Behaviour of the NORTA Method for Correlated Random Vector Generation as the Dimension Increases

SOUMYADIP GHOSH and SHANE G. HENDERSON

Cornell University

1. INTRODUCTION

There is a growing need to capture dependence between random variables that serve as primitive inputs to stochastic models. In a manufacturing setting, for example, the processing times of a single job at different stations may be correlated due to characteristics of the job such as size. In financial engineering, Das et al. [2001] claim that the risk profile of credit portfolios can be understated if correlation is ignored. Further applications have recently been reported in cost analysis [Lurie and Goldberg 1998], and in decision and risk analysis [Clemen and Reilly 1999]. If simulation is used with such models, then we need methods for efficiently generating samples of correlated random variables. We examine the case where the correlated primitive inputs of a model are finite in number and hence can be characterized

Authors' Address: School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, U.S.A.

The work of the first author was partially supported by an IBM Ph.D. Fellowship Award. This research was partially supported by NSF Grant Number DMI-0230528.

jointly as a random vector.

We focus on the NORmal To Anything method of random vector generation described by Cario and Nelson [1997]. The NORTA method belongs to a family of methods that address the problem of generating samples of a finite-dimensional random vector with a given set of distributions for the individual components (their *marginal* distributions) and a given measure of dependence between them, which in the NORTA case is either the product-moment or the rank correlation matrix.

Definition 1.1. The *product-moment* correlation matrix for a random vector $X = (X_1, \ldots, X_d)$ is the correlation matrix (i.e., a symmetric positive semidefinite matrix with unit diagonal elements) $\Sigma_X = (\Sigma_X(i,j) : 1 \leq i, j \leq d)$ where

$$\Sigma_X(i,j) = \frac{\mathrm{Cov}(X_i, X_j)}{(\mathrm{Var}X_i \ \mathrm{Var}X_j)^{1/2}}.$$

The *rank* correlation matrix is of the same form except that now

$$\Sigma_X(i,j) = \frac{\mathrm{Cov}(F_i(X_i), F_j(X_j))}{(\mathrm{Var}F_i(X_i) \ \mathrm{Var}F_j(X_j))^{1/2}},$$

where F_i and F_j are the distribution functions of X_i and X_j respectively.

Remark 1.2. The product-moment correlation is a measure of *linear* correlation between random variables. Rank correlation is often preferred to product-moment correlation as a measure of dependence for two reasons. First, it is always defined, even if the random variables involved have infinite variance. Second, it is invariant with respect to strictly increasing transformations of the random variables involved.

The philosophy of specifying marginals and correlations to model dependent random variables is clearly an approximate one, since the joint distribution is not completely specified by this information. One hopes (but is not guaranteed) to capture the essence of the dependence between the random variables. This approach has the advantage that it is very easy to specify the marginal distributions and correlation matrix based on information obtained from sample data. An alternative is to fully specify the joint distribution. The primary difficulty in this case is that a tremendous amount of information is typically required to specify (and fit) such

a joint distribution. Furthermore, special methods must be devised to generate random vectors with the given joint distribution, and this can be a practically insurmountable problem for a model of even moderate complexity (Law and Kelton 2000, p. 479). Another alternative is to settle for a parameteric family of distributions where the marginal distributions come from a restricted class [Devroye 1986; Johnson 1987].

Another argument in support of modelling random vectors using marginals and correlations relates to the use of diffusion approximations for modelling stochastic systems. In many cases the limiting diffusions depend only on the first two moments of the input distributions. Therefore, there is some insensitivity in performance measures computed from these models to the exact form of the input distributions. In general then, if a form of this insensitivity is present in a model, then the approach discussed here for modelling random vectors is quite reasonable.

The NORTA method produces a random vector with the desired properties via a componentwise transformation of a multivariate normal random vector, and capitalizes on the fact that multivariate normal random vectors are easily generated; see e.g., Law and Kelton [2000], p. 480. Cario and Nelson [1997] traced the roots of the method back to Mardia [1970] who studied bivariate distributions, and to Li and Hammond [1975] who concentrated on the case where all of the marginals have densities (with respect to Lebesgue measure). Iman and Conover [1982] implemented the same transformation procedure to induce a given rank correlation in the output. Their method is only approximate, in that the output will have only very approximately the desired rank correlation.

The NORTA method is a very efficient and easy to implement sampling method, and has seen use in a variety of contexts. Clemen and Reilly [1999] use the NORTA procedure to induce a desired rank correlation in the context of decision and risk analysis. Lurie and Goldberg [1998] implement a variant of the NORTA method for use in cost analysis. Henderson et al. [2000] adapt the NORTA method to generate samples of dependent quasi-random vectors. The NORTA method is also routinely

used in portfolio models in industry.

So the NORTA procedure is often the method of choice for generating random vectors with prescribed marginals and correlation matrix. But can the NORTA procedure be used in all circumstances that require sampling such random vectors? To define this question more precisely we need a definition.

Definition 1.3. A product-moment (rank) correlation matrix Σ is *feasible* for a given set of marginal distributions $F_1, \ldots, F_d$ if there exists a random vector X with marginal distributions $F_1, \ldots, F_d$ and product-moment (rank) correlation matrix Σ.

The question we examine in this paper is whether the NORTA procedure can match *any* feasible correlation matrix for a given set of marginals.

For 2-dimensional random vectors, the NORTA method can match any feasible correlation matrix. This follows immediately from the characterizations in Whitt [1976]. However, this does not hold for dimensions 3 and greater. Both Li and Hammond [1975] and Lurie and Goldberg [1998] postulate examples of 3-dimensional random vectors where the NORTA procedure might fail, but do not establish that the counterexamples exist, i.e., that the example correlation matrices are feasible.

A necessary, but not always sufficient, condition for feasibility is that the given matrix is a *correlation* matrix, i.e., a symmetric, positive semidefinite matrix with unit diagonal elements. To see why this is not sufficient, consider a random vector that consists of an exponential random variable with mean 1 and a uniform (0,1] random variable. The maximum (minimum) product-moment correlation that can be induced between the two random variables is $\pm\sqrt{3}/2$, which is strictly less (greater) than $+1$ (-1). (Note that if correlations of 1 or -1 were achievable, then one random variable would be a linear function of the other, which is clearly impossible given the distributions specified.) In general then, the range of correlations that can be achieved between two random variables is a subinterval of [-1,1]. Hence for the case of the exponential-uniform pair, any matrix

$$\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix},$$

where $\sqrt{3}/2 < |r| \leq 1$, defines a 2×2 correlation matrix, but not a feasible correlation matrix. Thus, determining the feasibility of a given matrix for a specified set of marginals is not trivial.

In Ghosh and Henderson [2001; Ghosh and Henderson [2002a], a computational procedure is developed based on chessboard distributions to determine whether a given correlation matrix is feasible for the marginal distributions or not. Using this procedure one can rigorously establish that such counterexamples do exist. Let us call feasible correlation matrices that cannot be matched using the NORTA method *NORTA defective* matrices.

