Math. Num. Sin.
19:1(1997), 91–102.

Chinese J. Num. Math. & Appl.
19:2(1997), 82–95.

DIMENSION REDUCTION AND L_1 APPROXIMATION FOR EVALUATION OF MULTIVARIATE NORMAL INTEGRAL*[1)]

Yang Zi-qiang Zhang Chun-ming
(*Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Beijing*)

Abstract

In the present paper, the authors suggest an algorithm to evaluate the multivariate normal integrals under the supposition that the correlation matrix R is quasi-decomposable, in which we have $r_{ij} = a_i a_j$ for most i, j, and $r_{ij} = a_i a_j + b_{ij}$ for the others, where b_{ij}'s are the nonzero deviations. The algorithm makes the high-dimensional normal distribution reduce to a 2-dimensional or 3-dimensional integral which can be evaluated by the numerical method with a high precision. Our supposition is close to what we encounter in practice. When correlation matrix is arbitrary, we suggest an approximate algorithm with a medium precision, it is, in general, better than some approximate algorithms. The simulation results of about 20000 high-dimensional integrals showed that the present algorithms were very efficient.

It is a classical and important topic to evaluate the multivariate normal integrals. The early work had been reviewed by Gupta (1963)[1], and the later work by Tong (1990)[2]. Although there were many papers published in this research field, one could not find a consentaneous and efficient algorithm for the *general* multivariate normal integrals except in the cases of bivariate and trivariate (Zhang-Yang, 1993[3]). For the dimension $m > 3$, the articles published were almost concerned in some special cases, such as a special region of integral or a special structure of correlation matrix. Otherwise, the Monte Carlo or quasi Monte Carlo method (i.e. number-theoretic method, Fang-Wang, 1994[4]) was employed. The rate of convergence of quasi Monte Carlo is not over $O(n^{-2/m})$, and it is difficult to obtain a precise evaluation if m is large.

In evaluation of multivariate normal integrals, a well known supposition for the correlation matrix R is that the correlation coefficients r_{ij} can be *decomposed* as $r_{ij} = a_i a_j$ for all $i \neq j$ and $a_i^2 < 1$. This, however, is a severe supposition. In

* Received May 10, 1996.

1) The Project Supported by National Natural Science Foundation of China.

practice a more natural supposition may be that $r_{ij} = a_i a_j$ is only for *most* of r_{ij}, and $r_{ij} = a_i a_j + b_{ij}$ for the others, where b_{ij}'s are the nonzero deviations. R is called a quasi-decomposable matrix if b_{ij}'s satisfy some restrictions (see 1 for definition).

The main results in the present paper are

(1) To establish two dimension reduction theorems used to evaluate the multivariate normal integrals under the supposition that R is a quasi-decomposable matrix. In virtue of theorems, a high-dimensional normal distribution can be reduced to a 2-dimensional or 3-dimensional integral which can be evaluated by the numerical method with a high precision.

(2) To establish an approximate evaluation of multivariate normal integral when R is an arbitrary correlation matrix.

(3) To suggest an efficient Gaussian quadrature formula to evaluate the integral from the dimension reduction theorems. This formula is more efficient than the classical Gauss-Hermite quadrature. Also, a procedure of decomposition of domain of integration is suggested based on this formula.

(4) To give an algorithm to solve the a_i's and b_{ij}'s for a given R. Especially, the L_1 approximation is used.

Our algorithms are complete, on one hand, when a precise evaluation is required, the supposition is natural and friendly, and the L_1 solution makes it easily hold true in many cases, on the other hand, when an approximate evaluation is required, there is no restriction on our algorithm and an evaluation with a medium precision can be obtained. The simulation results of about 20000 multivariate normal integrals showed that the present algorithms were very efficient.

1. Dimension Reduction Theorems

As the conventional notation in multivariate statistical analysis,

$$(X_1, \cdots, X_m)' \sim N_m(\mu, \Sigma)$$

means that the m-dimensional random vector $(X_1, \cdots, X_m)'$ is normally distributed with the mean vector $\mu = (\mu_1, \cdots, \mu_m)'$ and the covariance matrix $\Sigma = (\sigma_{ij})$. For convenience, suppose that $\mu_i = 0$, $\sigma_{ii} = 1$ for all i without loss of generality. Then the covariance matrix becomes a correlation matrix $R = (r_{ij})$. The normal density and distribution functions are denoted $\phi_m(x_1, \cdots, x_m)$ and $\Phi_m(x_1, \cdots, x_m)$ respectively. When $m = 1$, the subscript m can be omitted, hence we have $N(0,1)$, $\phi(x)$ and $\Phi(x)$.

In multivariate statistical analysis, we have a well known result as follows[1,2]:

Lemma. *Suppose that $Y_1, \cdots, Y_m, U$ are independent random variables, each one has a standard normal distribution $N(0,1)$, and*

$$X_i = \sqrt{1 - a_i^2}\, Y_i - a_i U, \qquad \text{constant } a_i^2 < 1,\ i = 1, \cdots, m, \tag{1.1}$$

then

$$(X_1, \cdots, X_m)' \sim N_m(0, R), \tag{1.2}$$

$$\Phi_m(x_1,\cdots,x_m)=\int_{-\infty}^{\infty}\Big[\prod_{i=1}^{m}\Phi\Big(\frac{x_i+a_iu}{\sqrt{1-a_i^2}}\Big)\Big]\phi(u)du \tag{1.3}$$

where

$$r_{ij}=a_ia_j,\quad \forall\, i\neq j. \tag{1.4}$$

From (1.3), $\sqrt{1-a_i^2}$ appears in the denominator, so $a_i^2<1$ is needed. Furthermore, in view of numerical computation, a_i^2 must not close 1. Because the integral

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$$

can be evaluated by a simple expression with a high precision (more than 14 decimal digits, see IMSL), we regard (1.3) as an 1-dimensional integral. The above lemma shows that a high-dimensional normal distribution can be reduced to an 1-dimensional integral. This lemma had been used by many statisticans[1,2]. But, it needs a severe supposition that all r_{ij} can be decomposed as (1.4).

