

***Computing a Nearest Correlation Matrix with  
Factor Structure***

Borsdorf, Rüdiger and Higham, Nicholas J. and Raydan,  
Marcos

2009

MIMS EPrint: **2009.87**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary  
School of Mathematics  
The University of Manchester  
Manchester, M13 9PL, UK

ISSN 1749-9097

# COMPUTING A NEAREST CORRELATION MATRIX WITH FACTOR STRUCTURE\*

RÜDIGER BORSDORF<sup>†</sup>, NICHOLAS J. HIGHAM<sup>‡</sup>, AND MARCOS RAYDAN<sup>§</sup>

**Abstract.** An  $n \times n$  correlation matrix has  $k$  factor structure if its off-diagonal agrees with that of a rank  $k$  matrix. Such correlation matrices arise, for example, in factor models of collateralized debt obligations (CDOs) and multivariate time series. We analyze the properties of these matrices and, in particular, obtain an explicit formula for the rank in the one factor case. Our main focus is on the nearness problem of finding the nearest  $k$  factor correlation matrix  $C(X) = \text{diag}(I - XX^T) + XX^T$  to a given symmetric matrix, subject to natural nonlinear constraints on the elements of the  $n \times k$  matrix  $X$ , where distance is measured in the Frobenius norm. For a special one parameter case we obtain an explicit solution. For the general  $k$  factor case we obtain the gradient and Hessian of the objective function and derive an instructive result on the positive definiteness of the Hessian when  $k = 1$ . We investigate several numerical methods for solving the nearness problem: the alternating directions method; a principal factors method used by Anderson, Sidenius, and Basu in the CDO application, which we show is equivalent to the alternating projections method and lacks convergence results; the spectral projected gradient method of Birgin, Martínez, and Raydan; and Newton and sequential quadratic programming methods. The methods differ in whether or not they can take account of the nonlinear constraints and in their convergence properties. Our numerical experiments show that the performance of the methods depends strongly on the problem, but that the spectral projected gradient method is the clear winner.

**Key words.** correlation matrix, factor structure, patterned covariance matrix, positive semidefinite matrix, Newton's method, principal factors method, alternating directions method, alternating projections method, spectral projected gradient method

**AMS subject classifications.** 65F30, 90C30

**1. Introduction.** In many practical applications involving statistical modelling it is required to adjust an approximate, empirically obtained correlation matrix so that it has the three defining properties of a correlation matrix: symmetry, positive semidefiniteness, and unit diagonal. Lack of definiteness can result from missing or asynchronous data which, in the case of financial modelling, may be due to a company being formed or ceasing to trade during the period of interest or markets in different regions trading at different times and having different holidays. Furthermore, stress testing may require individual correlations to be artificially adjusted, with subsequent Value-at-Risk analysis breaking down if the perturbed matrix is not a correlation matrix [9], [32]. In a variety of applications it is natural to replace the given empirical matrix by the nearest correlation matrix in the (weighted) Frobenius norm [15], [33], [37], [44]. This problem has received much attention in the last few years and can be solved using the alternating projections method [15] or a preconditioned Newton method [5], [31], the latter having quadratic convergence and being the method of choice.

---

\*Version of November 9, 2009. This work was supported by a Royal Society-Wolfson Research Merit Award to the second author.

<sup>†</sup>School of Mathematics, University of Manchester, Manchester, M13 9PL, UK (borsdorf@maths.man.ac.uk, <http://www.maths.manchester.ac.uk/~borsdorf/>). This author was supported by EPSRC, the School of Mathematics at the University of Manchester, and the Numerical Algorithms Group Ltd.

<sup>‡</sup>School of Mathematics, University of Manchester, Manchester, M13 9PL, UK (higham@maths.man.ac.uk, <http://www.maths.manchester.ac.uk/~higham/>).

<sup>§</sup>Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, Ap. 89000, Caracas 1080-A, Venezuela (mraydan@usb.ve).