Based on the numerical results obtained in Ghosh and Henderson [2002a] we conjectured that as the dimension grew, more and more feasible correlation matrices would be NORTA defective. This is the first aspect of the feasiblity problem that we investigate in this paper. We estimate, for each dimension, the probability that the NORTA procedure fails to work for a feasible rank correlation matrix chosen uniformly from the set of all feasible correlation matrices. Kurowicka and Cooke [2001] also looked at this problem, but they worked with a probability distribution that is not uniform over the set of all feasible correlation matrices. Our results confirm their finding that the probability the NORTA procedure fails to work grows rapidly with dimension. This suggests that the NORTA procedure is unlikely to be effective in high dimensional problems.

However, suppose that we are willing to accept a random vector with the prescribed marginals, and a correlation matrix that is, at least approximately, the required correlation matrix. In Ghosh and Henderson [2002a] we describe a semidefinite programming approach that can assist in this regard.

The proposed augmented NORTA method works in exactly the same manner as the original method unless a NORTA defective matrix is encountered. For such a matrix, a semidefinite program is solved, and the results are then used to modify the inputs given to the NORTA generation step in the hope that the generated random vector has a correlation matrix that is "close" to the desired one (it has the same

marginal distributions). The numerical results in Ghosh and Henderson [2002a] indicate that this holds for the 3-dimensional case. In this paper, we examine higher dimensions, exploring how the augmented NORTA method performs as the dimension increases. Specifically we estimate the average distance (measured in an appropriate sense) between the correlation matrix given by the augmented NORTA procedure and the desired (possibly NORTA defective) matrix.

The results indicate that NORTA can typically get very close to a target correlation matrix, even in very high dimensions. So in high dimensions, while NORTA is (very) unlikely to be able to exactly match a desired correlation matrix, it is typically able to match a correlation matrix that is very close to the desired one.

An important part of our analysis is the development of a method for sampling uniformly from the set of all correlation matrices of a given dimension. We choose to call this method the *Onion Method* for reasons that will be clear once the working of the method is explained. The method is easily generalized to generate from the set of symmetric positive definite matrices with arbitrary (fixed) positive diagonal entries. Thus a possible use of (an appropriately modified version of) the sampling scheme might be to study the performance of algorithms that operate on such matrices. In related work, Marsaglia and Olkin [1984] survey methods for sampling random correlation matrices, but none of the methods they mention samples *uniformly* over the set of all correlation matrices (of fixed dimension). Edelman [1989] shows how the distribution of a symmetric positive definite matrix can be expressed as a function of the distributions of the matrices of its eigenvalue decomposition (Theorems 3.1 and 3.2), although we do not use this result in our analysis. The onion method can also be used to sample from bounded non-uniform densities on sets of the kind mentioned above, via standard sampling frameworks like the acceptance-rejection method.

The onion method is simple to implement since it uses nothing more than standard tools from the simulation input modelling toolkit, and sample generation is very fast. Indeed, the most complex and computationally demanding part of the

method involves sampling from univariate beta distributions, which is a very well-studied problem with many efficient algorithms available (see Law and Kelton 2000, p. 467).

This paper, which is an outgrowth of Ghosh and Henderson [2002b], is organized as follows. The next section reviews the NORTA procedure and indicates why some matrices may be NORTA defective. Section 3 discusses our simulation framework for estimating the performance of NORTA as the dimension increases. Section 4 describes the random matrix sampling procedure. In Section 5, we return to the NORTA method, briefly describing the SDP augmentation proposed in Ghosh and Henderson [2002a], and studying how this augmented method performs in higher dimensions.

2. THE NORTA PROCEDURE

Suppose that we wish to generate i.i.d. replicates of a random vector $X = (X_1, \ldots, X_d)$ with prescribed marginal distributions

$$F_i(\cdot) = P(X_i \leq \cdot), i = 1, \ldots, d,$$

and product-moment or rank correlation matrix

$$\Sigma_X = \Sigma_X(i,j),\ 1 \leq i,j \leq d.$$

The NORTA method generates i.i.d. replicates of X through the following procedure.

(1) Generate an $\mathbb{R}^d$ valued joint normal random vector $Z = (Z_1, \ldots, Z_d)$ with mean vector 0 and covariance matrix $\Sigma_Z = (\Sigma_Z(i,j) : 1 \leq i,j \leq d)$, where $\Sigma_Z(i,i) = 1$ for $i = 1, \ldots, d$. (Since the variances of each component have been chosen to be 1, Σ_Z also represents the product-moment correlation matrix of Z.)

(2) Compute the vector $X = (X_1, \ldots, X_d)$ via

$$X_i = F_i^{-1}(\Phi(Z_i)), \tag{1}$$

for $i = 1, \ldots, d$, where Φ is the distribution function of a standard normal random

variable, and

$$F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}. \tag{2}$$

The vector X generated by this procedure will have the prescribed marginal distributions. To see this, note that each Z_i has a standard normal distribution, so that $\Phi(Z_i)$ is uniformly distributed on $(0, 1)$, and so $F_i^{-1}(\Phi(Z_i))$ will have the required marginal distribution.

The covariance matrix Σ_Z should be chosen, in a preprocessing phase, so that it induces the prescribed correlation matrix Σ_X on X. However, there is no general closed-form expression that gives Σ_Z in terms of Σ_X. Indeed, determining the right Σ_Z is the most difficult step in implementing the NORTA method.

Each component of Σ_X has been shown to depend only on the corresponding component of Σ_Z. As in Cario and Nelson [1997], we can define $c_{ij}(z) = \Sigma_X(i, j)$ to represent the correlation between X_i and X_j as a function of the correlation z between Z_i and Z_j, when X_i and X_j are generated as in (1). Cario and Nelson [1997] show that under certain very mild conditions $c_{ij}(\cdot)$ is a non-decreasing, continuous function. This result helps us perform an efficient numerical search for a value $\Lambda_Z(i, j)$ that solves

$$c_{ij}(\Lambda_Z(i, j)) = \Sigma_X(i, j). \tag{3}$$

Hence a numerical estimate Λ_Z of Σ_Z can be determined by solving a number of one-dimensional root-finding problems. Unless stated otherwise, we assume that a solution exists for (3) for all i and j. This assumption is without loss of generality, since if Σ_X is feasible, then (3) must have a solution for all i and j. Henderson et al. [2000] also show that under slightly stronger assumptions, $\Lambda_Z(i, j)$ in (3) is uniquely determined by $\Sigma_X(i, j)$.

The matrix Λ_Z is constructed in a way that does not necessarily ensure that it is positive semidefinite. It might indeed turn out to be indefinite, in which case it cannot be a valid covariance matrix for a joint normal distribution, and NORTA will fail. Li and Hammond [1975] postulated the following example to demonstrate this possibility. Suppose that $X = (X_1, X_2, X_3)$ is a random vector with uniform

$(0, 1]$ marginals, and correlation matrix

$$\Sigma_X = \begin{pmatrix} 1 & -0.4 & 0.2 \\ -0.4 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix}.$$

For the special case of uniform marginals, (3) can be solved analytically (Kruskal 1958) to give

$$\Lambda_Z(i, j) = 2 \sin\left(\frac{\pi}{6}\Sigma_X(i, j)\right). \tag{4}$$

The unique solution Λ_Z for the given Σ_X turns out to be indefinite.

