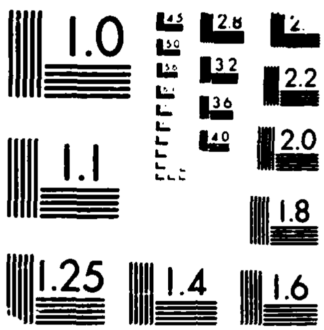
AD-A193 904 COMPUTING PARTIAL CORRELATIONS FROM THE DATA MATRIX(U) 1/1
YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE
J DELOSME ET AL. JUN 87 YALEU/DCS/RR-541
UNCLASSIFIED N00014-86-K-0310 F/G 12/3 NL



MICROCOPY RESOLUTION TEST CHART

1

Computing Partial Correlations from the Data Matrix

Jean-Marc Delosme[1], Ilse C.F. Ipsen[2]

Research Report YALEU/DCS/RR-541
June 1987

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

87 9 9 281

Abstract. The usual way of computing partial correlations is based on the formation of the covariance matrix, that amounts to squaring the data matrix, thus inviting a potential loss of numerical accuracy. This paper recommends the determination of partial correlations from the data matrix: the QR decomposition of the data matrix is computed and plane rotations are applied to the resulting upper triangular matrix, which is the Cholesky factor of the covariance matrix. ~~We show~~ that if rotations are applied to the triangular matrix so as to leave the number of its zero entries invariant, the sines of the rotation angles are partial correlations. Different ways of organizing the computations are presented for extracting any set of partial correlations.

Computing Partial Correlations from the Data Matrix

Jean-Marc Delosme[1], Ilse C.F. Ipsen[2]

Research Report YALEU/DCS/RR-541
June 1987

[1] Department of Electrical Engineering, Yale University
[2] Department of Computer Science, Yale University

The work presented in this paper was supported by the National Science Foundation under contract ECS-8314750, by the Office of Naval Research under contracts N000014-86-K-0310 and N00014-85-K-0461, and by the Army Research Office under contract DAAL03-86-K-0158.

1. Introduction

Classical texts [1] on multivariate statistics suggest the computation of partial correlations by first forming the *empirical covariance matrix* $\frac{1}{m-1}(A-\bar{A})^T(A-\bar{A})$, where A is the $m \times n$ *data matrix*, whose ith column is associated with random variable A_i, and $\bar{A} = \frac{1}{m}ee^TA$ is the *empirical mean matrix* (e is the $m \times 1$ vector of ones). Substantial loss of numerical accuracy is incurred by squaring the data matrix thus resulting in errors in the computed partial correlations (one can easily construct examples where a full-rank matrix $A-\bar{A}$ leads to a numerically indefinite covariance matrix). This loss of accuracy is inherent in the use of the covariance matrix and independent of the formulas and recursions employed to compute the partial correlations.

These shortcomings may be avoided with a method working directly on the data matrix and, in addition, employing orthogonal transformations. Our approach consists of two steps: after the QR decomposition of the matrix $A-\bar{A}$ the resulting upper triangular matrix U is transformed to lower triangular form L via plane rotations. The rotations are executed in a specific order which exploits the zero structure of the upper triangular matrix, and the values of their sines constitute the partial correlations between variables A_i and A_j, for $i < j$, holding variables $A_{i+1} \ldots A_{j-1}$ fixed.

By effecting the rotations on only a submatrix of U, the partial correlations between A_i and A_j where variables $A_1 \ldots A_{i-1}$ and $A_{i+1} \ldots A_{j-1}$ are held fixed are efficiently computed, without having to reorder the columns of either the data matrix A or the upper triangular matrix U. In general, we will present ways of organizing the computations so as to determine any set of partial correlations while keeping arbitrary collections of variables fixed.

2. Partial Correlation Coefficients

The column vector of m observed values a_{ki}, $1 \le k \le m$, of a real *random variable* A_i is denoted by

$$a_i \equiv \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}.$$

The centered, or zero-mean, data vector is $\alpha_i = a_i - \bar{a}_i$, where the barred quantity denotes the mean vector

$$\bar{a}_i \equiv \frac{1}{m} e(e^T a_i),$$

and e is the column vector of m ones.

The *empirical correlation coefficient* ρ_{ij} of two random variables A_i and A_j is defined as the cosine of the angle θ_{ij} between the centered data vectors α_i and α_j:

$$\rho_{ij} \equiv \cos\theta_{ij} = \frac{\alpha_i^T\alpha_j}{\sqrt{\alpha_i^T\alpha_i}\sqrt{\alpha_j^T\alpha_j}}.$$