In practice, a more natural supposition may be that

$$r_{ij}=\begin{cases} a_ia_j, & \text{for most of } r_{ij},\\ a_ia_j+b_{ij}, & \text{for a few of } r_{ij},\end{cases}$$

where b_{ij} are the nonzero deviations. For example, if only one r_{ij} is with a (nonzero) deviation as follows[1)]:

$$r_{ij}=\begin{cases} a_ia_j, & i>j,\ (i,j)\neq(i_1,j_1),\\ a_{i_1}a_{j_1}+b_{i_1j_1}, & i=i_1,\ j=j_1,\end{cases} \tag{1.5}$$

we have the following theorem.

Theorem 1. *Suppose that $Y_1,\cdots,Y_m,U,V$ are independent random variables, each one has a standard normal distribution $N(0,1)$, and*

$$\begin{cases} X_i=\sqrt{1-a_i^2}\,Y_i-a_iU, & i\neq i_1,j_1,\\ X_{i_1}=\sqrt{1-a_{i_1}^2-|b_{i_1j_1}|c_{i_1j_1}}\,Y_{i_1}-a_{i_1}U-\pm\sqrt{|b_{i_1j_1}|c_{i_1j_1}}\,V,\\ X_{j_1}=\sqrt{1-a_{j_1}^2-|b_{i_1j_1}|/c_{i_1j_1}}\,Y_{j_1}-a_{j_1}U-\sqrt{|b_{i_1j_1}|/c_{i_1j_1}}\,V,\end{cases} \tag{1.6}$$

where $c_{i_1j_1}>0$, and $\pm$ indicates the sign of radical, i.e. sign$(b_{i_1j_1})$, then

$$(X_1,\cdots,X_m)'\ \sim\ N_m(0,R),$$

[1)]Since correlation matrix is symmetric and $r_{ii}=1$, we only use its lower triangle, then $i>j$.

where R is defined as (1.5), and the distribution function

$$\Phi_m(x_1,\cdots,x_m)=\int_{-\infty}^{\infty} g_{i_1j_1}(u)[\prod_{\substack{i=1\\ i\neq i_1,j_1}}^{m}\Phi(\frac{x_i+a_iu}{\sqrt{1-a_i^2}})]\phi(u)du, \tag{1.7}$$

$$g_{i_1j_1}(u)=\int_{-\infty}^{\infty}\Phi(\frac{x_{i_1}+a_{i_1}u\pm\sqrt{|b_{i_1j_1}|c_{i_1j_1}}\,v}{\sqrt{1-a_{i_1}^2-|b_{i_1j_1}|c_{i_1j_1}}})\Phi(\frac{x_{j_1}+a_{j_1}u+\sqrt{|b_{i_1j_1}|/c_{i_1j_1}}\,v}{\sqrt{1-a_{j_1}^2-|b_{i_1j_1}|/c_{i_1j_1}}})\phi(v)dv. \tag{1.8}$$

This theorem reduces Φ_m to a 2-dimensional integral. The proof is similar to the lemma. Obviously, we have the following restrictions

$$a_i^2<1,\quad a_{i_1}^2+|b_{i_1j_1}|c_{i_1j_1}<1,\quad a_{j_1}^2+|b_{i_1j_1}|/c_{i_1j_1}<1. \tag{1.9}$$

$c_{i_1j_1}$ is usually equal to 1. When the above restrictions do not satisfy or their values are too close 1, we can set $c_{i_1j_1}$ the other positive value. In other words, $c_{i_1j_1}>0$ makes the restrictions easy.

The remainder of this section we discuss the case where the number of deviations is more than one. We are concerned with two kinds of deviations as follows:

(1) All subscripts in nonzero deviations are different, e.g. b_{21}, b_{53}, b_{76}, b_{94}.

(2) There are some common subscripts in the nonzero deviations, but any subscript cannot appear in more than two nonzero deviations, e.g. (b_{21}, b_{31}), (b_{74}, b_{76}), (b_{85}, b_{98}).

Let n_1 be the number of nonzero deviations in the first kind, and n_2 the second. Where n_2=0,2,4,6,$\cdots$. Two kinds of deviations are denoted as $b_{i_tj_t}$, $t=1,\cdots,n_1$ and ($b_{i'_sj'_s}$, $b_{k'_sj'_s}$), $s=1,\cdots,n_2/2$ respectively. Although, we suppose that the common subscript j'_s appears in the second position, the following discussion holds true for the other positions.

Definition. *$R_{m\times m}$ is called a quasi-decomposable matrix if it is a positive definite symmetric matrix and its elements are as follows:*

$$r_{ij}=\begin{cases}1, & i=j,\\ a_ia_j, & i>j,\ \text{and } i,j\neq i_t,j_t,i'_s,j'_s,k'_s,\\ a_{i_t}a_{j_t}+b_{i_tj_t}, & i=i_t,\ j=j_t,\ t=1,\cdots,n_1,\\ a_{i'_s}a_{j'_s}+b_{i'_sj'_s}, & i=i'_s,\ j=j'_s,\ s=1,\cdots,n_2/2,\\ a_{k'_s}a_{j'_s}+b_{k'_sj'_s}, & i=k'_s,\ j=j'_s,\ i'_s,j'_s,k'_s\neq i_t,j_t,\end{cases} \tag{1.10}$$

where the number of deviations is $n=n_1+n_2$, and $2n_1+3(n_2/2)\leq m$. Obviously, in a quasi-decomposable matrix, any subscript of nonzero deviations cannot appear in more than two nonzero deviations.

