Abstract. An $m \times m$ symmetric nonnegative definite matrix Σ has Cholesky factorization $\Sigma = U^T U$. By carrying out the factorization in a particular way for positive definite Σ, the Schur complements of all the leading principal submatrices of Σ are produced, as well as their Cholesky factors. It is shown how the same can be done for generalized Schur complements when Σ is singular. When Σ is the population covariance matrix of a multivariate random distribution, partial covariances and correlations can be defined in terms of the elements of such Schur complements. It follows that these can be produced efficiently and reliably from the Cholesky factorization.

When $n \times m$ A is given and $\Sigma = A^T A$, the Cholesky factor U may be found directly from the QR factorization $A = Q_1 U$, $Q_1^T Q_1 = I$, and this is preferable in many numerical computations. This QR factorization, or the modified Gram-Schmidt orthogonalization, produces projections of later columns of A onto spaces orthogonal to earlier columns. It is shown how the cosines of the angles between such projected vectors can be found using the elements of U. These cosines produced from A turn out to be the previously mentioned partial correlation coefficients produced from Σ, when $\Sigma = A^T A$. When A is obtained from observations of random variables, these are the *sample* correlation coefficients. It is shown how such correlation coefficients can be efficiently obtained when observations are added or deleted. This corresponds to altering all of A in a certain simple way, and adding or deleting rows.

The Cholesky Factorization, Schur Complements, Correlation Coefficients, Angles between Vectors, and the QR Factorization

J.-M. Delosme[1], I.C.F. Ipsen[2], C.C. Paige[3]

Research Report YALEU/DCS/RR-607
February 1988

[1] Department of Electrical Engineering, Yale University
[2] Department of Computer Science, Yale University
[3] Department of Computer Science, McGill University, Montreal, Canada

The work presented in this paper was supported by the National Science Foundation under contract ECS-8314750, by the Office of Naval Research under contract N000014-86-K-0310 and N00014-85-K-0461, by the Army Research Office under contract DAAL03-86-K-0158, and by the Natural Sciences and Engineering Research Council of Canada under grant A9236. Approved for public release: distribution is unlimited.

Notation

We will be dealing with vectors and matrices in their own right, as well as treating them as statistical objects. Because of the very different notation used in econometrics, statistics, and other areas which use matrix theory, we will use a standard notation of matrix theory and describe the convention we use within this for representing statistical objects. We describe these fully for possibly different audiences.

Matrix Theory Notation. Upper case greek and italic letters, like Σ and A, will denote matrices; lower case italics, like a_i and x, will denote vectors; except that i, j, k, m, n, p, q, r, s, t will represent integers or indices. Lower case greek letters will denote other scalars. Thus,

$$A(n \times m) = (a_1 \quad \dots \quad a_m) = (\alpha_{ij})$$

represents an $n \times m$ matrix A with columns $a_1, \dots, a_m$ and elements α_{ij}.

A ***principal submatrix*** of Σ is a square submatrix whose diagonal elements are also diagonal elements of Σ. A ***leading principal submatrix*** of Σ is a principal submatrix in the top left corner of Σ.

We will make frequent use of the following notions: Σ^T denotes the *transpose* of Σ; $\Re^n$ is the *real n-space* and $\|x\|_2 = \sqrt{x^T x}$ the two-norm of a vector x. The *range* or *image* or *column space* of a matrix A is denoted by $R(A)$, and its orthogonal complement by $R(A)^\perp$. $\Theta(a, b)$ is the *angle between two vectors* a and b of equal dimension.

Statistical Notation. This will always follow the matrix notation above. We will use the same notation for a random variable and an instance or observation of that variable. Thus x can represent a random vector or an observation of this random vector, the meaning will be clear from the context. $E(x)$ denotes the *expected value*, and $x \sim (a, \Sigma)$ denotes a vector x of random variables with mean a and variance-covariance matrix (covariance for short) Σ.

1. Introduction

For a given matrix

$$\Sigma = \begin{pmatrix} E & F \\ G & H \end{pmatrix} \tag{1.1}$$

with nonsingular E, the Schur complement of E in Σ, often written (Σ/E), is

$$S = (\Sigma/E) = H - GE^{-1}F. \tag{1.2}$$

There is an extensive literature on this, and it is a theoretical tool widely used by statisticians as can be seen from the bibliography prepared by Ouellette [12], see also [2, 3]. Perhaps the most common use by statisticians of the Schur complement occurs when Σ is a symmetric positive definite covariance matrix, and the Schur complement of a principal submatrix is required. In this case there is a simple relation between Schur complements and Cholesky factors. Suppose we partition the matrices

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = U^T U = \begin{pmatrix} U_{11}^T & 0 \\ U_{12}^T & U_{22}^T \end{pmatrix} \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix}, \tag{1.3}$$

where U is the upper triangular Cholesky factor of Σ. It follows that

$$\begin{aligned} S = (\Sigma/\Sigma_{11}) &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= U_{22}^T U_{22} + U_{12}^T U_{12} - U_{12}^T U_{11}(U_{11}^T U_{11})^{-1} U_{11}^T U_{12} \\ &= U_{22}^T U_{22}, \end{aligned} \tag{1.4}$$

so that the Cholesky factor of a symmetric positive definite Σ not only provides the Cholesky factors of all the leading principal submatrices $\Sigma_{11} = U_{11}^T U_{11}$ of Σ, but also the Cholesky factors of the Schur complements of all the leading principal submatrices $(\Sigma/\Sigma_{11}) = U_{22}^T U_{22}$ in Σ. By permuting rows, and permuting columns in the same way, any principal submatrix can be made a leading principal submatrix, and the above well known result suggests we might be able to use the Cholesky factor and ignore the Schur complement. This is not quite true, as we now point out.

When Σ is the covariance matrix of a multivariate distribution, the *population* partial correlation coefficients are defined in terms of the elements of Schur complements of principal submatrices of Σ, see [1] pp. 37,41, so these Schur complements are really needed. In §2 we show that if the Cholesky factorization is carried out in a particular order, then the Schur complements of all leading principal submatrices occur naturally as part of the factorization; that is, without having to form $U_{22}^T U_{22}$.