The correlation between two variables A_i and A_j can arise, in part, from the fact that both A_i and A_j show a correlation with a third variable A_k. The 'partial correlation between A_i and A_j given A_k' then represents the correlation between A_i and A_j after the dependence on A_k has been removed. Formally, the *empirical partial correlation coefficient* ρ_{ij}^k between variables A_i and A_j *given* (*conditioned with respect to*) variable A_k is defined to be the cosine of the angle θ_{ij}^k between α_i^k and α_j^k where

$$\alpha_i^k \equiv \alpha_i - (\alpha_i^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}\alpha_k, \quad \alpha_j^k \equiv \alpha_j - (\alpha_j^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}\alpha_k,$$

are the respective projections of α_i and α_j onto the subspace orthogonal to the *vector* α_k (note that superscripts here denote conditioning rather than powers). Substituting this into the formula for the cosine

$$\rho_{ij}^k = \frac{(\alpha_i^k)^T\alpha_j^k}{\sqrt{(\alpha_i^k)^T\alpha_i^k}\sqrt{(\alpha_j^k)^T\alpha_j^k}}$$

yields

$$\rho_{ij}^k = \frac{\alpha_i^T\alpha_j - (\alpha_i^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}(\alpha_k^T\alpha_j)}{\sqrt{\alpha_i^T\alpha_i - (\alpha_i^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}(\alpha_k^T\alpha_i)}\sqrt{\alpha_j^T\alpha_j - (\alpha_j^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}(\alpha_k^T\alpha_j)}}.$$

Note that all quantities in the expression for ρ_{ij}^k are of the form $\alpha_i^T\alpha_j$. This means that in the general case of n variables A_i with m observed data values each, the partial correlations ρ_{ij}^k can be computed from the elements of the $n \times n$ *empirical covariance matrix*

$$B = (b_{ij}) \equiv (A - \bar{A})^T(A - \bar{A}),$$

where $A - \bar{A}$ is the centered $m \times n$ data matrix, and the data matrix A and its mean matrix $\bar{A}$ are defined by

$$A = [\,a_1 \quad \dots \quad a_n\,], \quad \bar{A} = \frac{1}{m}e(e^TA)$$

(the empirical covariance matrix is usually defined as $(m-1)^{-1}B$; since partial correlations are normalized quantities, independent of the scaling by $(m-1)^{-1}$, we shall use here the more convenient unscaled expression). Denoting by

$$b_{ij}^k \equiv b_{ij} - b_{ik}b_{kk}^{-1}b_{kj} = \left(\alpha_i^T\alpha_j - (\alpha_i^T\alpha_k)(\alpha_k^T\alpha_k)^{-1}(\alpha_k^T\alpha_j)\right) \equiv (\alpha_i^k)^T\alpha_j^k$$

the elements of the Schur complement in B with respect to b_{kk} [2, 3] one has

$$\rho_{ij}^{k} = \frac{b_{ij}^{k}}{\sqrt{b_{ii}^{k}}\sqrt{b_{jj}^{k}}}.$$

In general, the conditioning may occur with respect to more than one variable, for instance, with respect to A_l, A_k and A_m or with respect to a sequence $A_k \ldots A_{k+l}$. In that case the involved vectors α_i and α_j are projected onto a subspace orthogonal to the subspace spanned by α_k, α_l and α_m or by $\alpha_k \ldots \alpha_{k+l}$, respectively. Denoting by $b_{ij}^{k,l,m}$ and $b_{ij}^{k:k+l}$ the elements of the respective Schur complements of

$$\begin{bmatrix} b_{kk} & b_{kl} & b_{km} \\ b_{lk} & b_{ll} & b_{lm} \\ b_{mk} & b_{ml} & b_{mm} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b_{kk} & \ldots & b_{k,k+l} \\ \vdots & & \vdots \\ b_{k+l,k} & \ldots & b_{k+l,k+l} \end{bmatrix}$$

in B, the partial correlation between A_i and A_j given A_k, A_l and A_m is

$$\rho_{ij}^{k,l,m} = \frac{b_{ij}^{k,l,m}}{\sqrt{b_{ii}^{k,l,m}}\sqrt{b_{jj}^{k,l,m}}}$$

and the partial correlation between A_i and A_j given $A_k, A_{k+1}, \ldots, A_{k+l}$ is

$$\rho_{ij}^{k:k+l} = \frac{b_{ij}^{k:k+l}}{\sqrt{b_{ii}^{k:k+l}}\sqrt{b_{jj}^{k:k+l}}}.$$

Our notation automatically incorporates the so-called quotient property for Schur complements [2], which essentially states that the effect of conditioning with respect to variables belonging to a set S can be accomplished by first conditioning with respect to variables that belong to a subset of S_1 of S followed by conditioning with respect to the remaining variables in $S - S_1$, the complement of S_1 in S. The quotient property for Schur complements yields readily recursive formulas for the computation of $\rho_{ij}^{k,l,m}$ or $\rho_{ij}^{k:k+l}$ and these formulas are the ones generally used to compute the partial correlation coefficients [1, 3]. An inconvenience with such formulas, that rely on the computation of Schur complements in the covariance matrix, is that construction of the covariance matrix itself implies squaring up the data, $A - \bar{A}$, and thus a doubling of the dynamic range and potential loss of accuracy as the subsequent example shows.