In this work we are interested in the nearness problem in which factor structure is imposed on the correlation matrix. Such structure arises in factor models of asset returns [6, Sec. 3.5], collateralized debt obligations (CDOs) [2], [11], [19], and multivariate time series [25]. To motivate this structure we consider the factor model<sup>1</sup>

$$(1.1) \quad \xi = X\eta + F\varepsilon$$

for the random vector  $\xi \in \mathbb{R}^n$ , where  $X \in \mathbb{R}^{n \times k}$ ,  $F \in \mathbb{R}^{n \times n}$  is diagonal, and  $\eta \in \mathbb{R}^k$  and  $\varepsilon \in \mathbb{R}^n$  are vectors of independent random variables having zero mean and unit variance, with  $\eta$  and  $\varepsilon$  independent of each other. In the terminology of factor analysis [28] the components of  $\eta$  are the factors and  $X$  is the loading matrix. With  $\text{cov}(\cdot)$  and  $E(\cdot)$  denoting the covariance matrix and the expectation operator, respectively, it follows that  $E(\xi) = 0$  and hence

$$(1.2) \quad \text{cov}(\xi) = E(\xi\xi^T) = XX^T + F^2.$$

If we assume that the variance of  $\xi_i$  is 1 for all  $i$  then  $\text{cov}(\xi)$  is the correlation matrix of  $\xi$  and (1.2) gives  $\sum_{j=1}^k x_{ij}^2 + f_{ii}^2 = 1$ , so that

$$(1.3) \quad \sum_{j=1}^k x_{ij}^2 \leq 1, \quad i = 1:n.$$

This model produces a correlation matrix of the form

$$(1.4a) \quad C(X) = D + \sum_{j=1}^k x_j x_j^T = D + XX^T,$$

$$(1.4b) \quad X = [x_1, \dots, x_k] = \begin{bmatrix} y_1^T \\ \vdots \\ y_n^T \end{bmatrix} = Y^T \in \mathbb{R}^{n \times k},$$

$$(1.4c) \quad D = \text{diag}(I - XX^T) = \text{diag}(1 - y_i^T y_i),$$

and we say  $C(X)$  has *k factor correlation matrix structure*. Note that  $C(X)$  can be written in the form

$$C(X) = \begin{bmatrix} 1 & y_1^T y_2 & \cdots & y_1^T y_n \\ y_1^T y_2 & 1 & \cdots & \vdots \\ \vdots & \cdots & \ddots & y_{n-1}^T y_n \\ y_1^T y_n & \cdots & y_{n-1}^T y_n & 1 \end{bmatrix},$$

where  $y_i \in \mathbb{R}^k$ . While  $C(X)$  can be indefinite for general  $X$ , the constraints (1.3) ensure that  $XX^T$  has diagonal elements bounded by 1, which means that  $C(X)$  is the sum of two positive semidefinite matrices and hence is positive semidefinite. In general,  $C(X)$  is of full rank; correlation matrices of low rank, studied in [13], [29], [46], for example, form a very different set.

The problem of computing a correlation matrix of  $k$  factor structure nearest to a given matrix is posed in the context of credit basket securities by Anderson,

<sup>1</sup>This model is referred to in [11] as the “multifactor copula model”.

Sidenius, and Basu [2], wherein an ad hoc iterative method for its solution is described. The problem is also discussed by Glasserman and Suchintabandit [11, Sec. 5] and Jäckel [19]. Here, we give theoretical analysis of the problem and show how standard optimization methods can be used to tackle it.

We begin in Section 2 by considering a correlation matrix depending on just one parameter, for which an explicit solution to the nearness problem is available. The one factor ( $n$  parameter) case is treated in Section 3, where results on the representation, determinant, and rank of  $C(X)$  are given, along with formulae for the gradient and Hessian of the relevant objective function and a result on the definiteness of the Hessian. In Section 4 we consider the general  $k$  factor problem and derive explicit formulae for the relevant gradient and Hessian.