This example is of course valid only if such a uniform random vector X exists. Li and Hammond [1975] did not show this. Ghosh and Henderson [2002a] develop a computational procedure that can determine, for *almost* any (in a Lebesgue measure sense) given correlation matrix, whether it is feasible for a given set of marginal distributions (assumed to be continuous and have bounded support) or not. Applying this procedure to the Li and Hammond example gives a construction of the random vector, so that it does, indeed, exist. Ghosh and Henderson [2002a] generate a number of such feasible matrices for three-dimensional uniform random vectors that are NORTA defective. The numerical results suggest a structure to the failure of NORTA. To explain this observation more carefully we need some notation.

Suppose that the marginal distributions $F_1, \ldots, F_d$ have densities with bounded support, and are fixed. With an abuse of notation, we can view a $d \times d$ correlation matrix as an element of a $d(d-1)/2$ dimensional vector space, since there are $d(d-1)/2$ elements above the diagonal, the matrix is symmetric, and the diagonal elements are equal to 1. Let Ω denote the set of feasible correlation matrices. (The definition of Ω depends on whether we consider product-moment or rank correlation, but we suppress this dependence in our discussion.) We view this set as a subset of $d(d-1)/2$ dimensional space. Ghosh and Henderson [2002a] prove that in this setting Ω is nonempty, convex, closed and full-dimensional.

Returning to the discussion above, we found that in 3 dimensions, NORTA defective matrices tended to occur near the boundary of Ω. Moreover, the indefinite correlation matrices Λ_Z determined for the joint normal distribution from (3) seemed to lie close to (but outside of) the set of symmetric positive semidefinite matrices. So NORTA defective matrices tended to occur near the boundary, and they were never too distant from a NORTA feasible matrix.

3. NORTA IN HIGHER DIMENSIONS

NORTA appears to fail most often when the correlation matrix is close to the boundary of the set Ω. Now, in a sense that can be made precise, "most" points in certain sets in high dimensions lie close to the boundary. For example, consider the interior of the unit hypercube $[-\frac{1}{2}, \frac{1}{2}]^d$ in $\mathbb{R}^d$ represented by the hypercube $[-\frac{1-\epsilon}{2}, \frac{1-\epsilon}{2}]^d$, where $\epsilon \in (0, 1)$. The ratio of the volume of the interior to that of the whole set is $(1-\epsilon)^d$, which decreases rapidly to 0 as d increases.

This suggests that feasible matrices within the set Ω may become increasingly likely to be NORTA defective as the dimension of the problem increases, so that the feasiblity problem that NORTA faces may become increasingly acute as the dimension increases.

Let us consider this possibility in the context of generating samples of a uniform random vector, i.e., a random vector with uniform $(0, 1]$ marginal distributions. This case has special significance to the NORTA method because, by construction, the method has to generate a uniform random vector $(\Phi(Z_1), \ldots, \Phi(Z_d))$ as an intermediary step. Furthermore, the rank correlation matrix of a NORTA-generated vector with continuous marginal distributions coincides with the product moment correlation matrix for the intermediate uniform random vector.

This special case also has two advantages. First, the function c_{ij} is explicitly known; see (4). Hence, a correlation matrix for the uniform random vector can be easily tested for NORTA feasibility. One simply computes the (symmetric) matrix Λ_Z as given by (4) and checks whether it is positive semidefinite or not. Note that if Λ_Z is positive semidefinite then a joint normal random vector Z with this

correlation matrix exists, and this vector transforms through NORTA to a random vector with the desired correlation matrix. On the other hand a correlation matrix must be positive semidefinite, so if Λ_Z is indefinite, then it is not a correlation matrix and NORTA will fail.

Second, it has recently been established [Kurowicka and Cooke 2001] that the set of all feasible correlation matrices for uniform marginals, say Ω, coincides with the set of all symmetric positive semidefinite matrices with ones on the diagonal. Thus the problem of estimating the probability of NORTA infeasibility reduces to the following algorithm.

(1) Let $n \geq 1$ be given.

(2) Let $\Sigma_X(1), \ldots, \Sigma_X(n)$ be an i.i.d. sample chosen uniformly from

$$\Omega = \{\Sigma : \Sigma = \Sigma^T, \Sigma \succeq 0, \Sigma_{jj} = 1 \quad j = 1, \ldots, d\}. \tag{5}$$

(3) For each $i = 1, \ldots, n$ let $\Lambda_Z(i)$ be obtained from $\Sigma_X(i)$ using the componentwise relation (4).

(4) Estimate the probability of NORTA infeasibility by the proportion of matrices in $\{\Lambda_Z(i) : i = 1, \ldots, n\}$ that are not positive semidefinite.

(The matrix inequality $A \succeq 0$ signifies a constraint that the matrix A be positive semidefinite.)

Note that in estimating the probability of NORTA infeasibility we have had to choose a probability distribution on Ω. The uniform distribution (with respect to Lebesgue measure) is a natural choice, and is the one we prefer to work with. Kurowicka and Cooke [2001] give estimates for the probability of NORTA feasibility using a different distribution on Ω.

A straightforward approach (and one that we adopted) to estimating the probability of NORTA infeasibility is to combine three well-known methods in simulation estimation: acceptance-rejection, importance sampling and ratio estimation. We used importance sampling and acceptance rejection on the hypercube $[-1, 1]^{\frac{d(d-1)}{2}}$ (Ω is a strict subset of this hypercube) to choose correlation vectors from Ω. We

then used ratio estimation (see, e.g., Henderson [2001]) to estimate the probability of NORTA infeasibility. The estimator of the probability of NORTA infeasibility is therefore of the form

$$\frac{\sum_{i=1}^{n}[I(\Sigma_X(i) \succeq 0, \Lambda_Z(i) \not\succeq 0)\ \frac{2^{-d(d-1)/2}}{\phi(\Sigma_X(i))}]}{\sum_{i=1}^{n}[I(\Sigma_X(i) \succeq 0)\ \frac{2^{-d(d-1)/2}}{\phi(\Sigma_X(i))}]}, \tag{6}$$

where I represents the indicator function (i.e., returns 1 if its arguement is true, and 0 otherwise), and the matrices $\Sigma_X(i)$ are chosen independently with density ϕ from the hypercube $[-1,1]^{\frac{d(d-1)}{2}}$. We chose the density ϕ in a heuristic fashion.

This method of estimation works well in lower dimensions but we found that it became excessively slow as the dimension increased. Indeed, it took more than two days to generate on the order of a thousand samples of positive definite matrices even for a dimension as low as $d = 12$. With a better choice of ϕ the algorithm would presumably be much faster, but it is not clear how to choose ϕ. A better sampling technique is needed.