Example. Suppose that the following zero-mean data matrix is given,

$$A - \bar{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & \epsilon \\ -1 & 1 & -\epsilon \\ \epsilon & \epsilon & 1 \\ -\epsilon & -\epsilon & -1 \end{bmatrix},$$

where ϵ is non-zero, and that the partial correlation

$$\rho_{13}^2 = \frac{b_{13}^2}{\sqrt{b_{11}^2}\sqrt{b_{33}^2}}$$

is to be determined. The corresponding covariance matrix is

$$B = \begin{bmatrix} 1+\epsilon^2 & -1+\epsilon^2 & 2\epsilon \\ -1+\epsilon^2 & 1+\epsilon^2 & 0 \\ 2\epsilon & 0 & 1+\epsilon^2 \end{bmatrix},$$

and in exact arithmetic, one has

$$b_{13}^2 = 2\epsilon, \quad b_{11}^2 = 1+\epsilon^2 - \frac{(-1+\epsilon^2)^2}{1+\epsilon^2}, \quad b_{33}^2 = 1+\epsilon^2$$

so that

$$\rho_{13}^2 = \mathrm{sign}(\epsilon) \equiv \begin{cases} 1 & \epsilon > 0 \\ -1 & \epsilon < 0 \end{cases}.$$

However, in finite precision floating point arithmetic and with ϵ chosen to be sufficiently small (e.g. ϵ is the largest number so that the *computed* $\mathrm{fl}(1 \pm 4\epsilon^2) = 1$), the computed quantities turn out to be

$$\mathrm{fl}(B) = \begin{bmatrix} 1 & -1 & 2\epsilon \\ -1 & 1 & 0 \\ 2\epsilon & 0 & 1 \end{bmatrix}$$

and

$$b_{13}^2 = 2\epsilon, \quad b_{11}^2 = 0, \quad b_{33}^2 = 1$$

so that ρ_{13}^2 is not a finite number.

The next section introduces a numerical method that achieves much higher accuracy by working directly on the data matrix $A - \bar{A}$.

In order to avoid squaring the data matrix one may try to work with the Cholesky factor of B. Indeed the $n \times n$ upper triangular Cholesky factor U can be obtained without squaring from the QR factorization of the scaled centered data matrix since, if $A - \bar{A}$ is decomposed into the product QU where Q has orthogonal columns,

$$B = (A-\bar{A})^T(A-\bar{A}) = (A-\bar{A})^T QQ^T (A-\bar{A}) = U^T U.$$

It is known that the elements of the Cholesky factor can be represented in terms of elements of Schur complements with respect to the leading principal submatrices of B:

Lemma 2.1. *The non-zero elements of the Cholesky factor U of B are of the form*

$$u_{ij} = (b_{ii}^{1:i-1})^{-1/2} b_{ij}^{1:i-1}, \qquad j \geq i.$$

Proof. Let $\tilde{U}$ be an upper triangular matrix with elements $\tilde{u}_{ij} = (b_{ii}^{1:i-1})^{-1/2} b_{ij}^{1:i-1}$, $j \geq i$. The (i,j)th element, $j \geq i$, of the symmetric matrix $\tilde{U}^T\tilde{U}$ is

$$\begin{aligned}
\sum_{k=1}^{i} \tilde{u}_{ki}\tilde{u}_{kj} &= \sum_{k=1}^{i} (b_{kk}^{1:k-1})^{-1/2} b_{ki}^{1:k-1} (b_{kk}^{1:k-1})^{-1/2} b_{kj}^{1:k-1} = \sum_{k=1}^{i} b_{ik}^{1:k-1} (b_{kk}^{1:k-1})^{-1} b_{kj}^{1:k-1} \\
&= b_{ij} - \left(b_{ij} - b_{i1}(b_{11})^{-1} b_{1j}\right) + \sum_{k=2}^{i} b_{ik}^{1:k-1} (b_{kk}^{1:k-1})^{-1} b_{kj}^{1:k-1} = b_{ij} - b_{ij}^{1} + \sum_{k=2}^{i} b_{ik}^{1:k-1} (b_{kk}^{1:k-1})^{-1} b_{kj}^{1:k-1} \\
&= b_{ij} - \left(b_{ij}^{1} - b_{i2}^{1}(b_{22}^{1})^{-1} b_{2j}^{1}\right) + \sum_{k=3}^{i} b_{ik}^{1:k-1} (b_{kk}^{1:k-1})^{-1} b_{kj}^{1:k-1} = b_{ij} - b_{ij}^{1:2} + \sum_{k=3}^{i} b_{ik}^{1:k-1} (b_{kk}^{1:k-1})^{-1} b_{kj}^{1:k-1} \\
&= \ \ldots = b_{ij} - \left(b_{ij}^{1:i-1} - b_{ii}^{1:i-1} (b_{ii}^{1:i-1})^{-1} b_{ij}^{1:i-1}\right) = b_{ij},
\end{aligned}$$

where the telescoping of the sum is achieved by making use of the quotient property of Schur complements. Hence $\tilde{U}^T\tilde{U} = B$, and the uniqueness of the Cholesky factor implies $\tilde{U} = U$.