When E is singular or not square in (1.1) Marsaglia and Styan [11] define a generalized Schur complement

$$S = (\Sigma/E) = H - GE^{-}F, \qquad EE^{-}E = E. \tag{1.5}$$

When Σ is symmetric nonnegative definite and possibly singular, this generalized Schur complement is unique and symmetric nonnegative definite, and we show in §2 how it, and its Cholesky factor, occur naturally as part of a particular Cholesky factorization of Σ. In fact the form and the factorization is as simple as that for the positive definite case. When it is also realized that such algorithms for producing Cholesky factors are very efficient, and are numerically reliable when Σ is positive definite (see for example [8], p. 89), and can nearly always be made reliable when Σ is singular, see [9], and that results involving (generalized) Schur complements of principal submatrices are easily derived using Cholesky factors, the argument for thinking largely in terms of Cholesky factors becomes quite strong, especially for those interested in computations.

In §3 we discuss correlation coefficients, partial correlation coefficients, and conditional correlation coefficients, and give an efficient and reliable way of obtaining these during the Cholesky factorization of the given covariance matrix Σ. Readers not interested in such statistical objects need only note that the aim is to obtain the coefficients $\rho_{jk}^{(i)}$ in (3.5), and this is done from the elements $\sigma_{jk}^{(i)}$ of $\Sigma^{(i)} = U_{22}^T U_{22}$ in (1.4).

The angle $\Theta(a,b)$ between two nonzero vectors a, $b \in \Re^n$ is defined by

$$\cos\Theta(a,b) = \frac{a^T b}{\|a\|_2 \|b\|_2}, \qquad 0 \le \Theta(a,b) \le \pi, \tag{1.6}$$

and §4 relates this to some orthogonal transformations.

It is possible that A is available where the covariance matrix is $\Sigma = A^T A$, and §5 contains the main theorem (Theorem 5.1) which relates the Cholesky factorization and generalized Schur complements of Σ in §2, and the covariance matrices and partial correlation coefficients of §3, with the angles between vectors arising in the QR factorization of A in §5; the initial work on this was reported in [5]. We now expand on this.

We know that it is not generally advisable to form $A^T A$ numerically if we have A available, and that the Cholesky factor U of $\Sigma = A^T A = U^T U$ can more reliably be obtained as the upper triangular matrix U in the QR factorization of A (see for example [8], Chapter 6)

$$\begin{aligned} Q^T A &= \begin{pmatrix} U \\ 0 \end{pmatrix}, \qquad QQ^T = Q^T Q = I, \\ A &= Q_1 U, \qquad Q = (Q_1 \quad Q_2). \end{aligned} \tag{1.7}$$

The question now arises as to whether the Schur complements of Σ, or their factors, and the corresponding partial correlation coefficients, can be obtained directly from the QR factorization of A. Section 5 answers this by considering how algorithms for the QR factorization of A successively produce the orthogonal projection of a_j onto the space orthogonal to a_1, $R((\,a_1 \quad a_2\,))$, ..., $R((\,a_1 \quad \ldots \quad a_{j-1}\,))$, for each $2 \le j \le m$. The scalar $\rho_{jk}^{(i)}$ (the partial correlation of ξ_j and ξ_k keeping $\xi_1, \ldots, \xi_{i-1}$ fixed), $i \le j,k \le m$, is shown to be the cosine of the angle between the projections of a_j and a_k onto $R((\,a_1 \quad \ldots \quad a_{i-1}\,))^{\perp}$. It is shown how such correlations can be found from the QR factorization of A without forming A^TA or U^TU. The Cholesky factors of the Schur complements of leading principal submatrices of $\Sigma = A^TA$ can also be produced directly.

In §6 we consider the special case of having observations of random variables, and estimating sample partial correlation coefficients from these. Thus, when x_1, $x_2 \in \Re^n$ are vectors of n observations on two possibly related random variables ξ_1, and ξ_2, and their means are subtracted to give

$$a_i = x_i - e(x_i^Te)/n, \qquad i = 1,2, \qquad e = (\,1 \quad \ldots \quad 1\,)^T, \tag{1.8}$$

then $\rho_{12} = \cos\Theta(a_1, a_2)$ is called the *sample* correlation coefficient between ξ_1 and ξ_2. For n observations on each of m such random variables $\xi_1, \ldots, \xi_m$, the means can be subtracted to give $A = (\,a_1 \quad \ldots \quad a_m\,)$, and if

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix} = A^TA, \tag{1.9}$$

the sample correlation coefficients are

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}, \qquad \text{when } \sigma_{ii}\sigma_{jj} \neq 0, \qquad 1 \le i,j \le m. \tag{1.10}$$

The sample *partial* correlation coefficients are then defined in terms of the elements of the Schur complements of principal submatrices of Σ, in a similar manner to the population partial correlation coefficients mentioned earlier. Note that $(n-1)^{-1}A^TA$ is usually taken as the estimate for the population covariance matrix, but we will be able to ignore the scaling since correlations are independent of scaling. By an intelligent representation of the data we show that the computed results may be updated efficiently for certain changes in the data.

In §7 we consider the costs of computations, and how sets of partial correlation coefficients could be computed. We illustrate the loss of accuracy caused by forming $\Sigma = A^TA$ in finite precision when A is given, so that working with A could be important in some cases despite the greater efficiency of working with Σ.

In §8 we try to summarize the various results of the paper and emphasize the relations among them.

2. Cholesky Factors and Generalized Schur Complements

If $\Sigma(m \times m)$ is symmetric nonnegative definite it has a Cholesky factorization with upper triangular factor U:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix} = U^TU, \qquad U = \begin{pmatrix} u_1^T \\ \vdots \\ u_m^T \end{pmatrix} = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1m} \\ & \ddots & \vdots \\ & & \mu_{mm} \end{pmatrix}, \quad \mu_{ii} \ge 0. \tag{2.1}$$

If Σ is positive definite this is unique, but if Σ is singular it need not be. We will make it unique (see for example [10], p. 124) by demanding

$$0 = \mu_{i,i+1} = \cdots = \mu_{im} \qquad \text{if } \mu_{ii} = 0. \tag{2.2}$$

Clearly the first row u_1^T of U can be found from (2.1) and (2.2), and if we have found rows u_1^T, ..., u_{i-1}^T, we can find the ith row from $\Sigma^{(i)}$ in

$$\Sigma - u_1 u_1^T - \ \ldots\ - u_{i-1}u_{i-1}^T = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma^{(i)} \end{pmatrix}, \tag{2.3}$$

$$\Sigma^{(i)} = \begin{pmatrix} \sigma_{ii}^{(i)} & \ldots & \sigma_{im}^{(i)} \\ \vdots & & \vdots \\ \sigma_{mi}^{(i)} & \ldots & \sigma_{mm}^{(i)} \end{pmatrix} = \begin{pmatrix} \mu_{ii} & & \\ \vdots & \ddots & \\ \mu_{im} & \ldots & \mu_{mm} \end{pmatrix} \begin{pmatrix} \mu_{ii} & \ldots & \mu_{im} \\ & \ddots & \vdots \\ & & \mu_{mm} \end{pmatrix}.$$