Theorem 2. *Suppose* $(X_1, \cdots, X_m)' \doteq N_m(0, R)$, *where* R *is a quasi-decomposable matrix with* n *nonzero deviations above, then we have*

$$\Phi_m(x_1, \cdots, x_m) = \int_{-\infty}^{\infty} \Big[\prod_{t=1}^{n_1} g_{i_t j_t}(u)\Big]\Big[\prod_{s=1}^{n_2/2} h_{i'_s j'_s k'_s}(u)\Big]\Big[\prod_{\substack{i\neq i_t, j_t \\ i'_s, j'_s, k'_s}}^{m} \Phi\Big(\frac{x_i + a_i u}{\sqrt{1-a_i^2}}\Big)\Big]\phi(u)du, \tag{1.11}$$

where

$$g_{i_t j_t}(u) = \int_{-\infty}^{\infty} \Phi\Big(\frac{x_{i_t} + a_{i_t}u \pm \sqrt{|b_{i_t j_t}|c_{i_t j_t}}\, v}{\sqrt{1 - a_{i_t}^2 - |b_{i_t j_t}|c_{i_t j_t}}}\Big)\Phi\Big(\frac{x_{j_t} + a_{j_t}u + \sqrt{|b_{i_t j_t}|/c_{i_t j_t}}\, v}{\sqrt{1 - a_{j_t}^2 - |b_{i_t j_t}|/c_{i_t j_t}}}\Big)\phi(v)dv, \tag{1.12}$$

$$\begin{aligned} h_{i'_s j'_s k'_s}(u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} &\Phi\Big(\frac{x_{i'_s} + a_{i'_s}u \pm \sqrt{|b_{i'_s j'_s}|c_{i'_s j'_s}}\, v}{\sqrt{1 - a_{i'_s}^2 - |b_{i'_s j'_s}|c_{i'_s j'_s}}}\Big)\Phi\Big(\frac{x_{k'_s} + a_{k'_s}u \pm \sqrt{|b_{k'_s j'_s}|c_{k'_s j'_s}}\, w}{\sqrt{1 - a_{k'_s}^2 - |b_{k'_s j'_s}|c_{k'_s j'_s}}}\Big) \\ &\cdot\Phi\Big(\frac{x_{j'_s} + a_{j'_s}u + \sqrt{|b_{i'_s j'_s}|/c_{i'_s j'_s}}\, v + \sqrt{|b_{k'_s j'_s}|/c_{k'_s j'_s}}\, w}{\sqrt{1 - a_{j'_s}^2 - |b_{i'_s j'_s}|/c_{i'_s j'_s} - |b_{k'_s j'_s}|/c_{k'_s j'_s}}}\Big)\phi(v)\phi(w)dvdw \end{aligned} \tag{1.13}$$

and the restrictions are

$$\begin{cases} a_i^2 < 1, \\ a_{i_t}^2 + |b_{i_t j_t}|c_{i_t j_t} < 1, \qquad a_{j_t}^2 + |b_{i_t j_t}|/c_{i_t j_t} < 1, \\ a_{i'_s}^2 + |b_{i'_s j'_s}|c_{i'_s j'_s} < 1, \qquad a_{k'_s}^2 + |b_{k'_s j'_s}|c_{k'_s j'_s} < 1, \\ a_{j'_s}^2 + |b_{i'_s j'_s}|/c_{i'_s j'_s} + |b_{k'_s j'_s}|/c_{k'_s j'_s} < 1, \\ c_{ij} > 0, \ \forall\, i \neq j. \end{cases} \tag{1.14}$$

The proof of the theorem is not difficult. It is based on theorem 1 and the following transform of variables concerning with the deviations in the second kind:

$$\begin{cases} X_{i'_s} = \sqrt{1 - a_{i'_s}^2 - |b_{i'_s j'_s}|c_{i'_s j'_s}}\; Y_{i'_s} - a_{i'_s}U \doteq \pm\sqrt{|b_{i'_s j'_s}|c_{i'_s j'_s}}\, V_s, \\ X_{k'_s} = \sqrt{1 - a_{k'_s}^2 - |b_{k'_s j'_s}|c_{k'_s j'_s}}\; Y_{k'_s} - a_{k'_s}U - \pm\sqrt{|b_{k'_s j'_s}|c_{k'_s j'_s}}\, W_s, \\ X_{j'_s} = \sqrt{1 - a_{j'_s}^2 - |b_{i'_s j'_s}|/c_{i'_s j'_s} - |b_{k'_s j'_s}|/c_{k'_s j'_s}}\; Y_{j'_s} \\ \qquad -a_{j'_s}U - \sqrt{|b_{i'_s j'_s}|/c_{i'_s j'_s}}\, V_s - \sqrt{|b_{k'_s j'_s}|/c_{k'_s j'_s}}\, W_s, \end{cases} \tag{1.15}$$

where $Y_{i'_s}$, $Y_{j'_s}$, $Y_{k'_s}$, U, V_s, W_s all are independent $N(0,1)$. The supposition in this theorem may be the most friendly and it is close to what we encounter in practice. In virtue of this theorem, the high-dimensional normal integral can be reduced to a 2-dimensional integral (for $n_2 = 0$) or 3-dimensional integral (for $n_2 \neq 0$), hence, it can be evaluated precisely by Gaussian quadrature (see Section 3).

In principle, we can establish other dimension reduction theorems used for the cases that a common subscript appears in more than two nonzero deviations. To

avoid the evaluation of high-dimensional integral in theorems we can, however, employ the approximate algorithm described in the next section.