∎

The partial correlation between A_i and A_j given the intermediate variables $A_1, \ldots, A_{i-1}$ can be expressed as

$$\rho_{ij}^{1:i-1} = (b_{ii}^{1:i-1})^{-1/2} b_{ij}^{1:i-1} (b_{jj}^{1:i-1})^{-1/2} = u_{ij}(b_{jj}^{1:i-1})^{-1/2}.$$

Thus it seems that the partial correlations may be computed as simple functions of the elements of the Cholesky factor. Yet unfortunately the desired quantity $(b_{jj}^{1:i-1})^{-1/2}$ differs from $u_{jj}^{-1} = (b_{jj}^{1:j-1})^{-1/2}$. Moreover, it is hard to see how to determine the quantities $(b_{jj}^{1:i-1})^{-1/2}$ without the use of squaring operations.

The next section introduces a numerical method that gets around this difficulty by applying plane rotations to the columns of the Cholesky factor.

3. New Algorithm

From the previous section it is clear that one must think of more subtle means to employ the Cholesky factor: our method determines partial correlations as cosines evaluated through inner products but instead as sines of rotations that zero out components of certain column vectors. The key idea for the new algorithm is based on the fact that the data vectors are initially represented by a triangular matrix, the Cholesky factor; and that the partial correlations may be computed by

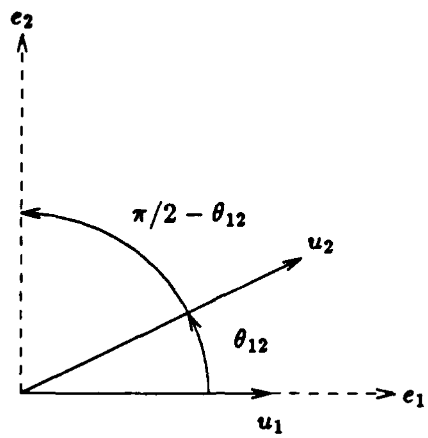


Figure 1: Angles in the 2×2 Example.

applying plane rotations in a particular order to the columns of the Cholesky factor. To see that consider a simple 2×2 example.

Let

$$U = \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

be the upper triangular factor (with positive diagonal elements) in the QR decomposition of a $m \times 2$ matrix $A - \bar{A}$. The (partial) correlation ρ_{12} between A_1 and A_2 is the cosine of the angle θ_{12} between the two columns of U. Because of the triangular structure of U its first column, $[\,u_{11} \quad 0\,]^T$, is a positive multiple of the first canonical vector $e_1 = [\,1 \quad 0\,]^T$ while its second column is a linear combination of e_1 and the second canonical vector $e_2 = [\,0 \quad 1\,]^T$.

The columns of the matrix U may be rotated in such a way that the second column becomes a positive multiple of e_2 thereby turning the first into a linear combination of e_1 and e_2:

$$L \equiv \Theta U = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}.$$

Suppose the angle between e_1 and e_2, denoted by $\angle(e_1, e_2)$, is $+\pi/2$ then the angle between the two columns can be defined as $\theta_{12} \equiv \angle(u_1, u_2)$. The fact that the first column is a positive multiple of e_1 implies $\angle(e_1, u_2) = \theta_{12}$. To turn the second column into a positive multiple of e_2 requires that all columns of U be rotated by the angle

$$\angle(u_2, e_2) = \angle(e_1, e_2) - \angle(e_1, u_2) = \pi/2 - \theta_{12},$$

see Figure 1. Hence the angle between the two columns of U is preserved under the rotation, and the angle of such a rotation

$$\Theta = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

completes θ_{12} to a right angle: $c = \cos(\pi/2 - \theta_{12})$ and

$$s = \sin(\pi/2 - \theta_{12}) = \cos\theta_{12} = \rho_{12}.$$

Consequently, the desired (partial) correlation is the sine of the rotation Θ.

The above suggests that, in general, certain partial correlations may be computed from the plane rotations that transform the upper Cholesky factor to the lower Cholesky factor. The triangular zero-structure of U makes it possible to rotate columns in a manner illustrated above and determine a partial correlation from the sine of a rotation.

brief look at the 3×3 case

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

illustrates the above.

At first, because the second column has only one more non-zero element than the first, the columns of U can be rotated in the (e_1, e_2)-plane so as to make the second column co-linear with e_2,

$$\begin{bmatrix} c_{12} & -s_{12} & 0 \\ s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} = \begin{bmatrix} * & 0 & * \\ * & * & * \\ 0 & 0 & u_{33} \end{bmatrix},$$

and $\rho_{12} = s_{12}$. Here, 'co-linear' is used to mean 'a positive multiple of' and $*$ denotes terms that are non-zero in general.

Next, to achieve conditioning of A_1 and A_3 with respect to A_2, the first and third columns are projected onto the subspace orthogonal to the second column. Due to the triangular structure of U and the effect of the previous rotation the second column is co-linear to e_2, and the subspace orthogonal to it is just the plane (e_1, e_3). The partial correlation ρ_{13}^2 can then be determined from that rotation that makes co-linear with e_3 the projection of the third column onto (e_1, e_3). Since this rotation takes place in a subspace orthogonal to the second column it does not affect the second column, and the zero element introduced by the previous rotation is preserved:

$$\begin{bmatrix} c_{13}^2 & 0 & -s_{13}^2 \\ 0 & 1 & 0 \\ s_{13}^2 & 0 & c_{13}^2 \end{bmatrix} \begin{bmatrix} * & 0 & * \\ * & * & * \\ 0 & 0 & u_{33} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ * & * & * \\ * & 0 & * \end{bmatrix}$$

and $\rho_{13}^2 = s_{13}^2$. Note that another non-zero element is introduced in the first column.