The particular variant of the Cholesky algorithm we are interested in then has the form, with $\sigma_{ij}^{(1)} = \sigma_{ij}$, (note that we need only compute the upper triangular part),

Cholesky Algorithm

$$\begin{array}{lll} i = 1, \ldots, m & \mu_{ii} := (\sigma_{ii}^{(i)})^{1/2} & \\ & k = i+1, \ldots, m \quad \mu_{ik} := \begin{cases} 0 & \text{if } \mu_{ii} = 0 \\ \sigma_{ik}^{(i)}/\mu_{ii} & \text{otherwise} \end{cases} & \\ & j = i+1, \ldots, k \quad \sigma_{jk}^{(i+1)} := \sigma_{jk}^{(i)} - \mu_{ij}\mu_{ik} & \end{array} \tag{2.4}$$

Here, the μ_{ik} can overwrite the $\sigma_{ik}^{(i)}$, and the $\sigma_{jk}^{(i+1)}$ can overwrite the $\sigma_{jk}^{(i)}$. Thus, the elements of $\Sigma = \Sigma^{(1)}$, $\Sigma^{(2)}$, ..., $\Sigma^{(m)} = \sigma_{mm}^{(m)}$ are all formed by the algorithm. This is essentially the basic structure of algorithm SCHDC in LINPACK, see §8 and Appendix C.99-C.102 in [6].

If we partition Σ as in (1.3) with $(i-1) \times (i-1)$ Σ_{11}, then from (2.3)

$$\Sigma^{(i)} = U_{22}^T U_{22}, \tag{2.5}$$

and if Σ_{11} is nonsingular, (1.4) shows that $\Sigma^{(i)}$ is the Schur complement of Σ_{11} in Σ. If Σ_{11} is singular a generalized Schur complement is, see (1.5), (1.1) and (1.3),

$$S = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-}\Sigma_{12} = U_{22}^T U_{22} + U_{12}^T U_{12} - U_{12}^T U_{11}(U_{11}^T U_{11})^{-} U_{11}^T U_{12}, \tag{2.6}$$

where Σ_{11}^{-} is any generalized inverse of Σ_{11} satisfying

$$\Sigma_{11}\Sigma_{11}^{-}\Sigma_{11} = \Sigma_{11} = U_{11}^T U_{11}(U_{11}^T U_{11})^{-} U_{11}^T U_{11}. \tag{2.7}$$

But our choice (2.2) ensures in (1.3) that

$$R((\,U_{11} \quad U_{12}\,)) = R(U_{11}), \tag{2.8}$$

so $U_{12} = U_{11}B_{12}$ for some B_{12}, and substituting this in (2.6) and using (2.7) shows that (2.5) is the generalized Schur complement of singular Σ_{11} in Σ.

Now from (2.8)

$$R((\,\Sigma_{11} \quad \Sigma_{12}\,)) = R(U_{11}^T\,(\,U_{11} \quad U_{12}\,)) = R(U_{11}^T U_{11}) = R(\Sigma_{11}),$$

and this is the necessary and sufficient condition for the generalized Schur complement of Σ_{11} in symmetric Σ to be independent of the choice of generalized inverse Σ_{11}^{-}, see for example [2], Proposition 1. We can now summarize this result.

Result 2.1. *For any $m \times m$ symmetric nonnegative definite matrix Σ with elements $\sigma_{ij}^{(1)}$, the Cholesky algorithm (2.4) produces $\Sigma^{(i)}$ in (2.3), $i = 2, \ldots, m$, as the unique (generalized) Schur complement of the leading principal $i-1$ square submatrix in Σ. The algorithm then proceeds to find the Cholesky factor, see (2.3), of this generalized Schur complement.*

This is a simple and clean theoretical result, involving no generalized inverses, just a variant of the Cholesky factorization. It provides a strong argument for using this Cholesky factor as a powerful theoretical tool in this area.

3. Covariance Matrices and Correlation Coefficients

If $x = (\xi_1 \quad \ldots \xi_m)^T$ is a vector of random variables with $x \sim (a, \Sigma)$, which indicates x has mean

$$E[x] = a = (\alpha_1 \quad \ldots \quad \alpha_m)^T \tag{3.1}$$

and covariance

$$E[(x-a)(x-a)^T] = \Sigma = (\sigma_{ij}), \tag{3.2}$$

then a good measure of the degree of dependence between ξ_i and ξ_j is their covariance standardized by their standard deviations. This is called their *correlation coefficient*

$$\rho_{ij} = \frac{E[(\xi_i - \alpha_i)(\xi_j - \alpha_j)]}{\{E[(\xi_i - \alpha_i)^2]E[(\xi_j - \alpha_j)^2]\}^{1/2}} = \frac{\sigma_{ij}}{\sigma_{ii}^{1/2}\sigma_{jj}^{1/2}}. \tag{3.3}$$

Thus all the correlation coefficients are obtained directly from the covariance matrix.

If x and Σ are partitioned conformably

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \qquad \Sigma_{11} \text{ is } (i-1) \times (i-1), \tag{3.4}$$

and Σ is nonsingular, then Anderson [1], p. 41 states that the partial covariance of x_2 given x_1 can be defined as the covariance of the residual of x_2 from its regression on x_1. Anderson [1] §2.5 shows that this is just the Schur complement of Σ_{11} in Σ, which we have shown is $\Sigma^{(i)}$ in (2.5), the elements of which are the $\sigma_{jk}^{(i)}$ produced in (2.4). The *partial* correlation coefficients are defined in terms of the elements of this *partial* covariance matrix, just as was done in (3.3) for ordinary correlation coefficients.

Definition 3.1. The *partial correlation* between ξ_j and ξ_k, $j, k = i, \ldots, m$, given $\xi_1, \ldots, \xi_{i-1}$ is defined to be

$$\rho_{jk}^{(i)} = \frac{\sigma_{jk}^{(i)}}{(\sigma_{jj}^{(i)})^{1/2}(\sigma_{kk}^{(i)})^{1/2}}, \qquad \sigma_{jj}^{(i)}\sigma_{kk}^{(i)} \neq 0, \tag{3.5}$$

and is a measure of dependence between ξ_j and ξ_k when the effects of $\xi_1, \ldots, \xi_{i-1}$ have been removed.