4. THE ONION METHOD

Our goal is to construct a method that samples *exactly*, and very quickly, from the uniform distribution on the set Ω_d as defined in (5), when viewed as a subset of $\mathbb{R}^{d(d-1)/2}$. We use the suffix d to emphasize the dependence on the dimension d. We thus have to construct a procedure that samples uniformly from the convex, closed, compact and full-dimensional set Ω_d (these properties are established in Ghosh and Henderson 2002a), i.e., generate samples from the density

$$f(\Sigma)\ \propto\ 1\,,\ \forall\ \Sigma \in \Omega_d, \tag{7}$$

where f is a function of the $d(d-1)/2$ upper-diagonal elements of Σ.

For a matrix Σ let $\Sigma_{\mathbf{k}}$ represent its $k \times k$ dimensional principal leading minor (i.e., the upper-left $k \times k$ submatrix of Σ), and $f_{\mathbf{k}}$ represent the marginal density of $\Sigma_{\mathbf{k}}$ when Σ has the joint density (7). Let $\mathbf{q}$ be the vector such that

$$\Sigma_{\mathbf{k}} = \begin{bmatrix} \Sigma_{\mathbf{k-1}} & \mathbf{q} \\ \mathbf{q}^t & 1 \end{bmatrix}.$$

We call $\mathbf{q}$ the *completion* of $\Sigma_{\mathbf{k-1}}$ in $\Sigma_{\mathbf{k}}$.

The onion method is iterative in that it starts with the one-dimensional matrix 1 and then "grows out" the matrix to the dimension desired by successively adding an extra row (and the corresponding mirrored column) chosen from an appropriate distribution. This successive layering approach is the inspiration behind its name. Marsaglia and Olkin [1984] use a similar matrix-growing approach in their algorithm to sample correlation matrices with a given set of eigenvalues, but they apply it to transform diagonal elements of arbitrary positive definite matrices to 1 in order to form correlation matrices from them. Ouellette [1981] points out some other uses of the layering approach, notable among them being the numerical method proposed by Guttman [1946] to compute inverses of large non-singular matrices.
To be more precise the onion method is as follows.

(1) Let $\Sigma_{\mathbf{1}}$ be the 1×1 matrix 1.

(2) For $k = 2, \ldots, d$

(a) Let $\mathbf{q}$ be a column vector in $\mathbb{R}^{k-1}$ chosen, independently of all else, from density $\varphi_k(\cdot; \Sigma_{\mathbf{k-1}})$ say.

(b) Set

$$\Sigma_{\mathbf{k}} = \begin{bmatrix} \Sigma_{\mathbf{k-1}} & \mathbf{q} \\ \mathbf{q^t} & 1 \end{bmatrix}.$$

(c) Next k.

The densities φ_k, which determine the kth layer, are conditional densities that depend on the partial matrix $\Sigma_{\mathbf{k-1}}$ constructed thus far. We now state the key result that motivates the iterative sampling scheme, and in particular provides the form of the φ_ks.

PROPOSITION 4.1. *Let $f_{\mathbf{k}}$ be the marginal density of $\Sigma_{\mathbf{k}}$ when Σ is distributed as in (7). Then*

$$f_{\mathbf{k}}(\Sigma_{\mathbf{k}}) \propto \ (\det(\Sigma_{\mathbf{k}}))^{\frac{d-k}{2}} \qquad \forall \ \Sigma_{\mathbf{k}} \in \Omega_k \, , \ \forall \, 2 \leq k \leq d.$$

The marginal density $f_{\mathbf{k}}$ represents the *joint* marginal density of the components $\Sigma_{\mathbf{k-1}}$ and $\mathbf{q}$ of $\Sigma_{\mathbf{k}}$, and Proposition 4.1 expresses $f_{\mathbf{k}}$ as a function of $\Sigma_{\mathbf{k-1}}$ and q (through $\Sigma_{\mathbf{k}}$). The density φ_k of the completion $\mathbf{q}$ can then be obtained from $f_{\mathbf{k}}$ by conditioning on a fixed $\Sigma_{\mathbf{k-1}}$, $k = 2, \ldots, d$. The key to the generation scheme is the fact that the expression obtained for φ_k by this conditioning arguement can be unravelled in terms of $\Sigma_{\mathbf{k-1}}$ and $\mathbf{q}$ in a way that allows for easy generation of $\mathbf{q}$ for a fixed $\Sigma_{\mathbf{k-1}}$.

We describe an efficient scheme to sample $\mathbf{q}$ from φ_k after we prove Proposition 4.1. We need two preliminary results for the proof. Let $I(\cdot)$ be the indicator function.

LEMMA 4.2. *If $m \geq 0$ and A is some symmetric p.d. matrix in Ω_d, then*

$$C = \int_{\mathbb{R}^d} I(x^t A x \leq 1)\ (1 - x^t A x)^m dx = L(m,d) \cdot \det(A)^{-\frac{1}{2}},$$

where $0 < L(m,d) < \infty$ is independent of A.

PROOF. Since A is symmetric and positive definite, it has a unique upper triangular Cholesky factor $A^{1/2}$ say, so that $A = (A^{\frac{1}{2}})^t A^{\frac{1}{2}}$. Applying the linear change of variables $w = A^{\frac{1}{2}} x$ gives

$$\begin{aligned} C &= |\det(A)^{-\frac{1}{2}}| \cdot \int_{\mathbb{R}^d} I(w^t w \leq 1)\ (1 - w^t w)^m dw \\ &= \det(A)^{-\frac{1}{2}} \cdot L(m,d). \end{aligned}$$

The function $g(w) = I(w^t w \leq 1)\ (1 - w^t w)^m$ is non-negative, bounded and non-zero only over the compact region that forms the unit ball in $\mathbb{R}^d$. Hence $0 \leq L(m,d) < \infty$. Since $g(0) = 1$ and g is continuous about 0, $L(m,d) > 0$. □

For the second result that we use in the proof of Proposition 4.1, first note that any positive definite symmetric $d \times d$ matrix A can be written as a product of two matrices, akin to a first step in an LU factorization of A, as

$$A = \begin{bmatrix} B & q \\ q^t & 1 \end{bmatrix} = \begin{bmatrix} B & 0 \\ q^t & 1 \end{bmatrix} \begin{bmatrix} I & B^{-1}q \\ 0 & 1 - q^t B^{-1} q \end{bmatrix}, \tag{8}$$

where $B = A_{\mathbf{d-1}}$. The quantity $1 - q^t B^{-1} q$ is called the *Schur complement* of B

in A. Ouellette [1981] is a useful source of literature on Schur complements. In particular, Ouellette [1981] points to the result obtained in Guttman [1946] that $\text{rank}(A) = \text{rank}(B) + I(1 - q^t B^{-1} q)$. This immediately gives us the following lemma.

LEMMA 4.3. *A necessary and sufficient condition for A to be positive definite is that B be positive definite, and $1 - q^t B^{-1} q > 0$.*

Ouellette [1981] also describes a result that follows from (8) and was first shown by Frobenius [1968]:

$$\det(A) = \det(B)\,(1 - q^t B^{-1} q). \tag{9}$$

We are now ready to prove Proposition 4.1.

PROOF OF PROPOSITION 4.1. We use induction on k from d to 2 to complete the proof. The result is immediate for $k = d$, since the density (7) in this case is the density we are aiming for in the first place. This establishes the base case.