Again, because of the triangular structure of U and the effect of the second rotation the zero-structure of the second and third columns is the same save for one element, the second column is co-linear with e_2 while the third is a linear combination of e_2 and e_3. Thus the columns of the matrix can be rotated to yield ρ_{23} by applying a rotation that makes the whole third column co-linear with e_3, and turns the second column into a linear combination of e_2 and e_3.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & -s_{23} \\ 0 & s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} l_{11} & 0 & 0 \\ * & * & * \\ * & 0 & * \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$$

and $\rho_{23} = s_{23}$.

Theorem 3.1. *If the elements in the Cholesky factor U of the covariance matrix B are eliminated in the order*

$$\begin{bmatrix} * & 1 & 2 & \dots & n-1 \\ & * & n & & 2n-3 \\ & & \ddots & & \vdots \\ & & & * & n(n-1)/2 \\ & & & & * \end{bmatrix},$$

that is, proceeding row after row from top to bottom, and within each row from left to right, then the sine of the rotation that eliminates element (i,j), $j > i$, is equal to the partial correlation $\rho_{ij}^{i+1:j-1}$.

Proof. The proof proceeds by induction.

The induction basis comprises the computation of partial correlations between A_1 and all other variables. To start with, the matrix U is of the form

$$\begin{bmatrix} * & * & * & * & \dots & * \\ & * & * & * & \dots & * \\ & & * & * & \dots & * \\ & & & * & \dots & * \\ & & & & \ddots & \vdots \\ & & & & & * \end{bmatrix}.$$

From the 2×2 case one can see that elimination of element $(1,2)$ in the upper triangular matrix U by a rotation in plane (e_1, e_2) provides ρ_{12}. The second column of the resulting matrix becomes co-linear to e_2 while the first column becomes a linear combination of e_1 and e_2. Hence, there is a new non-zero element in the first column and a zero has been introduced in the first row:

$$\begin{bmatrix} * & & * & * & \cdots & * \\ * & * & * & * & \cdots & * \\ & & * & * & \cdots & * \\ & & & * & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & * \end{bmatrix}.$$

The 3×3 case showed that the correlation ρ_{13}^{2} between A_1 and A_3 given A_2 could be computed by rotating the first and third column and thereby introducing a non-zero element in position $(3, 1)$ and a zero in position $(1, 3)$:

$$\begin{bmatrix} * & & & * & \cdots & * \\ * & * & * & * & \cdots & * \\ * & & * & * & \cdots & * \\ & & & * & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & * \end{bmatrix}.$$

Continuing this argument, the partial correlation $\rho_{1j}^{2:j-1}$ between A_1 and A_j, given $A_2, \ldots, A_{j-1}$ is computed by peforming a rotation in plane (e_1, e_j) thereby creating a zero element in position $(1, j)$. Thus, once all correlations involving A_1 have been computed the first column of the matrix has totally filled in, and the first row is zero except for the first element:

$$\begin{bmatrix} * & & & & & \\ * & * & * & * & \cdots & * \\ * & & * & * & \cdots & * \\ * & & & * & \cdots & * \\ \vdots & & & & \ddots & \vdots \\ * & & & & & * \end{bmatrix}.$$

Assume that the partial correlations $\rho_{kl}^{k+1:l-1}$ have already been computed for $k \leq i$ and

$k < l < j$. The corresponding matrix is of the form

$$\left[\begin{array}{c|ccccc|c} L_0 & & & & & & \\ \hline * & v_{ii} & & & & v_{i,j} & * \\ * & v_{i+1,i} & v_{i+1,i+1} & \cdots & v_{i+1,j-1} & v_{i+1,j} & * \\ * & \vdots & & \ddots & \vdots & \vdots & * \\ * & v_{j-1,i} & & & v_{j-1,j-1} & v_{j-1,j} & * \\ * & & & & & v_{j,j} & * \\ \hline * & & & & & & U_0 \end{array}\right],$$

where L_0 is lower triangular and U_0 is upper triangular.