The work here has shown that a numerically reliable and efficient way of obtaining these is to use (3.5) once the $\sigma_{jk}^{(i)}$ have been obtained directly from the Cholesky factorization (2.4). If x also has a multivariate *normal* distribution, so that $x \sim (a, \Sigma)$ completely determines the distribution, then Anderson [1] §2.5 shows that $\Sigma^{(i)}$ is the *conditional* covariance of x_2 given x_1, and then we can call (3.5) a *conditional* correlation.

The nonzero condition in (3.5) automatically holds when Σ is nonsingular, but we have included it for full generality. If Σ is singular and $\sigma_{jj}^{(i)} = 0$, then it is straightforward to show $\sigma_{jk}^{(i)} = \sigma_{kj}^{(i)} = 0$, $k = i, \ldots, m$, and $\rho_{jk}^{(i)}$ is undefined.

Although this is not a statistical paper, we will now show that the generalized Schur complement $\Sigma^{(i)}$ is the partial covariance matrix of x_2 given x_1, when Σ_{11} is singular. We do this to show the full generality of the results here, and to illustrate how easy it is to prove such results just using the Cholesky factorization (2.4) which provides $\Sigma^{(i)}$ and its factors so elegantly.

If $x \sim (a, \Sigma)$ and the Cholesky factor U of $\Sigma = U^T U$ is as in (2.4), then we may remove the zero rows of U to obtain $\bar{U}$ of full row rank and such that $\Sigma = \bar{U}^T \bar{U}$. It follows from Anderson [1], pp. 32-33 that with probability 1

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} \bar{U}_{11}^T & 0 \\ \bar{U}_{12}^T & \bar{U}_{22}^T \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = a + \bar{U}^T v, \qquad v \sim (0, I),$$

where $\bar{U}_{11}^T$ and $\bar{U}_{22}^T$ each have full column rank. If x_1 is given (such x_1 must lie in the linear variety $a_1 + \bar{U}_{11}^T v_1$, for some v_1, to be consistent) then v_1 is completely determined, and

$$x_2 = b_2 + \bar{U}_{22}^T v_2, \qquad b_2 = a_2 + \bar{U}_{12}^T v_1 \text{ given}, \quad v_2 \sim (0, I).$$

Now using the superscript $+$ to denote the Moore-Penrose generalized inverse,

$$b_2 = a_2 + \bar{U}_{12}^T (\bar{U}_{11}^T)^+ (x_1 - a_1)$$

can be defined as the regression of x_2 on x_1 ([1], p.41), so the residual $\bar{U}_{22}^T v_2 \sim (0, \Sigma^{(i)})$, and the partial covariance matrix is $\Sigma^{(i)}$.

It follows that (3.5) can also be used to define the partial correlation coefficients when Σ_{11} is singular, and so the Cholesky factorization (2.4) produces the partial covariance matrices $\Sigma^{(i)}$ and its factors, see (2.3), and then the partial correlation coefficients are given by (3.5), for any symmetric nonnegative definite covariance matrix $\Sigma = (\sigma_{ij}^{(1)})$.

4. Angles Between Vectors, and Orthogonal Transformations

We now consider some results which at first glance seem unrelated to §§2 and 3. We exhibit the connection at the end of §5.

We have defined the angle $\Theta(a, b)$ between two nonzero vectors $a, b \in \Re^n$ to satisfy

$$\cos \Theta(a, b) = \frac{a^T b}{\|a\|_2 \|b\|_2}, \qquad 0 \le \Theta(a, b) \le \pi. \tag{4.1}$$

If either a or b is zero the angle is not defined. We will be interested in orthogonal transformations, and the following obvious result indicates how angles are preserved.

Lemma 4.1. *For any matrix $Q_1(n \times r)$ such that $Q_1^T Q_1 = I$, and any nonzero vectors $c, d \in \Re^r$*

$$\Theta(Q_1 c, Q_1 d) = \Theta(c, d). \tag{4.2}$$

The next result exhibits a more interesting connection between orthogonal transformations and angles, and provides an alternative to (4.1) for obtaining angles.

Theorem 4.1. *Let $a, b \in \Re^n$ be nonzero vectors, and $(\,a \quad b\,) = Q_1 U$ a QR factorization with*

$$Q^T(\,a \quad b\,) = \begin{pmatrix} U \\ 0 \end{pmatrix}, \quad U = \begin{pmatrix} \mu_{11} & \mu_{12} \\ 0 & \mu_{22} \end{pmatrix}, \qquad \mu_{11} > 0, \quad \mu_{22} \geq 0, \tag{4.3}$$

and $Q = (\,Q_1 \quad Q_2\,)$ an $n \times n$ orthogonal matrix. Consider the orthogonal rotation through ϕ which transforms U to lower triangle, with

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \mu_{12} \\ \mu_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_{22} \end{pmatrix}, \qquad \lambda_{22} = (\mu_{12}^2 + \mu_{22}^2)^{1/2} > 0. \tag{4.4}$$

Then the angle between a and b may be obtained from either

$$\cos\Theta(a,b) = \sin\phi = \mu_{12}/\lambda_{22}, \tag{4.5}$$

$$\sin\Theta(a,b) = \cos\phi = \mu_{22}/\lambda_{22}. \tag{4.6}$$

The proper of the two solutions for (4.6) is retrieved by having the angle $\Theta(a,b)$ satisfy $\operatorname{sign}(\frac{\pi}{2} - \Theta(a,b)) = \operatorname{sign}(\mu_{12})$.

Proof. From Lemma 4.1, (4.1), and (4.4)

$$\cos\Theta(a,b) = \frac{\mu_{11}\mu_{12}}{\mu_{11}(\mu_{12}^2 + \mu_{22}^2)^{1/2}} = \frac{\mu_{12}}{\lambda_{22}} = \sin\phi,$$

the last step coming from multiplying (4.4) by the transpose of the rotation, which also gives $\mu_{22} = \lambda_{22}\cos\phi$, showing that $\cos\phi \geq 0$. But from (4.5) $\sin\Theta(a,b) = \pm\cos\phi$, whereas the definition (4.1) ensures $\sin\Theta(a,b) \geq 0$ and so (4.6) follows.