Let $\Psi_k = \{q \in \mathbb{R}^k |\ q^t (\Sigma_{\mathbf{k}})^{-1} q \leq 1\}$. Then, by Lemma 4.3, Ψ_k represents the set of all completion vectors $\mathbf{q}$ of $\Sigma_{\mathbf{k}}$ in $\Sigma_{\mathbf{k+1}}$. For any general k, assuming that the induction hypothesis holds for $k + 1$, we get

$$\begin{aligned}
f_{\mathbf{k}}(\Sigma_{\mathbf{k}}) &= \int_{\Psi_k} f_{\mathbf{k+1}}(\Sigma_{\mathbf{k+1}})\, dq && (10)\\
&\propto \int_{\Psi_k} \det(\Sigma_{\mathbf{k+1}})^{\frac{d-k-1}{2}}\, dq && (11)\\
&= \int_{\Psi_k} \det(\Sigma_{\mathbf{k}})^{\frac{d-k-1}{2}}\, (1 - q^t \Sigma_{\mathbf{k}}^{-1} q)^{\frac{d-k-1}{2}}\, dq && (12)\\
&\propto \det(\Sigma_{\mathbf{k}})^{\frac{d-k-1}{2}} \cdot \det(\Sigma_{\mathbf{k}})^{\frac{1}{2}} && (13)\\
&= \det(\Sigma_{\mathbf{k}})^{\frac{d-k}{2}}.
\end{aligned}$$

The first step (10) above expresses the marginal density of $\Sigma_{\mathbf{k}}$ as the function that results from integrating out the $(k+1)$th column, the completion vector $\mathbf{q}$, from the marginal density of $\Sigma_{\mathbf{k+1}}$ over the set Ψ_k. The inductive hypothesis gives (11). The equality (12) uses (9), and (13) follows from Lemma 4.2.

Thus the induction hypothesis holds for k, and hence is true for all k from d to 2. □

We now determine the densities φ_k used in the iterative generation procedure from the marginal densities of Proposition 4.1. As mentioned before, the densities $f_{\mathbf{k}}$ represent the joint densities of $\Sigma_{\mathbf{k-1}}$ and its corresponding completion vector $\mathbf{q}$ in $\Sigma_{\mathbf{k}}$. Hence, if $\Sigma_{\mathbf{k-1}}$ were fixed at A say, we would have that

$$\begin{aligned}
\varphi_k(q) &= f_{\mathbf{k}|\{\Sigma_{\mathbf{k-1}}=A\}}(\Sigma_{\mathbf{k}}) \\
&\propto \det\left(\begin{bmatrix} A & q \\ q^t & 1 \end{bmatrix}\right)^{\frac{d-k}{2}} \\
&= \det(A)^{\frac{d-k}{2}} \cdot (1 - q^t A^{-1} q)^{\frac{d-k}{2}} \ .
\end{aligned}$$

Therefore, given $\Sigma_{\mathbf{k-1}}$, the conditional density for its completion vector $\mathbf{q}$ is

$$\varphi_k(q) \ \propto \ (1 - q^t \Sigma_{\mathbf{k-1}}^{-1} q)^{\frac{d-k}{2}} \ \forall \ q \in \Psi_{k-1}. \tag{14}$$

Next comes the question of generating from densities of the form (14). For this we employ a sequence of variable transformations. First we apply the linear transformation $w = \Sigma_{\mathbf{k-1}}^{-1/2} q$, where $\Sigma_{\mathbf{k-1}}^{-1/2}$ represents the upper triangular Cholesky factor of $\Sigma_{\mathbf{k-1}}^{-1}$, to get that

$$\varphi_k(q)dq \ \propto \ \tilde{\varphi}_k(w)dw,$$

where $\tilde{\varphi}_k(w) \propto (1 - w^t w)^{(d-k)/2}$, and $w \in \mathbb{B}^{k-1}$, the unit ball in $\mathbb{R}^{k-1}$ (the constant Jacobian term that arises out of the transformation is included in the proportionality constant). Hence, to sample $\mathbf{q}$ from φ_k we could equivalently generate a $\mathbf{w}$ from $\tilde{\varphi}_k$ and set $\mathbf{q}$ to be the appropriate linear transformation of $\mathbf{w}$.

Now, $\tilde{\varphi}_k$ is radially symmetric, as is the set $\mathbb{B}^{k-1}$. Thus if we apply a polar transformation $w = (r, \theta)$, where r is the l_2-norm of w and $\theta = (\theta_1, \ldots, \theta_{k-2})$ represents the angles of the polar transformation (refer Kendall [1961] p. 15 for a treatment of polar transformations in higher dimensions), then

$$\begin{aligned}
\tilde{\varphi}_k(w)dw &\propto (1 - r^2)^{\frac{d-k}{2}} \ J(r, \theta) \ dr d\theta_1 \ldots d\theta_{k-2} \\
&= (1 - r^2)^{\frac{d-k}{2}} \ r^{k-2} dr \ (\cos\theta_1)^{k-3} \, (\cos\theta_2)^{k-4} \ldots \cos\theta_{k-3} \ d\theta_1 \ldots d\theta_{k-2} \\
&\propto h(r)dr
\end{aligned}$$

where $J(r,\theta)$ represents the Jacobian term of the variable transformation and expands out as given in the second equation, and $h(r) = (1-r^2)^{\frac{d-k}{2}} r^{k-2}$. The second equation implies that the distribution of r is independent of the distributions of the angles θ_i. Moreover, the radial symmetry of the integrand also gives us that $\tilde{\varphi}_k(w)$ affects only the distribution of r, and the angles need to be sampled such that a point is chosen uniformly on the surface of the unit hyperball $\mathbb{B}^{k-1}$.

This suggests that we can sample a $\mathbf{w}$ from $\tilde{\varphi_k}$ by instead first sampling a radius from a normalized version of h and then multiplying by a point chosen *uniformly* over the surface of the unit ball $\mathbb{B}^{k-1}$. Such a point can be generated by normalizing a joint-normal independent random vector (i.e., one with the identity matrix as its correlation matrix) to have unit norm. The radius has to be sampled from h, but note that under yet another change of variable $y = r^2$, we have that

$$h(r)dr \;\propto\; y^{\alpha_1 - 1}(1-y)^{\alpha_2 - 1}dy,$$

which (after normalization) is simply a univariate beta density function with parameters α_1 and α_2. For our case, the parameters α_1 and α_2 turn out to be $(k-1)/2$ and $(d-k)/2$ respectively. Law and Kelton (2000 p. 467) points to extensive literature on generating from beta distributions.

To recap, at the kth stage of the iterative generating procedure, to generate a sample q of $\mathbf{q}$ from φ_k given the matrix $\Sigma_{\mathbf{k-1}}$ already constructed, we do the following:

— Sample $y = r^2$ from a beta distribution with $\alpha_1 = (k-1)/2$ and $\alpha_2 = (d-k)/2$,

— Sample a unit vector θ uniformly from the surface of $\mathbb{B}^{k-1}$,

— Set $w = r\theta$, and finally

— Set $q = \Sigma_{\mathbf{k-1}}^{\frac{1}{2}} w$.