By induction hypothesis the entire lower triangular part of the leading $i-1$ columns is non-zero, and the ith column has $j-2$ non-zeros in its lower triangular part due to the computation of $\rho_{i,i+1}, \ldots, \rho_{i,j-1}^{i+1:j-2}$. In order to compute the next correlation $\rho_{ij}^{i+1:j-1}$ the corresponding columns $v_i, \ldots, v_j$ of the current matrix must be projected onto a subspace orthogonal to the subspace spanned by $A_{i+1}, \ldots, A_{j-1}$. Due to the initial 'nesting' of the column subspaces (i.e. the original upper triangular structure of U) the trailing components $j, \ldots, n$ of $v_{i+1}, \ldots, v_{j-1}$ are zero; and due to the rotations performed in order to retrieve previous partial correlations (i.e. the appearing lower triangular structure of L) the leading components $1, \ldots, i$ of $v_{i+1}, \ldots, v_{j-1}$ are zero. Hence the subspace spanned by $A_{i+1}, \ldots, A_{j-1}$ is the space spanned by $e_{i+1}, \ldots, e_{j-1}$, and the space orthogonal to it is the space spanned by $e_1, \ldots, e_i, e_j, \ldots, e_n$. Similarly, components $1, \ldots, i-1, j, \ldots, n$ of v_i and components $1, \ldots, i-1, j+1, \ldots, n$ of v_j are zero; and the projections of v_i and v_j onto $e_1, \ldots, e_i, e_j, \ldots, e_n$ are respectively co-linear to e_i and a linear combination of e_i and e_j. Thus, $\rho_{ij}^{i+1:j-1}$ is obtained by applying the rotation in plane (e_i, e_j) that makes the projection of v_j co-linear with e_i; $\rho_{ij}^{i+1:j-1}$ is the sine of that rotation. After the rotation the matrix has the form

$$\left[\begin{array}{c|ccccc|c} L_0 & & & & & & \\ \hline * & v'_{ii} & & & & & * \\ * & v_{i+1,i} & v_{i+1,i+1} & \cdots & v_{i+1,j-1} & v_{i+1,j} & * \\ * & \vdots & & \ddots & \vdots & \vdots & * \\ * & v_{j-1,i} & & & v_{j-1,j-1} & v_{j-1,j} & * \\ * & v'_{j,i} & & & & v'_{j,j} & * \\ \hline * & & & & & & U_0 \end{array}\right].$$

Remark. If the matrix L is not needed about half of the arithmetic operations can be saved by applying the rotations merely to the trailing principal submatrix of interest.

Returning to the example of the previous section it becomes clear that the new method can avoid the loss of accuracy associated with the explicit formation of the covariance matrix.

Example. Performing a QR decomposition of the matrix $A - \bar{A}$ yields 4×3 matrix Q with orthonormal columns and a 3×3 upper triangular factor

$$U = \frac{1}{\sqrt{1+\epsilon^2}} \begin{bmatrix} 1+\epsilon^2 & -1+\epsilon^2 & 2\epsilon \\ 0 & 2|\epsilon| & \mathrm{sign}(\epsilon)(1-\epsilon^2) \\ 0 & 0 & 0 \end{bmatrix}$$

in exact arithmetic and

$$\mathrm{fl}(U) = \begin{bmatrix} 1 & -1 & 2\epsilon \\ 0 & 2|\epsilon| & \mathrm{sign}(\epsilon) \\ 0 & 0 & 0 \end{bmatrix}$$

in finite precison arithmetic with the same choice of ϵ as before. The first rotation with $\mathrm{fl}(s_{12}) = -1$ and $\mathrm{fl}(c_{12}) = 2|\epsilon|$ yields

$$\mathrm{fl}(U') = \begin{bmatrix} 2|\epsilon| & 0 & \mathrm{sign}(\epsilon) \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the second rotation gives $c_{13}^2 = 0$ and $s_{13}^2 = \mathrm{sign}(\epsilon)$ so that $\mathrm{fl}(\rho_{13}^2) = \mathrm{sign}(\epsilon)$ indicates a linear dependence between the three columns of $A - \bar{A}$ as in the true computation.

4. Computation of Arbitrary Partial Correlations

Subject to a certain inital ordering of the random variables $A_1, \ldots, A_n$ our algorithm computes the partial correlations $\rho_{ij}^{i+1:j-1}$ between A_i and A_j given $A_{i+1}, \ldots, A_{j-1}$ by completely reducing the upper triangular matrix U to a lower triangular matrix L.

Other partial correlations may be computed by performing only a partial reduction. For instance, consider the following 6×6 example

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} \\ & u_{22} & u_{23} & u_{24} & u_{25} & u_{26} \\ & & u_{33} & u_{34} & u_{35} & u_{36} \\ & & & u_{44} & u_{45} & u_{46} \\ & & & & u_{55} & u_{56} \\ & & & & & u_{66} \end{bmatrix}.$$

The leading three columns of U span the subspace of A_1, A_2 and A_3, and this is equal to the space spanned by the first three canonical vectors e_1, e_2 and e_3 due to the triangular structure of U. The space orthogonal to it is the one spanned by e_4, e_5, e_6 and is, because of the triangular structure, equal to that of columns 4 through 6 of U with components 1 to 3 set to zero. This means that the correlation $\rho_{45}^{1:3}$ between A_4 and A_5, given A_1, A_2 and A_3, can be computed by a rotation of U in plane (e_4, e_5). The resulting matrix has a new zero in column five and a fill-in in column four:

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} \\ & u_{22} & u_{23} & u_{24} & u_{25} & u_{26} \\ & & u_{33} & u_{34} & u_{35} & u_{36} \\ & & & * & & * \\ & & & * & * & * \\ & & & & & u_{66} \end{bmatrix}.$$