∎

To illustrate that this is a result of general computational value, consider floating point computations with precision δ, so that if $|\gamma| < \delta$ the computed result of adding γ to 1 is $\mathrm{fl}(1+\gamma) = 1$. With

$$(\,a \quad b\,) = \begin{pmatrix} 1 & 1 \\ 0 & \eta \end{pmatrix}, \qquad \eta \text{ a computer number}, \quad \eta^2 < \delta$$

computing (4.1) or (4.5) would give

$$\mathrm{fl}(\cos\Theta(a,b)) = 1,$$

leading us to believe $\Theta(a,b) = 0$. The alternative (4.6) would give

$$\mathrm{fl}(\sin\Theta(a,b)) = \eta,$$

telling us $\Theta(a,b) \approx \eta$, which is far more desirable.

Note, that by following the work of Wilkinson [13] the computations for $\cos\Theta(a,b)$ in (4.1) and (4.5), and for $\sin\Theta(a,b)$ in (4.6), can be shown to be numerically stable, so if we want the cosine or sine then these give all the accuracy we can expect. However $|\cos\Theta| \approx 1$ is relatively insensitive to small changes in Θ, and if we want Θ in such cases then we should avoid (4.1) and (4.5). On the other hand $|\sin\Theta| \approx 1$ is also relatively insensitive to small changes in θ, and (4.6) should not be used to compute Θ in such cases. Replacing b by $(\,\eta \quad 1\,)^T$ in the above example will illustrate this. In summary, if we want to compute $\Theta(a,b)$ we could use (4.1) or (4.5) when $|\cos\Theta(a,b)| \leq 2^{-1/2}$

and (4.6) otherwise. If the choice is between (4.5) and (4.6) only, this criterion would lead to (4.5) when $|\mu_{12}| \leq \mu_{22}$ and (4.6) otherwise.

Note that computing (4.1) takes about $3n$ flops (for a definition of 'flop' see [8], p. 32), while computing (4.3) takes about $4n$ flops at best, [8], p. 148.

In the remaining sections we will only discuss the *computation* of cosines, and so either (4.1) or (4.5) will suffice. However, we will present theory related to angles in general, so anyone intending to compute these actual angles can use the results of this section.

5. The QR Factorization and Angles Between Projected Vectors

Suppose $A(n \times m) = (a_1 \quad \dots \quad a_m)$, which need not have rank m. Here we will partition $A = (A_1 \quad A_2)$ with $A_1(n \times i-1)$, and describe how the QR factorization produces projections of columns of A_2 onto the space $R(A_1)^\perp$. We will later show that partial correlation coefficients are the cosines of angles between such projections when $\Sigma = A^T A$.

Suppose $Q(n \times n) = (Q_1 \quad Q_2)$ is an orthogonal matrix such that

$$A = (A_1 \quad A_2) = (Q_1 \quad Q_2) \begin{pmatrix} E & F \\ 0 & G \end{pmatrix}, \tag{5.1}$$

where $E(q \times i-1)$ is of rank q. This could be computed by the first q steps of the QR factorization, or the first q steps of the modified Gram-Schmidt (MGS) algorithm (see [8], Chapter 6), but we only require the zero block and the full row rank of E, giving

$$A_1 = Q_1 E, \qquad A_2 = Q_1 F + Q_2 G.$$

Let us define the two symmetric idempotent matrices

$$P_1 = Q_1 Q_1^T, \qquad P_2 = Q_2 Q_2^T.$$

Since E has full row rank, $R(A_1) = R(Q_1) = R(P_1)$ so P_1 is the orthogonal projector onto $R(A_1)$, $P_2 = I - P_1$ is the orthogonal projector onto $R(A_1)^\perp$, and

$$P_2 A_2 = Q_2 G = A_2 - Q_1 F \tag{5.2}$$

is the projection of the columns of A_2 onto $R(A_1)^\perp$. That is, the columns of $Q_2 G$ correspond to the columns of A_2 after the dependence on $a_1, \dots, a_{i-1}$ has been removed, and it is the correlation between these that is of interest in defining partial correlations. We first consider the angles between these vectors.

Definition 5.1. Let h_j be the projection of a_j onto the space orthogonal to $a_1, \dots, a_{i-1}$, where $A = (a_1 \quad \dots \quad a_m)$. We define $\Theta_{jk}^{(i)}(A)$ to be the angle between any pair of nonzero such h_j and h_k, $j, k = i, \dots, m$.

If we write $G = (g_i \quad \dots \quad g_m)$ we have $h_j = Q_2 g_j$, and from Lemma 4.1

$$\Theta_{jk}^{(i)}(A) = \Theta(Q_2 g_j, Q_2 g_k) = \Theta(g_j, g_k) \quad \text{if} \quad \|g_j\|_2 \|g_k\|_2 \neq 0. \tag{5.3}$$

The angle is undefined if either projection is zero. It follows that these angles are just the angles between columns of G in (5.1).

Lemma 4.1 in §4 described an invariance of angles under certain transformations, and the following result extends this.

Lemma 5.1. *For any matrix $\tilde{Q}_1(p \times n)$ such that $\tilde{Q}_1^T \tilde{Q}_1 = I$, if $B = (b_1 \quad \dots \quad b_m) = \tilde{Q}_1 A$ then $\Theta_{jk}^{(i)}(B)$ is defined if and only if $\Theta_{jk}^{(i)}(A)$ is defined, and in this case*

$$\Theta_{jk}^{(i)}(B) = \Theta_{jk}^{(i)}(\tilde{Q}_1 A) = \Theta_{jk}^{(i)}(A), \qquad j,k = i,\dots,m.$$

Proof.

$$B = \tilde{Q}_1 A = \tilde{Q}_1 Q \begin{pmatrix} E & F \\ 0 & G \end{pmatrix}$$

has the same G as the orthogonal factorization in (5.1) so the result follows from (5.3). ∎

Since (5.3) only depends on the full row rank of E and the zero block in (5.1), it is unaffected by orthogonal transformations applied to the left of $(E \quad F)$, or of G. In particular, orthogonal transformations of the form (4.3) and (4.4) could be used on G to find the $\Theta_{jk}^{(i)}(A)$.

Note that q steps of the QR factorization for computing (5.2) will provide Q_2 and G separately, whereas MGS only provides $Q_2 G$ when it is stopped after q steps. From (5.3) we see either will do for computing angles. In [8], Chapter 6 it is shown that for a complete QR factorization the Householder and fast Givens algorithms each require about $m^2(n - m/3)$ flops, while MGS requires about $m^2 n$. Often $n \gg m$ and there will be little difference in cost, but we would generally choose the QR factorization for its excellent numerical properties.