This completes the description of the onion sampling method.

This exact sampling method is very competitive when estimating statistical properties of the set Ω_d when compared to methods like the one described in Section 3. First, since sampling from φ_i can be reduced to the problem of sampling from a

univariate beta distribution (and a joint-normal independent random vector), this method scales very well with dimension. In our study we were able to generate samples consisting of many thousands of matrices up to dimension $d = 25$ in a matter of hours. Second, this method does not involve a ratio-estimation step, which means that the estimation is more straightforward to implement. For a given sample size, we also found the results to be more accurate.

As noted before, this method can be generalized and applied very easily to generate uniformly from sets of symmetric positive definite matrices with any arbitrary (fixed) positive diagonal elements. One simply has to modify the method by substituting the diagonal values of 1 assumed in this section with the corresponding positive values at the appropriate places (the definition of Ψ_k in the proof of Proposition 4.1 is one such place). The constants of the variate generation method would be affected accordingly (for instance, the beta variate generation would not be over $(0, 1]$).

One can also use this method to sample from any bounded non-uniform density f defined on a set of symmetric positive definite matrices of the kind mentioned above. One simply uses the acceptance-rejection framework of random variate generation to do this, namely by first generating a point s uniformly from the set and then checking whether $f_{\max} * U \leq f(s)$ (where U is an independent uniform random variable, and $f_{\max} = \max_x f(x)$) in order to accept s as a sample.

The beta distribution used to sample the polar variable r above can be replaced with any distribution over the positive real line. Thus the onion method can also sample from any member of the family of distributions on the set of symmetric p.s.d. random matrices that are radially symmetric under an affine transformation (the first transformation in the sequence above).

5. FIXING NORTA

We used the exact sampling approach of Section 4 to estimate the probability of NORTA infeasibility for various dimensions. Our results are given in Figure 1, where the probability is plotted against dimension. The plot establishes that the

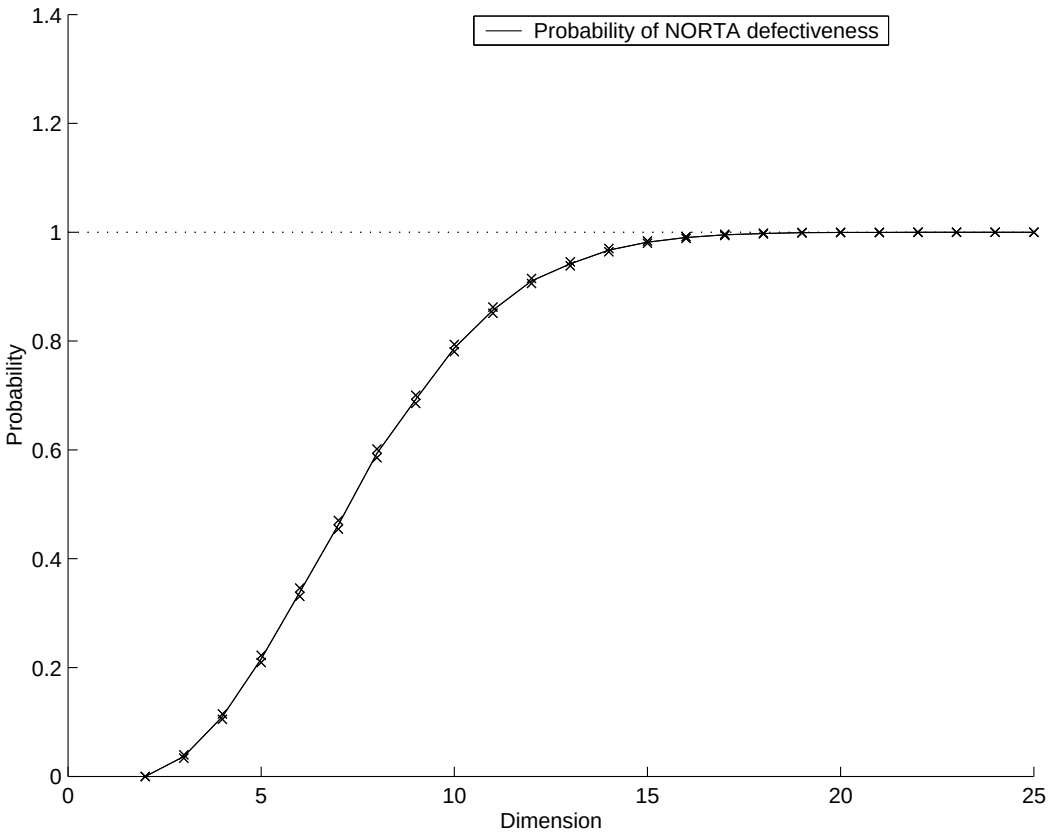

Fig. 1. Probability of NORTA infeasibility, based on sampling 15,000 matrices uniformly from Ω in each dimension. Also shown are 95% confidence intervals.

feasiblity problem rapidly becomes acute as the dimension increases. It would seem that NORTA can only be successful in low-dimensional problems, but this is in fact, not the case.

Recall that in Ghosh and Henderson [2002a] the indefinite matrices Λ_Z were observed to lie very close to the set of feasible correlation matrices for joint normal random vectors (i.e., the set of positive semidefinite matrices with ones on the diagonal). This led to the suggestion in Ghosh and Henderson [2002a] that the setup stage of NORTA be augmented with an SDP that is used, if Λ_Z turns out indefinite, to find a matrix Σ_Z that is "close" to Λ_Z and is positive semidefinite. The matrix Σ_Z is then used within the NORTA method.

Why is this approach reasonable? In Theorem 2 of Cario and Nelson [1997] it is shown that under a certain moment condition, the output correlation matrix is a continuous function of the input covariance matrix Σ_Z used in the NORTA procedure. So if Σ_Z is "close" to Λ_Z, then we can expect the correlation matrix of the NORTA generated random vectors to be close to the desired matrix Σ_X. The moment condition always holds when we are attempting to match rank correlations, and we can expect it to hold almost invariably when matching product-moment correlations. Therefore, it is eminently reasonable to try and minimize some measure of distance $r(\Lambda_Z, \Sigma_Z)$ between Λ_Z and Σ_Z.

The SDP falls under the broad class of matrix completion problems; see Alfakih and Wolkowicz [2000], or Johnson [1990]. For this case, given Λ_Z as data, we wish to choose a symmetric matrix Σ_Z to

$$\begin{aligned} \text{minimize} \quad & r(\Sigma_Z, \Lambda_Z) \\ \text{subject to} \quad & \Sigma_Z \succeq 0, \\ & \Sigma_Z(i,i) = 1. \end{aligned} \tag{15}$$

The metric $r(\cdot,\cdot)$ can be chosen as desired. In particular, choosing either the L_1 metric

$$r(A,B) = \sum_{i>j} |A_{ij} - B_{ij}|$$

or the L_∞ metric

$$r(A,B) = \max_{i>j} |A_{ij} - B_{ij}|$$

makes the minimization problem an SDP-constrained problem with a linear objective function. Efficient algorithms, and public domain codes implementing them, are available for solving semidefinite problems of this type; see Wolkowicz et al. [2000].