The next correlation that can be computed is $\rho_{46}^{1:3,5}$ with a rotation in plane (e_4, e_6), the subspace orthogonal to A_1, A_2, A_3 and A_5:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} \\ & u_{22} & u_{23} & u_{24} & u_{25} & u_{26} \\ & & u_{33} & u_{34} & u_{35} & u_{36} \\ & & & l'_{11} & & \\ & & & * & * & * \\ & & & * & & * \end{bmatrix}.$$

The last correlation $\rho_{56}^{1:3}$ is determined by completing the transformation of the 3×3 trailing principal submatrix to lower triangular form:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} \\ & u_{22} & u_{23} & u_{24} & u_{25} & u_{26} \\ & & u_{33} & u_{34} & u_{35} & u_{36} \\ & & & l'_{11} & & \\ & & & l'_{21} & l'_{22} & \\ & & & l'_{31} & l'_{32} & l'_{33} \end{bmatrix}.$$

In general, the correlation $\rho_{ij}^{1:\alpha,i+1:j-1}$ for $i > \alpha$ and $i < j \leq n$ can be determined be preserving the leading α rows and columns of U and transforming the trailing principal submatrix of order $n-\alpha$

to lower triangular form L_α by appropriate plane rotations:

$$
\begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & \vdots \\ & & * & \cdots & \cdots & * \\ & & & * & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & * \end{bmatrix} \to \begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & \vdots \\ & & * & \cdots & \cdots & * \\ & & & * & & \\ & & & \vdots & \ddots & \\ & & & * & \cdots & * \end{bmatrix}.
$$

Similarly, the computation of $\rho_{ij}^{i+1:j-1,n-\beta+1:n}$ for $j < n-\beta+1$ and $1 \leq i < j$ is accomplished by transforming U to lower triangular form L (or obtaining directly a QL factorization of $A - \bar{A}$) and then transforming the leading $\beta \times \beta$ principal submatrix of L to upper triangular form U_β:

$$
\begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & \vdots \\ & & * & \cdots & \cdots & * \\ & & & * & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & * \end{bmatrix} \to \begin{bmatrix} * & & & & & \\ \vdots & \ddots & & & & \\ * & \cdots & * & & & \\ * & \cdots & \cdots & * & & \\ \vdots & & & & \ddots & \\ * & \cdots & \cdots & \cdots & \cdots & * \end{bmatrix} \to \begin{bmatrix} * & \cdots & * & & & \\ & \ddots & \vdots & & & \\ & & * & & & \\ * & \cdots & \cdots & * & & \\ \vdots & & & & \ddots & \\ * & \cdots & \cdots & \cdots & \cdots & * \end{bmatrix}.
$$

Combining the two above strategies makes it possible to determine $\rho_{ij}^{1:\alpha,i+1:j-1,n-\beta+1:n}$ for $\alpha < i < j < n-\beta+1$ by transforming the trailing $(n-\alpha) \times (n-\alpha)$ principal submatrix of U to block lower triangular form L_α (see the sketch below) and subsequently transforming the leading $(\beta-\alpha) \times (\beta-\alpha)$ triangular submatrix of L_α to upper triangular form $U_{\alpha,\beta}$:

$$
\begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & & & & \vdots \\ & & * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & & & * & \cdots & \cdots & \cdots & \cdots & * \\ & & & & \ddots & & & & \vdots \\ & & & & & * & \cdots & \cdots & * \\ & & & & & & * & \cdots & * \\ & & & & & & & \ddots & \vdots \\ & & & & & & & & * \end{bmatrix} \to \begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & & & & \vdots \\ & & * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & & & * & & & & & \\ & & & \vdots & \ddots & & & & \\ & & & * & \cdots & * & & & \\ & & & * & \cdots & \cdots & \cdots & \cdots & * \\ & & & \vdots & & & & & \vdots \\ & & & * & \cdots & \cdots & \cdots & \cdots & * \end{bmatrix}
$$

$$
\to \begin{bmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & & & & \vdots \\ & & * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & & & * & \cdots & * & & & \\ & & & & \ddots & \vdots & & & \\ & & & & & * & & & \\ & & & * & \cdots & \cdots & \cdots & \cdots & * \\ & & & \vdots & & & & & \vdots \\ & & & * & \cdots & \cdots & \cdots & \cdots & * \end{bmatrix}
$$

If it is known in advance which partial correlations are to be determined then the columns of the $m \times n$ data matrix may be ordered so as to minimize the number of arithmetic operations succeeding the computation of the Cholesky factor.

For instance, a lower bound on the number of arithmetic operations in the computation of ρ_{ij}^S, where S is a subset of $k \geq 0$ numbers in $1 \ldots n$ not containing i and j, is $O(n-k)$ since our method requires at least one rotation to compute a partial correlation and the dimension of the space involved is $n-k$. This lower bound is attained by ordering the columns so that the set S represents the leading k columns of the data matrix followed by columns A_i and A_j. The correlation ρ_{ij}^k can then be determined by one rotation in the plane (e_{k+1}, e_{k+2}) that, due to the triangular structure of the Cholesky factor, involves $O(n-k)$ non-zero element pairs.