2. Approximate Algorithm – Indirect Application of Theorems

When R is not a quasi-decomposable matrix or we do not want to evaluate the 3-dimensional integral in (1.11), the following approximate algorithm can be employed.

Approximate formula. Let $(X_1, \cdots, X_m)' \sim N_m(0, R)$. Its distribution function $\Phi_m(x_1, \cdots, x_m)$ is to be evaluated. Suppose that n correlation coefficients contain nonzero deviations as follows:

$$r_{ij} = \begin{cases} a_i a_j, & i > j, \text{ and } (i,j) \neq (i_t, j_t), \\ a_{i_t} a_{j_t} + b_{i_t j_t}, & i = i_t,\ j = j_t,\ t = 1, \cdots, n. \end{cases} \tag{2.1}$$

Here, the deviations may not belong to the two kinds discussed in Section 1, especially, the common subscripts can appear in more than two nonzero deviations. In other words, R can be an arbitrary correlation matrix. Based on R we construct $n+1$ correlation matrices $R^{(0)}$, $R^{(t)}$, $t = 1, 2, \cdots, n$. Firstly, $R^{(0)}$ is such one that all deviations are eliminated, namely, $R^{(0)}$ can be decomposed completely. Let $\Phi_m^{(0)}(x_1, \cdots, x_m)$ be the distribution of $N_m(0, R^{(0)})$, it can be evaluated by the lemma in Section 1. Secondly, in $R^{(t)}$, all but one of elements are equal to $R^{(0)}$, and the different element is $b_{i_t j_t}$. The distribution of $N_m(0, R^{(t)})$ denoted $\Phi_m^{(t)}(x_1, \cdots, x_m)$ can be evaluated by the theorem 1 in Section 1. Suppose that the contribution of each deviation $b_{i_t j_t}$ is added up independently, we have

$$\begin{aligned} &\Phi_m(x_1, \cdots, x_m) \\ &\quad = \Phi_m^{(0)}(x_1, \cdots, x_m) + \sum_{t=1}^{n} [\Phi_m^{(t)}(x_1, \cdots, x_m) - \Phi_m^{(0)}(x_1, \cdots, x_m)]. \end{aligned} \tag{2.2}$$

Usually, the above supposition does not satisfy and (2.2) is only an approximate evaluation for a general multivariate normal integral.

Special Case. The Formula (2.2) has a special case, in which $a_i = 0$ for all i, i.e. $b_{ij} = r_{ij}$ for all $i \neq j$, then, $n = m(m-1)/2$, and $\Phi_m^{(0)}(x_1, \cdots, x_m)$ is the product of m independent $\Phi_1(x_i)$. That is just the method suggested by Drezner[6], and Drezner used Monte Carlo simulation to show the availability of his algorithm. According to his simulation, the absolute errors were not greater than 0.001 when $|r_{ij}| = |b_{ij}| \leq 0.2$. For our formula (2.2), $a_i \neq 0$ and $n < m(m-1)/2$, it should be better than Drezner's, and this suggestion had been confirmed by our simulation (see Section 5). In one word, our approximate algorithm is useful when deviations $|b_{i_t j_t}|$, $t = 1, \cdots, n$ are not large.

Generalization. The approximation (2.2) has a generalized type, in which every $R^{(t)}$ is a quasi-decomposable matrix as (1.10), especially, we have only

one $R^{(t)}$. The following is an example. Suppose that there exist 4 deviations as b_{21}, b_{31}, b_{42}, b_{43}, we let $R^{(1)}$ contain (b_{21}, b_{43}), and $R^{(2)}$ contain (b_{31}, b_{42}). By the way, such a grouping is better than others because there is no common subscript in each group, and only 2-dimensional integrals appear in theorem 2.

3. Numerical Method for Integral from Theorem 2

3.1 More Efficient Quadrature Formula

Each dimension in integral generated by the dimension reduction theorems has the form $\int_{-\infty}^{\infty} f(x)e^{-x^2}dx$, and the classical Gauss-Hermite quadrature formula

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx = \sum_{i=1}^{N} w_i^* f(x_i^*) \tag{3.1}$$

was employed by most people, where N is the number of nodes, x_i^* and w_i^* are the node and weight respectively. The table of $w_i^* - x_i^*$ can be found in many books of numerical computation, e.g., Stroud-Secrest[7]. Gauss-Hermite quarature is of order $(2N-1)$, i.e. it is exact if $f(x)$ is a polynomial of degree at most $2n-1$. Steen-Byrne-Gelbard (abbreviated SBG)[8] Suggested another Gaussian quadrature as follows:

$$\int_{0}^{\infty} f(x)e^{-x^2}dx = \sum_{i=1}^{N} w_i f(x_i). \tag{3.2}$$

It is order $(2N-1)$ in the interval $(0,\infty)$ (not $(-\infty,\infty)$). Obviously, (3.2) can also be used for the integral in $(-\infty,\infty)$, and

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}dx = \int_{0}^{\infty} [f(x)+f(-x)]e^{-x^2}dx = \sum_{i=1}^{N} w_i[f(x_i)+f(-x_i)]. \tag{3.3}$$