We now give the result which relates the work in §§2 and 3 with that in §§4 and 5 so far, see [5], and motivated the lengthy title for this paper.

Theorem 5.1. *If*

$$A^T A = \Sigma = (\sigma_{ij}^{(1)}), \tag{5.4}$$

then for $i = 1,\dots,m$ and $j,k = i,\dots,m$

$$\rho_{jk}^{(i)} = \cos \Theta_{jk}^{(i)}(A),$$

when either of these is defined. Here the $\rho_{jk}^{(i)}$ are defined by (3.5), where the $\sigma_{jk}^{(i)}$ are elements of the Schur complement $\Sigma^{(i)}$ of the leading principal $(i-1) \times (i-1)$ submatrix of Σ, and arise naturally in the Cholesky factorization (2.4) of Σ. The $\Theta_{jk}^{(i)}(A)$ are as in Definition 5.1 above.

When Σ is a scalar multiple of the (population *or* sample) covariance matrix of some distribution, these $\rho_{jk}^{(i)}$ are the corresponding partial correlation coefficients, see Definition 3.1. As a result partial correlation coefficients can also be defined in terms of angles between projected vectors. These projected vectors arise naturally from the QR factorization of A.

Proof. The $\rho_{jk}^{(i)}$ in (3.5) are a function of the elements $\sigma_{jk}^{(i)}$ of $\Sigma^{(i)} = U_{22}^T U_{22}$ obtained from the Cholesky factorization (2.4),

$$\Sigma = U^T U = \begin{pmatrix} U_{11}^T U_{11} & U_{11}^T U_{12} \\ U_{12}^T U_{11} & U_{12}^T U_{12} + U_{22}^T U_{22} \end{pmatrix}, \qquad U_{11}^T U_{11} \text{ is } (i-1 \times i-1), \tag{5.5}$$

where these products are unaffected if all the zero rows are removed from U, so we can henceforth assume U_{11} has full row rank.

In (5.3) we showed that $\Theta_{jk}^{(i)}(A)$ is the angle between columns g_j and g_k of G, where from (5.4) and (5.1)

$$\Sigma = A^T A = \begin{pmatrix} E^T E & E^T F \\ F^T E & F^T F + G^T G \end{pmatrix}, \tag{5.6}$$

where $E(q \times i-1)$ is of rank q, so that $F = EC$, $C = E^T(EE^T)^{-1}F$.

We now equate (5.5) and (5.6) to show $G^TG = U_{22}^TU_{22}$. The (1,1) and (1,2) blocks give

$$U_{11} = (U_{11}U_{11}^T)^{-1}U_{11}E^TE$$
$$U_{12} = (U_{11}U_{11}^T)^{-1}U_{11}E^TF = U_{11}C$$

so that

$$U_{12}^TU_{12} = C^TU_{11}^TU_{11}C = C^TE^TEC = F^TF,$$

and the desired result follows from equating the (2,2) blocks. This shows $g_j^Tg_k = \sigma_{jk}^{(i)}$, and combining (5.3), the definition of angle (4.1), and (3.5), proves the theorem.

∎

This result is essentially based on the uniqueness relation between the Cholesky factor of A^TA and the upper triangular matrix in the QR factorization of A. For full column rank A the proof would be briefer, but we have given the result in its full generality.

To summarize one practical advantage of this theorem, if we are given A where we know $\Sigma = A^TA$ is the sample or population covariance matrix of a distribution, then we can compute the corresponding partial correlation coefficients $\rho_{jk}^{(i)}$ directly from A, first by carrying out a QR-like factorization of the form (5.1), then finding $\cos\Theta(g_j, g_k)$ using either (4.1) or (4.3) with (4.5).

6. Observations of Random Variables

Suppose we have m random variables $\xi_1, \ldots, \xi_m$, and for $j = 1, \ldots, m$, $x_j = (\,\xi_{1j} \quad \ldots \quad \xi_{nj}\,)^T$ is a vector of n observations of ξ_j. Then for $j = 1, \ldots, m$

$$a_j = x_j - \frac{1}{n}e(x_j^Te), \qquad e = (\,1 \quad \ldots \quad 1\,)^T, \tag{6.1}$$

is x_j adjusted for its mean, and

$$\Sigma = \frac{1}{(n-1)}A^TA, \qquad A = (\,a_1 \quad \ldots \quad a_m\,),$$

is the usual estimate for the covariance matrix of the distribution of these random variables, and is called the *sample covariance matrix*. From this we could compute the sample partial covariance matrices and sample partial correlation coefficients exactly as in §3, and these would be estimates of the corresponding population values. However it is numerically preferable to work with A directly, and so these could also be computed as described in §§4 and 5.

A difficulty with this approach is that if we obtain $(\,\xi_{n+1,1} \quad \ldots \quad \xi_{n+1,m}\,)$, a new observation for each variable, and wish to update our values, then every mean has to be adjusted and every element of A is changed, as well as having a new row added. To avoid this, note that (6.1) can be written

$$a_j = Px_j, \qquad P = I - \frac{1}{n}ee^T,$$

where P is symmetric and idempotent, and is the orthogonal projector onto the space orthogonal to e. But from (5.1) and (5.2), one step of the QR factorization

$$(\,e \quad X\,) = (\,q_1 \quad Q_2\,)\begin{pmatrix}\alpha & b^T\\ 0 & B\end{pmatrix}, \qquad X = (\,x_1 \quad \ldots \quad x_m\,),$$

gives a matrix Q_2B whose columns are the projections of the columns of X onto the space orthogonal to e, and so are the required a_j. Now since $A = PX = Q_2Q_2^TX = Q_2B$, and all the values we

want can be found from $A^TA = B^TB$, we can work with B instead of A, and continuing the QR factorization of $(\,e \quad X\,)$ will produce everything needed in §5.

This is not just a nice way of effectively adjusting the x_j for their means. If we already have the QR factorization of $(\,e \quad X\,)$, we can add the new row $(\,1 \quad \xi_{n+1,1} \quad \dots \quad \xi_{n+1,m}\,)$ to $(\,e \quad X\,)$ and update the factorization using standard efficient and numerically reliable techniques, see for example [8] §12.6. We can also update the factorization efficiently when we discard a row of $(\,e \quad X\,)$, while adding or deleting columns (corresponding to random variables) is particularly easy. We could also update the Cholesky factor of given Σ when a row and corresponding column is added or deleted, see also [7, 10]. It is a straightforward exercise to compute the new partial correlation coefficients following such computations, however there is a danger of numerical errors in the case of discarding a row.

