8.
- [26] Rousseeuw PJ, Molenberghs G. The shape of correlation matrices. *The American Statistician* 1994;48(4):276–9.