Should we use either formula (3.1) or (3.3) to evalute the integral generated by the dimension reduction theorems? From a first glance, in order to attain order $(2N-1)$ in $(-\infty,\infty)$, SBG (3.3) needs $2N$ values of integrand when $f(x)$ is not an even function, while the classical Gauss-Hermite formula (3.1) needs N values only, and (3.1) should, in general, be used. However, the conclusion may be changed when $f(x)$ is not sufficient smooth. For example, $\int_{-\infty}^{\infty} |x|e^{-x^2}dx = 1$. If we set $n = 2, 4$ and 6 in numerical computation, Gauss-Hermite results contain the relative errors 25.3%, 11.3% and 7.3% respectively, while SBG results are all precise[8]. It is worth to point out that the superior of SBG formula also appears in the integral from our dimension reduction theorems. The present author (Yang [9]) had computed 400 8-dimensional normal integrals generated by random method and found that the total executive time of SBG (3.3) was shorter than classical (3.1) (ratio was about 1/2). Also, Yang found that the

classical formula slightly better than SBG only in the case that the condition of correlation matrix R was very good (*i.e.* $|R| > 0.5$). But the case $|R| > 0.5$ are rare in 8-dimensional normal integral (In fact, if all correlations are equal to r, then $|R| > 0.5$ corresponds to $|r| < 0.2$, and its frequency in simulation was only 7%.) and two executive times have not significant difference for a small number of nodes. Conversely, in case $|R| \leq 0.5$ (its frequency in simulation was 93%), the number of nodes used for classical formula was larger than SBG, and the difference between two executive times would increase quickly if the condition of R was bad.

In one word, Steen-Byrne-Gelbard formula was more efficient than the classical one in the numerical computation of high-dimensional normal integrals. By the way, the original table $w_i - x_i$ given by Steen-Byrne-Gelbard was small ($N \leq 15$). The present author (Yang [9]) had given a new table ($N \leq 60$).

3.2 Decomposition of Domain of Integration based on SBG

Now, we discuss the decomposition of domain of integration. We, first, partition the $m-$dimensional domain as 2^k ($k = 1, 2, \cdots$) sub-domains with the equal volume, *e.g.* two-dimensional plane can be partitioned as 4 quadrants, then, evaluate each integral in the sub-domains, and sum them. Because the features of integrand in sub-domains are not the same, the number of nodes used for evaluation of each sub-integral is not the same too, and the total number will be less than that based on a full domain. For the SBG formula in Section 3.1 is based on a half infinite interval and with the above advantages, it is proper to used this formula to evaluate each sub-integral.

The following we explain the decomposition of domain of integration based on SBG by an example. In (1.11), let $n_1 = 1$, $n_2 = 2$ and denote the integral as

$$\Phi_m = \int_{-\infty}^{\infty} g_v(u) h_{v^* w}(u) f(u) du, \tag{3.4}$$

where the subscript v of $g_v(u)$ is used to indicate the integral variable (see (1.12)). Let v_- and v_+ indicate the integrals in $(-\infty, 0)$ and $(0, \infty)$ respectively. We have $g_v(u) = g_{v_-}(u) + g_{v_+}(u)$. Similarly, let $h_{v^* w}(u) = h_{v_-^* w}(u) + h_{v_+^* w}(u)$, the integral (3.4) becomes a sum of 4 sub-integrals:

$$\begin{aligned}\Phi_m = \int_{-\infty}^{\infty} [&g_{v_-}(u) h_{v_-^* w}(u) + g_{v_-}(u) h_{v_+^* w}(u) \\ &+ g_{v_+}(u) h_{v_-^* w}(u) + g_{v_+}(u) h_{v_+^* w}(u)] f(u) du.\end{aligned}$$

We abbreviate the above formula as $\Phi_m = I_{--} + I_{-+} + I_{+-} + I_{++}$. Similarly, if we also partition the integral variable u, Φ_m becomes a sum of 8 sub-integrals: $\Phi_m = I_{---} + I_{--+} + I_{-+-} + I_{-++} + I_{+--} + I_{+-+} + I_{++-} + I_{+++}$. If we further partition variable w in $h_{v^* w}(u)$, Φ_m consists of sum of 16 sub-integrals. In a formula, each sub-domain of integration has a same volume, and each variable is

in $(-\infty, 0)$ or $(0, \infty)$. Obviously, the above procedure of decomposition of domain can be used for not only a single processor but also the multiple processors.

It is worth to point out that the excessive decomposition of domain could not reach a high efficiency in parallel process for the repeated operations. In serial process, you can save some results created in the preceding computing to avoid the repeated operations. It is the best to decompose variable u because of no repeated operation in this case. From experience (see Section 5), we could use three decompositions as follows: (1) variable u; (2) both u and v for $n_1 = 1$; (3) both u and v^* for $n_2 = 2$.

4. Decomposition of Correlation Matrix Under L_1

In this section, we discuss a method to decompose a correlation matrix $R_{m\times m}$ and get $\boldsymbol{a} = (a_1, \cdots, a_m)'$ and b_{ij}'s which satisfy $r_{ij} = a_i a_j + b_{ij}$. We want such a method that it could give the exact $\boldsymbol{a}$ and b_{ij}'s if R was a quasi-decomposable matrix as (1.10), otherwise, it could minimize the number of nonzero b_{ij}'s as well as the values of $|b_{ij}|$'s. The latter makes the approximation (2.2) more efficient. But, it is very difficult to find such a mathematical method. In the present section, a L_1 model

$$\begin{cases} Q(\boldsymbol{a}) = \sum\limits_{i=1}^{m}\sum\limits_{\substack{j=1\\ j\neq i}}^{m} |r_{ij} - a_i a_j| = \min, \\ |a_i| < 1, \quad i = 1, \cdots, m \end{cases} \tag{4.1}$$

serves as a substitute for such method. We select L_1 rather than L_2, i.e. the traditional least square method, because the latter makes the deviations b_{ij}'s appear in almost every r_{ij} if R is not completely decomposable. However, the L_1 solution can make some deviations zero, which means that the quasi-decomposable supposition can be satisfied easily or the approximation (2.2) can be improved. For example, $R_{6\times 6}$ is a quasi-decomposable matrix, and $r_{ij} = a_i a_j$ except r_{21} and r_{41}. Let

$$\boldsymbol{a} = (0.32, 0.45, 0.61, -0.85, 0.52, -0.95),\ b_{21} = -0.2027,\ b_{41} = 0.2807.$$

Under L_2 we get b_{ij}'s as follows:

$$\begin{array}{cccccccc} -.142 & .077 & .001 & .166 & -.019 & -.003 & .066 & .002 \\ -.012 & -.004 & -.123 & -.010 & .012 & .021 & .009 & \end{array}$$

They are all inequal to zero. But, using L_1 algorithm described below we obtain the exact $\boldsymbol{a}$ and b_{ij}'s as they are. Sometimes, we could not obtain such an ideal solution, but we always have several zero deviations and the minimum $\sum |b_{ij}|$. The simulation study in Section 5 shows that the L_1 method is good.