Numerical studies conducted in Ghosh and Henderson [2002a] indicate that in 3 dimensions this SDP augmentation yields NORTA generated random vectors with correlation matrices that are close to the desired ones. One might then ask whether this remains the case as the dimension increases.

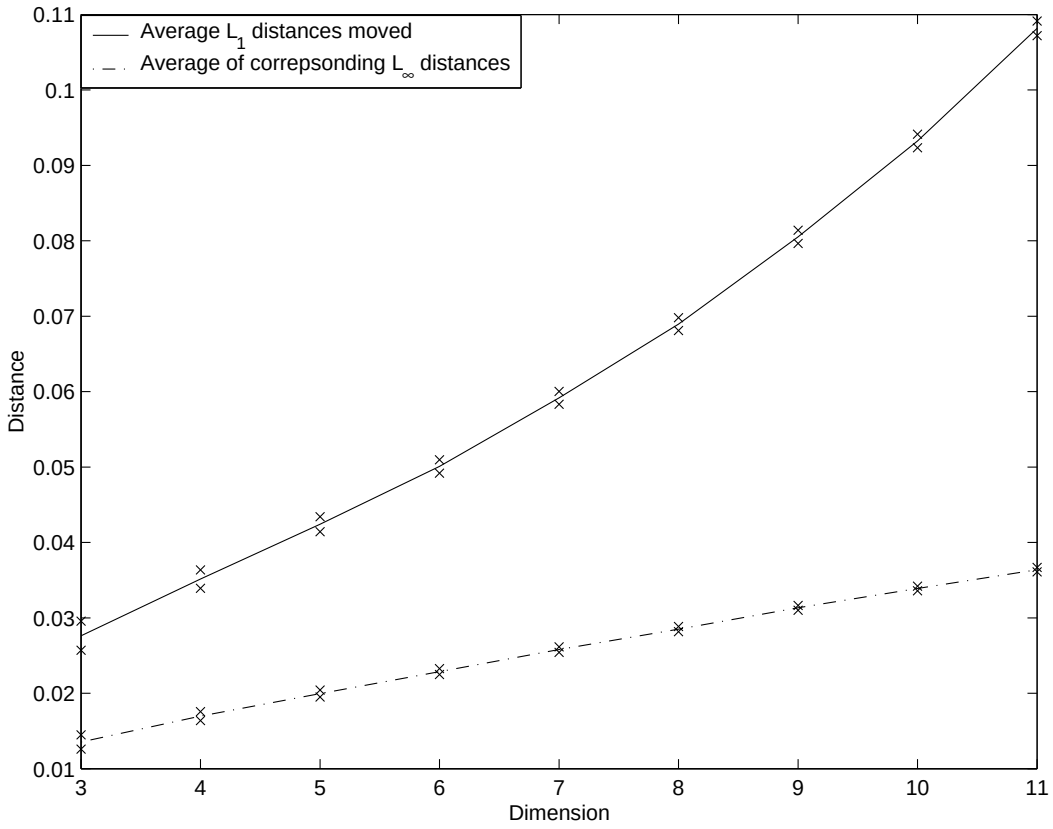

Fig. 2. Performance of the SDP-augmented NORTA in higher dimensions. 15,000 matrices were generated uniformly from Ω and the semidefinite program, with r taken as the L_1 distance, solved for the NORTA defective cases. The solid line gives the expected L_1 distance with 95% confidence intervals as marked, with the average taken only over NORTA defective matrices. The dotted line gives the corresponding expected distance as measured in the L_∞ metric.

We use a setting identical to that used in Section 3 for this study, and our measure of performance is the expected L_1 distance that we have to move from the desired correlation matrix to reach a NORTA feasible one. This means that the minimization problem (15) is solved with $r(\cdot, \cdot)$ as the L_1 metric.

Figure 2 plots the results. We see that the expected L_1 distance increases as the dimension d increases at what might be perceived as a linear rate, although a superlinear rate seems more likely. If the rate of increase is indeed linear, then,

since there are $d(d-1)/2$ matrix entries above the diagonal, the *average* change per entry is (eventually) decreasing with dimension. Of course, it is possible that a small number of entries change by a large amount. The L_∞ distance is also shown, and we see that indeed, at least one entry is changed by an increasing amount as the dimension increases.

It might be preferable from a modelling standpoint to instead minimize the L_∞ distance, so that one tries to minimize the maximum deviation between the achievable and target correlations. The results in this case are shown in Figure 3. We see that the expected L_∞ distance appears to remain constant at around 0.005 or even decrease with dimension. The corresponding L_1 distance appears to still grow at a superlinear rate.

Table I shows the maximum deviation observed in correlations (the L_∞ distance) for a sample of 1000 matrices. The matrices that were found NORTA defective were put through both the SDP-augmented methods, and the maximum deviation in correlations observed is recorded in the table. For the second column the SDP was solved with r as the L_1 metric, while the third column takes r to be the L_∞ metric. The trends observed in the expected values of the L_∞ distances in Figures 2 and 3 are reflected in the maximum values of the L_∞ distances in Table I. The maximum deviation for the case where the SDP is solved for the L_1 metric is seen to grow at a linear rate, while in the L_∞ case the maximum remains constant at around 0.015 or even decreases with dimension. This indicates that the corrective step involved in the SDP-augmented NORTA method appears to do well even in worst cases.

While the total absolute change in correlations seem to grow in either case, Figure 3 suggests that one could attempt a hybrid of the L_1 and L_∞ approaches by, for instance, minimizing the L_1 distance subject to an upper bound on the L_∞ distance, and thus avoid changing any single component of the correlation matrix by too high a value, while keeping the total change within reasonable limits.

We remark here that the SDP framework used here in searching for a "close"