7. Some Computational Considerations

So far we have considered computing $\rho_{jk}^{(i)}$, the partial correlation coefficient of ξ_j and ξ_k, keeping $\xi_1, \dots, \xi_{i-1}$ fixed where $j,k = i,\dots,m$; and we have seen how to compute these for the fixed ordering $i = 1,\dots,m$. Of course any initial ordering of the columns of X in §6 or A in §5 or any symmetric ordering of columns and rows of Σ could be chosen, and during the factorizations pivoting can always be used on the unfactored part to determine which coefficients will be produced next.

Let us write ρ_{ij}^{S} where S is some set of indices, to denote the partial correlation coefficient of ξ_i and ξ_j keeping ξ_k fixed for all $k \in S$. When several lots of such coefficients are required it is usually the case that we want $\rho_{ij}^{S_k}$, $k = 1,2,\dots,t$, where $S_1 \subset S_2 \subset \cdots \subset S_t$. In this case we can order columns of A for example to give

$$(\,A_1 \quad A_2 \quad \dots \quad A_t \quad A_{t+1}\,) = \tilde{A}$$

where $(\,A_1 \quad \dots \quad A_k\,)$ has just those columns with indices in S_k, $k = 1,\dots,t$ and A_{t+1} contains the remaining columns. Applying the QR factorization to $\tilde{A}$ will allow us to compute everything in the correct order.

If we already have upper triangular U in $A = Q_1U$ (or $\Sigma = U^TU$), and we wish to compute values for a different ordering, we can always reorder columns (and rows) and update U to produce the required form. For example suppose $U = (\,u_1 \quad \dots \quad u_5\,)$ is the nonsingular upper triangular 5×5 matrix obtained from the QR factorization of A, and we now want to compute $\rho_{1,5}^{S}$ with $S = \{2,4\}$. We could arrange the columns and transform $(\,u_2 \quad u_4 \quad u_1 \quad u_5\,)$ by eliminating elements (2,1), (4,2) and (3,2):

$$\left(\begin{array}{cc|cc} * & * & * & * \\ * & * & & * \\ \hline & * & & * \\ & * & & * \\ & & & * \end{array}\right) \rightarrow \left(\begin{array}{cc|cc} * & * & * & * \\ & * & * & * \\ \hline & & * & * \\ & & & * \\ & & & * \end{array}\right) = \begin{pmatrix} E & F \\ 0 & G \end{pmatrix}$$

and $\rho_{1,5}^{S}$ is the cosine of the angle between the two columns of G. This could be found using (4.1), or by applying rotations as in (4.3) and (4.4), see (4.5).

We would tend to use the above approach to maintain as much numerical information as possible when we initially know A or X, but when we are given Σ, it could be worth considering more efficient ways of updating U, see [7]. Note that after $i-1$ steps of the Cholesky algorithm (2.4) all the $\sigma_{jk}^{(i)}$ are at hand, and so *all* the $\rho_{jk}^{(i)}$ in (3.5) can be computed in $m-i$ square roots and

$(m-i)(m-i+1)/2$ multiplications and divisions. On the other hand if we use the QR factorization, or already have

$$U = \begin{pmatrix} E & F \\ 0 & G \end{pmatrix},$$

we still need to find the cosines of the angles between the columns of G. If G is $n-i+1 \times m-i+1$ and full this would take $(n-i+1)(m-i+1)(m-i)/2$ flops to compute all the $g_j^T g_k$ in (4.1). The cost would be significantly less if $n \gg m$ and G was already upper triangular, but it would still be on the order of $(m-i)^3$, and so this QR approach is inherently more costly than the Cholesky approach if many partial correlation coefficients are required. However there are situations where the loss of accuracy in forming $A^T A$ can be significant, as we now show, see also [5].

Example. Suppose we are given

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & -2\epsilon \\ \epsilon & \epsilon & 1+\epsilon \\ -\epsilon & -\epsilon & -1+\epsilon \end{pmatrix},$$

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

$$\rho_{23}^{(2)} = \frac{\sigma_{23}^{(2)}}{\sqrt{\sigma_{22}^{(2)}}\sqrt{\sigma_{33}^{(2)}}}$$

is to be determined. The corresponding covariance matrix $A^T A$ is

$$\Sigma = \begin{pmatrix} 1+\epsilon^2 & -1+\epsilon^2 & 0 \\ -1+\epsilon^2 & 1+\epsilon^2 & 2\epsilon \\ 0 & 2\epsilon & 1+3\epsilon^2 \end{pmatrix},$$

and in exact arithmetic, one has

$$\sigma_{23}^{(2)} = 2\epsilon, \quad \sigma_{22}^{(2)} = 1+\epsilon^2 - \frac{(-1+\epsilon^2)^2}{1+\epsilon^2}, \quad \sigma_{33}^{(2)} = 1+3\epsilon^2$$

so that

$$\rho_{23}^{(2)} = \text{sign}(\epsilon)\sqrt{\frac{1+\epsilon^2}{1+3\epsilon^2}}.$$

However, in finite precision floating point arithmetic and with ϵ chosen to be sufficiently small ($2\epsilon < \delta^{1/2}$, where δ is the floating point precision so that the *computed* $\text{fl}(1 \pm 4\epsilon^2) = 1$), the computed quantities turn out to be

$$\text{fl}(\Sigma) = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 2\epsilon \\ 0 & 2\epsilon & 1 \end{pmatrix}$$

and

$$\sigma_{23}^{(2)} = 2\epsilon, \quad \sigma_{11}^{(2)} = 0, \quad \sigma_{33}^{(2)} = 1$$

so that $\rho_{23}^{(2)}$ is not a finite number.

Performing a QR decomposition of the matrix A yields a 4×3 matrix Q with orthonormal columns and a 3×3 upper triangular factor

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

in exact arithmetic. In finite precison arithmetic with the same choice of ϵ as before, one has

$$\text{fl}(U) = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2|\epsilon| & \text{sign}(\epsilon) \\ 0 & 0 & \sqrt{2}|\epsilon| \end{pmatrix},$$

and the sine of the rotation that transforms the submatrix

$$\begin{pmatrix} 2|\epsilon| & \text{sign}(\epsilon) \\ 0 & \sqrt{2}|\epsilon| \end{pmatrix}$$

to lower triangular form is equal to $\text{fl}(\rho_{23}^{(2)}) = \text{sign}(\epsilon)$, the exact value to machine precision.