(4.1) can be transformed to an iteratively reweighted least square model as follows:

$$\begin{cases} Q(a) = \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} w_{ij}(r_{ij} - a_i a_j)^2 = \min, \\ w_{ij} = 1/(|r_{ij} - a_i a_j| + \text{eps}), \quad i \neq j = 1, \cdots, m, \\ a_i^2 < 1, \quad i = 1, \cdots, m, \end{cases} \tag{4.2}$$

where eps is a small positive value to avoid the zero denominator. The readers can see other papers given by the present author (Yang [10,11]) for the solution of L_1 model.

5. Simulation Study

In this section, we show the performance of the algorithm we suggest. About 20000 multivariate normal integrals generated by Monte Carlo method were evaluated, in which there are over 4000 integrals with dimension $m \geq 8$. Simulation study showed that the present algorithms were efficient.

Quadrature formula and table $w_i - x_i$. A product rule was used to evaluate the high-dimensional integral. For each dimension, we used Steen-Byrne-Gelbard quadrature (3.2) or (3.3), and the table $w_i - x_i$ expanded by Yang[9]. Noting $0 < f(x) < 1/\sqrt{\pi}$, where $f(x)$ is an integrand, we can omit some values of table $w_i - x_i$ for w_i too small. This would save executive time and space. According to the simulated results by Yang[9], if we omitted the terms corresponding to $w_i < 0.1 \times 10^{-9}$, the error of evaluation would not exceed 0.1×10^{-9}.

Convergence criterion and N – required number of nodes. The computation of integral started with N_0 nodes, and then increased 2 nodes each time until the difference between two neighbour results was less than eps (tolerance). To avoid a false convergence in computing, we required another evaluation with one more node. If and only if the new difference was also less than eps, the procedure stopped, and the number of nodes used for the last evaluation was recorded as N. It is worth to point out that N was only the number of nodes in one dimensional interval $(-\infty, 0)$ or $(0, \infty)$. Hence, the number of nodes in $(-\infty, \infty)$ was $2N$. Because the same N was used for every dimension in computing, the number of nodes was N^2 or $4N^2$ for 2-dimensional integral and N^3 or $8N^3$ for 3-dimensional integral.

5.1 Three-Dimensional Normal Integrals

When dimension $m = 3$, the correlation matrix R is always quasi-decomposable and the Theorem 1 in Section 1 can be used directly. First, for a famous example discussed by Steck[12], we obtained $\Phi_3(1.2, 1.0, -0.5) = 0.220609581$. In this example, $R(r_{21}, r_{31}, r_{32}) = (0.7, 0.2, -0.4)$, we took $a_1 = 0.95$, $a_2 = r_{21}/a_1$, $a_3 = r_{31}/a_1$, $b_{32} = r_{32} - a_3 a_2$, $c_{32} = 1.5$ and tolerance eps $= 0.5 \times 10^{-8}$. In computing, $N = 22$ when the domain of integration did not partition and the executive time was 0.55" (Pentium 90). When the domain was partitioned as four quadrants, we needed $N = (16, 18, 22, 18)$ and 0.44". If eps $= 0.5 \times 10^{-6}$, we had the evaluation

0.2206096, and the values of N all reduced 4 and the times were 0.34" and 0.27".

Secondly, we generated 10000 R matrices, in which $r_{ij} \sim U(-1,1)$, i.e. r_{ij} was an uniform random number in interval $(-1,1)$. Using a similar procedure (sometimes, we started with an appropriate value of a_2 or a_3), we obtained a precise evaluation for every positive definite matrix R without ill-condition. In computing, we regarded R as the ill-conditioned matrix if its minimal eigenvalue λ_3 failed in the interval $(0, 0.0015)$. In above 10000 random matrices, the number of positive definite matrices was 6195, and the number of ill-conditioned matrices was 21 only.

5.2 Case of Dimension $m \geq 4$

Two groups of examples. We listed two groups of examples of high-dimensional normal integrals, their dimensions were up to 12. In these examples, the correlation matrices were all quasi-decomposable as (1.10), we could, then, evaluate every integral using Theorem 2. The maximum number of nonzero deviations in examples was up to 6, and the second group of examples contained a common subscript 1. Perhaps, these examples could be used to assess some algorithms for lack of an available benchmark of high-dimensional normal integrals. The parameters a_i's, deviations b_{ij}'s, upper bounds x_i's and the evaluations of integrals were all listed in the table 5.1. Where Φ_m's were the evaluations with more than 8 significant digits given by Theorem 2, and A_m's were the approximations from (2.2). According to the table, the error of approximation A_m is small (about 10^{-5}). So, (2.2) can be used in a lot of cases. By the way, due to limited space, we made the difference between parameters in neighbour integrals only appear in the last dimensions, e.g. Φ_{10} is from Φ_{12} in which the last two dimensions are omitted.