Not only the ordering of the columns is important but also the sequence in which particular correlations are computed. Consider the computation of a partial correlation between two variables A_i and A_j with successively more variables fixed: $\rho_{ij}^{S_1}, \ldots, \rho_{ij}^{S_k}$, where $S_1 \subset \ldots \subset S_k$ and $i, j \notin S_k$. It seems that the following order of rotations constitutes the simplest way of determining the above correlations. It is illustrated by means of a 5×5 example for the computation of ρ_{12}, ρ_{12}^3, $\rho_{12}^{3:4}$, and $\rho_{12}^{3:5}$. At first the columns of the data matrix are ordered so that i and j represent the first two columns followed by the columns of S_1, the columns of $S_2 - S_1$, the columns of $S_3 - S_2 - S_1$, etc. In the example this amounts to the 'natural' ordering $A_1 \ldots A_5$ of the variables. The first correlation ρ_{12} can now be computed with one rotation from the Cholesky factor U. To compute ρ_{12}^3 columns two and three of U are exchanged and a rotation in plane (e_2, e_3) results in the Cholesky factor U' corresponding to the data matrix with variables in the order A_1, A_3, A_2, A_4, A_5. Since A_3 is situated between A_1 and A_2 two rotations suffice for the computation of ρ_{12}^3. The effect of these steps on the matrix is depicted below:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} & u_{15} \\ & u_{22} & u_{23} & u_{24} & u_{25} \\ & & u_{33} & u_{34} & u_{35} \\ & & & u_{44} & u_{45} \\ & & & & u_{55} \end{bmatrix} \to \begin{bmatrix} u_{11} & u_{13} & u_{12} & u_{14} & u_{15} \\ & u_{23} & u_{22} & u_{24} & u_{25} \\ & u_{33} & & u_{34} & u_{35} \\ & & & u_{44} & u_{45} \\ & & & & u_{55} \end{bmatrix} \to \begin{bmatrix} u_{11} & u_{13} & u_{12} & u_{14} & u_{15} \\ & u'_{22} & u'_{23} & u'_{24} & u'_{25} \\ & & u'_{33} & u'_{34} & u'_{35} \\ & & & u_{44} & u_{45} \\ & & & & u_{55} \end{bmatrix}$$

$$
\rightarrow \begin{bmatrix} * & & * & * & * \\ * & * & * & * & * \\ & & u'_{33} & u'_{34} & u'_{35} \\ & & & u_{44} & u_{45} \\ & & & & u_{55} \end{bmatrix} \rightarrow \begin{bmatrix} * & & & * & * \\ * & * & * & * & * \\ * & & * & * & * \\ & & & u_{44} & u_{45} \\ & & & & u_{55} \end{bmatrix}.
$$

Similarly, exchanging columns three and four of U', performing a rotation in plane (e_3, e_4) to get the Cholesky factor U'' of the data matrix corresponding to the ordering A_1, A_3, A_4, A_2, A_5, and performing three more rotations on U'' results in the extraction of $\rho_{12}^{3:4}$. In general, if the sets S_l differ by more than one index, more columns of the Cholesky factor must be exchanged to ensure that all fixed variables are situated between A_i and A_j.

As for arbitrary sequences of partial correlations, the determination of the column ordering of the data matrix as well as the computation sequence of the partial correlations so as to minimize the number of arithmetic operations seems to be an NP-complete problem. The use of heuristics, such as the following greedy approach, might lead to acceptable operation counts: the random variables are ordered so that as many partial correlations as possible can be determined from the resulting Cholesky factor. Repeatedly, the columns of the Cholesky factor are then re-ordered according to the same strategy, the matrix returned to upper triangular form, and appropriate rotations performed until all correlations have been computed.

5. An Open Problem

Adding a row a^T to the $m \times n$ data matrix A results in a rank-two update to the Cholesky factor U of $A - \bar{A}$. Suppose the QR-factorization of $A - \bar{A}$, $A - \bar{A} = QU$ where Q is $m \times n$ with orthogonal columns is available. At first the rank-one $n \times n$ matrix $Q^T e(\frac{1}{m} e^T A - \frac{1}{m+1}(e^T A + a^T))$, which can be computed in $O(mn)$ operations, is added to U and then the row $a^T - \frac{1}{m+1}(e^T A + a^T)$ is appended as the $(n+1)$st row. The so augmented Cholesky factor can be reverted to upper triangular form U' by means of rotations in $O(n^2)$ operations. The quantities of interest, the partial correlations of the updated matrix, can be computed from U'. However, instead of starting all over from U', a more effective approach could be to use the augmented Cholesky factor as starting point in order to 'update' the partial correlations computed from U. A similar problem arises after the deletion of a row from the data matrix.

References

[1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis,* Wiley, 1958.

[2] Cottle, R.W., *Manifestations of the Schur Complement,* Linear Algebra and its Applications, 8 (1974), pp. 189–211.

[3] Ouellette, D.V., *Schur Complements and Statistics,* Linear Algebra and its Applications, 36 (1981), pp. 187–295.

END
DATE
FILMED
7-88
DTIC