8. Conclusions

In this paper we have shown several relationships between theoretical objects and algorithms, and used these to suggest approaches to numerical computations. Several different topics have been considered, and it is important to see how they all fit together.

It is a standard result that the Cholesky factors of the Schur complements $\Sigma^{(i)}$ of all leading principal submatrices of symmetric positive definite Σ are produced by the Cholesky factorization of Σ. It is probably less well known (at least we know of no reference) that not only the Cholesky factors, but also the Schur complements themselves are produced directly (that, without having to form the products of their factors) with the correct organization of the Cholesky algorithms, see (2.4). We have extended this result to show it also holds for singular Σ when we consider the *generalized* Schur complements $\Sigma^{(i)}$. This is summarized in Result 2.1.

When $\Sigma = (\sigma_{ij})$ above is the nonsingular covariance matrix of a vector of random variables $x \sim (a, \Sigma)$, then $\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$ are the correlation coefficients. The Schur complement $\Sigma^{(i)} = (\sigma_{jk}^{(i)})$, $j,k = i,\ldots,m$ of the leading $(i-1) \times (i-1)$ submatrix of Σ is called the partial covariance of elements i to n of x, given the first $i-1$ elements. The corresponding partial correlation coefficients are then $\rho_{jk}^{(i)} = \sigma_{jk}^{(i)}/(\sigma_{jj}^{(i)}\sigma_{kk}^{(i)})^{1/2}$. Since the $\sigma_{jk}^{(i)}$ are given directly by the Cholesky algorithm (2.4), the $\rho_{jk}^{(i)}$ can be computed from them. We have shown that even when Σ is singular, the generalized Schur complement $\Sigma^{(i)}$ given by the Cholesky algorithm (2.4) is still the required partial covariance matrix, and so the partial correlation coefficients can be computed as indicated above whenever $\sigma_{jj}^{(i)}\sigma_{kk}^{(i)}$ is nonzero. If x has a multivariate normal distribution these partial covariances and correlations are also conditional covariances and correlations.

If A is available where $\Sigma = A^TA$, then the upper triangular Cholesky factor U of Σ is given by the QR factorization of A. It was shown in [5] how partial correlation coefficients could be computed from U without forming U^TU. In §5 we rephrased those results to emphasize that partial correlation coefficients are the cosines of angles between vectors appearing in the QR factorization of A. As a result, partial correlation coefficients for different sets of given elements of the random vector x can be obtained by permuting the columns of A and updating the QR factorization, see [5]. Again this did not require Σ to be nonsingular.

The two main sequences of relationships have therefore been, for a (possibly singular) covariance matrix $\Sigma = A^T A$:

1. Cholesky factorization of Σ, as implemented in (2.4)
   - (generalized) Schur complements of leading submatrices of Σ
   - partial covariance matrices
   - partial correlation coefficients.
2. QR (or MGS) factorization of A
   - projections of later columns of A onto spaces orthogonal to earlier columns
   - cosines of angles between these projections
   - partial correlation coefficients.

Of course both the Cholesky factorization of Σ and the QR factorization of A produce Cholesky factors of each leading principal submatrix of Σ and of the corresponding (generalized) Schur complement, but these are well known and not part of the above sequences.

When the population covariance matrix for the random vector x is not available, and A is obtained from n observations of each of the random variables as shown in §6, then $\Sigma = (n-1)^{-1} A^T A$ is the sample covariance matrix. The sample partial covariance matrices are the generalized Schur complements of the principal submatrices of Σ, and the sample partial correlation coefficients can be found from the Cholesky factorization of Σ or the QR factorization of A, just as was described for the population partial correlation coefficients. In this case however, we may obtain further observations on each random variable, and §6 indicated how the computations may be arranged to include these, and update the factorization and corresponding partial correlation coefficients. Ways of computing different sets of partial correlation efficiently were considered in §7.

Reliable numerical computation has been a guide in our search for connections and algorithms here, and in §4 we made some general observations on computing angles between vectors, and the possible effect of finite precision. More importantly it was pointed out in [5] and repeated in §7 here, how working with the QR factorization of A rather than the Cholesky factorization of $A^T A$ can be numerically crucial in some cases of computing partial correlation coefficients, even though the $A^T A$ approach may be faster.

In a forthcoming paper will be presented a derivation of this QR approach from the parallel implementation of an algorithm for solving symmetric positive definite linear systems by hyperbolic rotations, as shown in [4].

References

[1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis, Second Edition,* Wiley, 1984.

[2] Carlson, D., *What are Schur Complements, Anyway?,* Linear Algebra and its Applications, 74 (1986), pp. 257–75.

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

[4] Delosme, J.-M. and Ipsen, I.C.F., *Parallel Solution of Symmetric Positive Definite Systems with Hyperbolic Rotations,* Linear Algebra and its Applications, 77 (1986), pp. 75–111.

[5] ————, *Computing Partial Correlations from the Data Matrix,* Research Report 541, Dept Computer Science, Yale University, 1987.

[6] Dongarra, J.J., Bunch, J.R., Moler, C.B. and Stewart, G.W., *LINPACK Users' Guide,* SIAM, Philadelphia, 1979.

[7] Gill, P.E., Golub, G.H., Murray, W. and Saunders, M.A., *Methods for Modifying Matrix Factorizations,* Math. Comp., 28 (1974), pp. 505–35.

[8] Golub, G.H. and van Loan, C.F., *Matrix Computations,* The Johns Hopkins Press, Baltimore, MD, 1983.

[9] Higham, N.J., Analysis of the Cholesky Decomposition of a Semi-Definite Matrix, Cox, M. and Hammarling, S. eds., *Reliable Numerical Computation,* Oxford University Press, 1988. To appear.

[10] Lawson, C.L. and Hanson, R.J., *Solving Least Squares Problems,* Prentice Hall, 1974.

[11] Marsaglia, G. and Styan, G.P.H., *Equalities and Inequalities for Ranks of Matrices,* Linear and Multilinear Algebra, 2 (1974), pp. 269–92.

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

[13] Wilkinson, J.H., *The Algebraic Eigenvalue Problem,* Clarendon Press, Oxford, 1965.