Table 5.1 Two groups of examples of high-dimensional normal integrals

Group 1					Group 2				
i	a_i	b_{ij}	x_i	Φ_m and A_m	i	a_i	b_{ij}	x_i	Φ_m and A_m
1	-.95		2.46		1	-.33		-.05	
2	-.63	b_{21}=.06	2.06		2	.81	b_{21}=-.02	1.79	
3	.19		-.33	.354965422 Φ_4	3	-.80		2.93	
4	-.82	b_{43}=-.11	2.35	.354976 A_4	4	-.21	b_{43}=.05	-.11	.213636260 Φ_5
5	.42		1.64	.322708218 Φ_6	5	.80	b_{51}=.29	3.39	.213628 A_5
6	-.17	b_{65}=-.34	1.69	.322718 A_6	6	.25		1.31	.190555838 Φ_7
7	-.84		2.31	.238884528 Φ_8	7	.81	b_{76}=.16	3.39	.190557 A_7
8	-.62	b_{87}=.11	.72	.238891 A_8	8	-.73		2.44	.159269374 Φ_9
9	.27		2.38	.236778173 Φ_{10}	9	.09	b_{98}=.17	1.01	.159269 A_9
10	-.49	$b_{10,9}$=.39	3.43	.236784 A_{10}	10	-.77		2.30	.106984813 Φ_{11}
11	-.74		.41	.152603476 Φ_{12}	11	.44	$b_{11,10}$=.38	0.56	.106984 A_{11}
12	-.46	$b_{12,11}$=-.37	1.46	.152604 A_{12}	12				

Decomposition of domain of integration. Using Monte Carlo method we generated 100 normal integrals with dimension 8. The upper bounds of integral were uniformly distributed in interval $(-0.5, 3.5)$, i.e. $x_i \sim U(-0.5,\ 3.5)$. The correlation matrices R were quasi-decomposable, and $a_i \sim U(-1,1)$, $i = 1, 2, \cdots, 8$.

They contained three nonzero deviations b_{21}, b_{43} and b_{53} ($n_1 = 1,\ n_2 = 2$). From the construction of a_i's above, the correlation matrix, in general, had not a good condition for the probability $P(\text{at least one}|a_i| > 0.9) = 1 - 0.9^8 = 0.57$. In order to avoid the non-positive definite R, we formed the deviations as follows:

$$|b_{21}| = \frac{1}{2}\min(1 - a_1^2, 1 - a_2^2), \quad |b_{43}| = 2^{-k}(1 - a_4^2), \quad |b_{53}| = 2^{-k}(1 - a_5^2),$$

where k was the minimum integer satisfied $1 - a_3^2 - 2^{-k}[(1 - a_4^2) + (1 - a_5^2)] > 0$. The computing was on PVM (Parallel Virtual Machine) combined several SGI-INDIGO computers. We started from $N_0 = 6$ and let eps $= 0.5 \times 10^{-4}$, the job time had been showed in the Table 5.2. According the table, decomposition of domain could save job time even in the case of single processor, and the speed up had a good value $46.78/28.06 = 1.67$ in two processors. But the value 2.45, in four processors, was not very good. Perhaps, the job time depended on that sub-domain in which N, the number of nodes, was the maximum, especially, there were several integrals in which the correlation matrices were with a bad condition (more than one $|a_i| > 0.95$). In this case, value N in some one of sub-domains was considerably larger than others, hence, most processors were in wait state for a long time.

Table 5.2 Job time of 100 normal integrals with dimension 8

procedure	Job time Number of processors			Speed-up Number of processors	
procedure	1	2	4	2	4
no decomposing	63.09"				
decomposing (u, v^*)	46.78"	28.06"	19.09"	1.67	2.45

Table 5.3 Error frequency of approximation (2.2)

deviations b_{ij}	Two deviations (b_{21}, b_{43})							$\frac{1}{2}m$ deviations			5 b_{ij}'s
dimension m	4	5	6	7	8	10	12	8	10	12	10
sample size	200	200	200	200	200	200	200	200	250	300	250
no. of integrals	151	163	155	165	162	163	156	120	137	159	130
freq $\in$ [0, .000001)	50	34	30	29	48	59	62	8	13	8	12
[.000001, .000005)	21	18	20	21	18	10	16	12	12	16	15
[.000005, .000010)	5	10	13	10	7	5	4	12	9	10	13
[.000010, .000025)	5	12	10	11	10	13	8	20	12	16	7
[.000025, .000050)	6	6	6	10	7	3	3	9	15	11	12
[.000050, .000075)	3	4	3	4	2	4	1	11	9	9	9
[.000075, .000100)	1	2	4	1	2	0	1	7	5	4	5
[.000100, .000250)	7	9	8	10	3	4	3	12	15	14	14
[.000250, .000500)	1	3	5	1	1	1	1	6	4	8	6
[.000500, .000750)	1	1	0	1	1	1	1	3	3	1	3
[.000750, .001000)		1	1	1	1				2	1	3
[.001000, .002500)				1					1	2	1
cumulative freq.	100	100	100	100	100	100	100	100	100	100	100

Quasi-decomposable matrix. Two simulations based on dimension reduction theorem and the approximation (2.2) were included. Let $r_{ij} = a_i a_j + b_{ij}$,

a_i's be random, and $a_i \sim U(-1,1)$; the nonzero deviation b_{ij}'s were random too, $b_{ij} \sim U(-.45,\ .45)$. For simplicity of programming, we set the number and the positions of b_{ij}'s as follows: (1) two deviations (b_{21}, b_{43}) for $m \geq 4$; (2) $\frac{1}{2}m$ deviations $b_{21}, b_{43}, b_{65}, b_{87}, \cdots$ (without common subscript) for $m = 8, 10, 12$; (3) five deviations $b_{43}, b_{52}, b_{62}, b_{87}, b_{91}$ (with a common subscript 2) for $m = 10$. In simulation, 200 (or 250, 300, see Table 5.3) random matrices were generated for every dimension m. Because the non-positive or ill-conditioned matrices (the minimal eigenvalue $\lambda_{min} < 0.0015$) should be deleted, the number of integrals to be evaluated was less than 200. In computing, we first used Theorem 2 to evaluate an integral, and the result was regarded as a 'true' value, and then we used approximation (2.2) (each $R^{(t)}$ contained only one deviation) to evaluate the same integral again, in the latter, a_i's and b_{ij}'s were regarded as 'unknown', and they must be solved by L_1 algorithm described in Section 4. The error in Table 5.3 was the difference between the 'true' value and the approximation. We had used several seeds in simulation, but their results had not significant difference, and the Table 5.3 was arbitrary one taken from them.