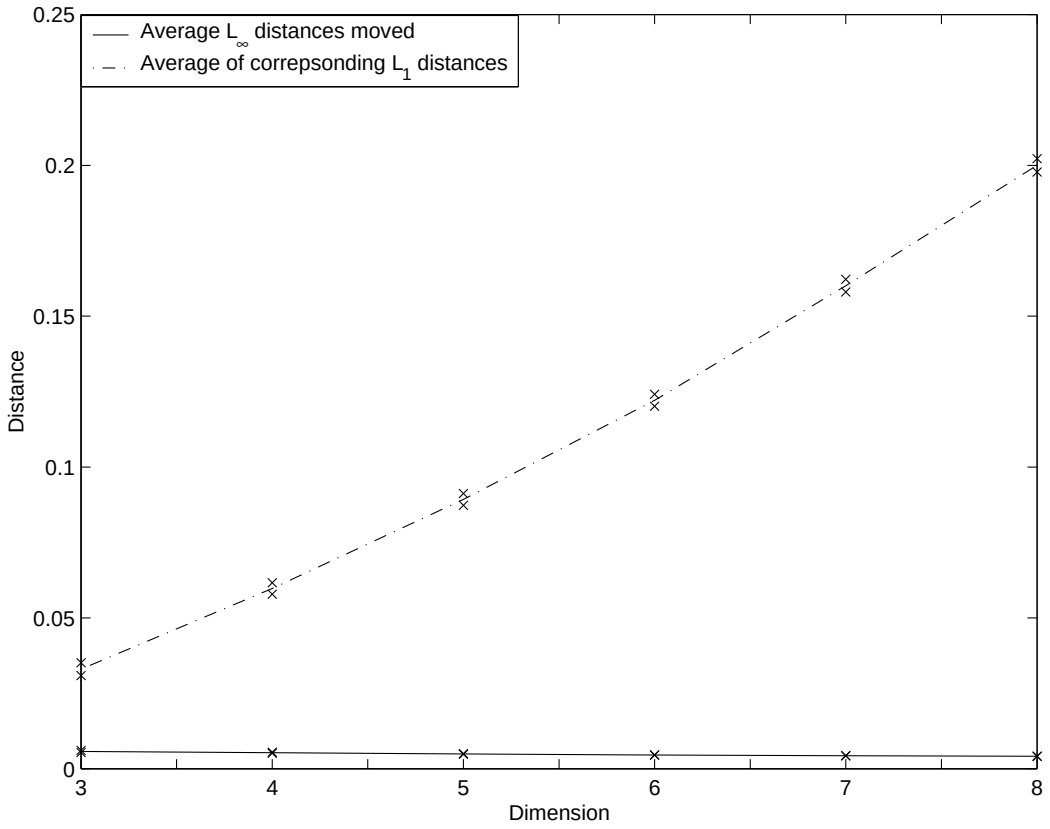

Fig. 3. The SDP-augmented NORTA in higher dimensions. For each dimension, 15,000 matrices were generated uniformly from Ω and the semidefinite program, with r taken as the L_∞ distance, solved for NORTA defective matrices. The solid line gives the expected L_∞ distance with 95% confidence intervals as marked, with the average taken only over NORTA defective matrices. The dotted line gives the corresponding expected distance as measured in the L_1 metric.

positive definite matrix in the SDP problem (15) allows us a certain degree of control on how the search is performed. For instance, we can restrict the change in certain components by adding additional constraints on the SDP.

Thus, the SDP-augmented NORTA problem performs well on average even in higher dimensions. It generates random vectors with correlation matrices which are close to the desired ones, while keeping changes to the individual correlations within reasonable limits.

Table I. Maximum deviations in correlations (L_∞ distances)

	Taking r as	
d	L_1	L_∞
3	0.04042	0.01671
4	0.04513	0.01268
5	0.05527	0.01455
6	0.06901	0.01295
7	0.07227	0.01247
8	0.07893	0.01190
9	0.07676	0.01113
10	0.09149	0.00898
11	0.09876	0.00971

We conclude that despite the feasibility problem, the NORTA method *is* a viable method even in high dimensional problems.

Computational results also show that the SDP problem in the SDP-augmented method is solved within a very reasonable amount of time when the L_1 metric is used as r. However we find that the L_∞ SDP problems, formulated as in Problem 15, become increasingly harder to solve as the dimension increases. So we are presently considering methods to improve on the present formulation. We will also be evaluating other alternatives to the SDP formulation presented above.

REFERENCES

Alfakih, A. and Wolkowicz, H. 2000. Matrix completion problems. In *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*, H. Wolkowicz, R. Saigal, and L. Vandenberghe, Eds. Kluwer, Boston, 533–545.

Cario, M. C. and Nelson, B. L. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Tech. rep., Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Clemen, R. T. and Reilly, T. 1999. Correlations and copulas for decision and risk analysis. *Management Science 45*, 208–224.

Das, S. R., Fong, G., and Geng, G. 2001. Impact of correlated default risk on credit portfolios. *The Journal of Fixed Income 11,* 3, 9–19.

DEVROYE, L. 1986. *Non-uniform Random Variate Generation*. Springer-Verlag, New York.

EDELMAN, A. 1989. Eigenvalues and condition numbers of random matrices. Ph.D. thesis, Department of Mathematics, MIT, Cambridge, MA.

FROBENIUS, F. G. 1968. *Gessammelte Abhandlungen*. Vol. 3. Springer, Berlin.

GHOSH, S. AND HENDERSON, S. G. 2001. Chessboard distributions. In *Proceedings of the 2001 Winter Simulation Conference*, B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, Eds. IEEE, Piscataway NJ, 385–393.

GHOSH, S. AND HENDERSON, S. G. 2002a. Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research 50,* 5, 820–834.

GHOSH, S. AND HENDERSON, S. G. 2002b. Properties of the NORTA method in higher dimensions. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C. H. Chen, J. L. Snowdon, and J. M. Charnes, Eds. IEEE, Piscataway NJ, 263–269.

GUTTMAN, L. 1946. Enlargement methods for computing the inverse matrix. *Annals of Mathematical Statistics 17*, 336–343.

HENDERSON, S. G. 2001. Mathematics for simulation. In *Proceedings of the 2001 Winter Simulation Conference*, B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, Eds. IEEE, Piscataway NJ, 83–94.

HENDERSON, S. G., CHIERA, B. A., AND COOKE, R. M. 2000. Generating "dependent" quasi-random numbers. In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, Eds. IEEE, Piscataway NJ, 527–536.

IMAN, R. AND CONOVER, W. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics: Simulation and Computation 11*, 311–334.

JOHNSON, C. R. 1990. Matrix completion problems: a survey. *Proceedings of Symposia in Applied Mathematics 40*, 171–198.

JOHNSON, M. E. 1987. *Multivariate Statistical Simulation*. Wiley, New York.

KENDALL, M. G. 1961. *A Course in the Geometry of n Dimensions*. Charles Griffin and Co., London.

KRUSKAL, W. 1958. Ordinal measures of associaton. *Journal of the American Statistical Association 53*, 814–861.

KUROWICKA, D. AND COOKE, R. M. 2001. Conditional, partial and rank correlation for the elliptical copula; dependence modelling in uncertainty analysis. Tech. rep., Delft University of Technology, Mekelweg 4, 2628CD Delft, Netherlands.

Law, A. M. and Kelton, W. D. 2000. *Simulation Modeling and Analysis*, 3rd ed. McGraw-Hill, New York.

Li, S. T. and Hammond, J. L. 1975. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man, and Cybernetics 5*, 557–561.

Lurie, P. M. and Goldberg, M. S. 1998. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science 44*, 203–218.

Mardia, K. V. 1970. A translation family of bivariate distributions and Fréchet's bounds. *Sankhya A32*, 119–122.

Marsaglia, G. and Olkin, I. 1984. Generating correlation matrices. *SIAM Journal of Scientific and Statistical Computing 5*, 470–475.

Ouellette, D. V. 1981. Schur complements and statistics. *Linear Algebra and Its Applications 36*, 187–295.

Whitt, W. 1976. Bivariate distributions with given marginals. *The Annals of Statistics 4*, 1280–1289.

Wolkowicz, H., Saigal, R., and Vandenberghe, L., Eds. 2000. *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*. Kluwer, Boston.