In the table, there was an interesting phenomenon that the error was not monotone increasing with the dimension m if the number of deviations was not changed (see the case of two deviations). Perhaps, the reason could be found by considering the process of evaluation of integral. In fact, before using approximation (2.2), we employed the decomposition technique under L_1 to get the a_i's and b_{ij}'s. When m was large enough, the degrees of freedom of L_1 solution was large too because the sum of deviations added to the R was not changed. In other words, $\sum |b_{ij}|$ from L_1 might less than the original value assigned. This phenomenon implicated that our algorithm had a strong power when dimension m was large.

Arbitrary high-dimensional normal integral. A difficulty to asses an algorithm used for the arbitrary high-dimensional normal integrals is that we do not know the accurate results of those integrals. To asses our approximation (2.2) we defined the ***half-difference*** as follows: supposed that the correlations r_{ij}'s $(i \neq j)$ had only two different values a and $a+b$, and the two sides were equal in number denoted n. There were two procedures to evaluate an integral. (1) Forward computing. In this case, the elements of matrix $R^{(0)}$ were $R^{(0)}_{ij} = a$, $\forall (i \neq j)$. The value of b was regarded as a deviation, hence we had n matrices $R^{(t)}$, the difference between $R^{(0)}$ and $R^{(t)}$ was only in one element, and this element had value $a+b$. (2) Backward computing. $R^{(0)}_{ij} = a+b$, $\forall (i \neq j)$. The value of deviation was $-b$. We also had n matrices $R^{(t)}$, the difference between $R^{(0)}$ and $R^{(t)}$ was only in one element, and this element had value a. Denoted the above two evaluations for the same integral as Φ^+_m and Φ^-_m respectively, and let $\frac{1}{2}(\Phi^+_m + \Phi^-_m)$ be the final approximation, the half-difference $\frac{1}{2}|\Phi^+_m - \Phi^-_m|$ was then used for assessment. The simulation results were listed in Table 5.4, in which

the sample size was 100, and the dimension m made the number of correlations even. From a first glance, each deviation $|b| = 0.1$ was not large, but the number of deviations $\frac{1}{4}m(m-1)$ was large, hence $\sum|b_{ij}|$ was large too. For example, $\sum|b_{ij}| = 3.3$ if $m = 12$. On the other hand, we had not adjusted a_i and b_{ij} to reduce the level of deviation as in the case of quasi-decomposable matrix. In one word, our approximation (2.2) had a good efficiency.

Table 5.4 Half-difference frequency of approximation (2.2)

dimension m	4	5	8	9	12
correlation a	.30	.50	.60	.60	.20
deviation b	.10	-.10	.10	-.10	-.10
$\Sigma\|b_{ij}\|$	.30	0.50	1.40	1.80	3.30
freq $\in$ [0, .000005)	16	9	2	3	2
[.000005, .000010)	15	9	6	1	1
[.000010, .000050)	31	22	9	6	12
[.000050, .000100)	21	27	10	11	8
[.000100, .000250)	17	22	28	34	22
[.000250, .000500)		11	30	36	47
[.000500, .000750)			12	7	5
[.000750, .001000)			3	2	3
cumulative freq.	100	100	100	100	100

References

[1] S.S. Gupta, Probability integrals of multivariate normal and multivariate t, *Annals of Mathematical Statistics*, **34**:3 (1963), 792-828.

[2] Y.L. Tong, The Multivariate Normal Distribution, Springer-Verlag, New York, 1990.

[3] C.M. Zhang, Z.Q. Yang, Evaluation of multivariate normal probabilities, Proceedings of the Asian Conference on Statistical Computing, Beijing, October, 1993.

[4] P. Moran, The Monte-Carlo evaluation of orthant probabilities for multivariate normal distributions, *Australian J. of Stat.*, **26** (1984), 39-44.

[5] K.T. Fang, Y. Wang, Number-Theoretic Methods in Statistics, Chapman & Hall, London, 1994.

[6] Z. Drezner, Approximations to the multivariate normal integral, *Commun. Statist. Simula.* **19**:2 (1990), 527-534.

[7] A.H. Stroud, D.H. Secrest, Gaussian Quadrature Formulas, Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

[8] N.M. Steen, G.D. Byrne, E.M. Gelbard, Gaussian quadratures for the integrals, *Mathematics of Computation*, **23**:2 (1969), 661-671.

[9] Z.Q. Yang, Comparison of two gaussian quadratures in multivariate normal integrals, *Commun. Statist. Simula.*, to appear.

[10] Z.Q. Yang, On some computation methods of the nonlinear L_1 norm regression, *Chinese J. of Num. Math. and App.*, **16**:3 (1994), 31-38.

[11] Z.Q. Yang, An algorithm for nonlinear L_1 curve-fitting based on the smooth approximation, *Computational Statistics & Data Analysis*, **19**:1 (1995), 45-52.

[12] G.P. Steck, A table for computing trivariate normal probability, *Ann. Math. Statist.*, **29** (1958), 780-800.