Several suitable numerical methods are presented in Section 5. We show that the principal components-based method proposed in [2] is an alternating projections method and explain why it cannot be guaranteed to converge. Other methods considered are an alternating directions method, a spectral projected gradient method, and Newton and sequential quadratic programming (SQP) methods. We also derive a good starting matrix for use with any method. In Section 6 we give numerical experiments to compare the performance of the methods. Conclusions are given in Section 7.

Throughout, we will use the Frobenius norm  $\|A\|_F = \langle A, A \rangle^{1/2}$  on  $\mathbb{R}^{n \times n}$ , where the inner product  $\langle A, B \rangle = \text{trace}(B^T A)$ .

**2. One parameter problem.** We begin by considering a one parameter matrix  $C(w)$  that has unit diagonal and every off-diagonal element equal to  $w \in \mathbb{R}$ :

$$(2.1) \quad C(w) = (1 - w)I + wee^T = I + w(ee^T - I),$$

where  $e = [1, 1, \dots, 1]^T$ . This matrix is more general than the special case  $C(\theta e)$  of the one factor matrix considered in the next section because in that case  $w \equiv \theta^2$  is forced to be nonnegative. This structure corresponds to a covariance matrix with constant diagonal and constant off-diagonal elements—a simple but frequently occurring pattern [1], [17], [20, p. 55], [23], [36], [45].

LEMMA 2.1.  $C(w) \in \mathbb{R}^{n \times n}$  ( $n \geq 2$ ) is a correlation matrix if and only if

$$(2.2) \quad \frac{-1}{n-1} \leq w \leq 1.$$

*Proof.*  $C(w)$  is a correlation matrix precisely when it is positive semidefinite. The eigenvalues of  $C(w)$  are  $1 + (n-1)w$  and  $n-1$  copies of  $1-w$ , so  $C(w)$  is positive semidefinite precisely when (2.2) holds.  $\square$

We can give an explicit solution to the corresponding nearness problem,

$$(2.3) \quad \min\{\|A - C(w)\|_F : C(w) \text{ is a correlation matrix}\}.$$

THEOREM 2.2. For  $A \in \mathbb{R}^{n \times n}$ ,

$$\min_w \|A - C(w)\|_F^2 = \|A - I\|_F^2 - \frac{(e^T A e - \text{trace}(A))^2}{n^2 - n}$$

and the minimum is attained uniquely at

$$(2.4) \quad w_{\text{opt}} = \frac{e^T A e - \text{trace}(A)}{n^2 - n}.$$

The problem (2.3) has a unique solution given by the projection of  $w_{\text{opt}}$  onto the interval  $[-1/(n-1), 1]$ .

*Proof.* We want the global minimizer of

$$\begin{aligned} f(w) &:= \|A - (I + w(ee^T - I))\|_F^2 \\ &= \|A - I\|_F^2 + w^2 \|ee^T - I\|_F^2 - 2\text{trace}((A - I)w(ee^T - I)) \\ &= \|A - I\|_F^2 + w^2(n^2 - n) - 2w\text{trace}(Aee^T - A - ee^T + I) \\ &= \|A - I\|_F^2 + w^2(n^2 - n) - 2w(e^T Ae - \text{trace}(A)). \end{aligned}$$

Since  $f'(w) = 2w(n^2 - n) - 2(e^T Ae - \text{trace}(A))$ ,  $f$  has a unique stationary point at  $w_{\text{opt}}$  given by (2.4). From  $f''(w) = 2(n^2 - n) > 0$  it follows that  $f$  is strictly convex so  $w_{\text{opt}}$  is a local and hence global minimizer. The last part follows from the convexity of  $f$ .  $\square$

It is known [15, Thm. 2.5] that if  $a_{ii} \equiv 1$  and  $A$  has  $t$  nonpositive eigenvalues then the solution to  $\min\{\|A - X\|_F : X \text{ is a correlation matrix}\}$  has at least  $t$  zero eigenvalues. By contrast, from Theorem 2.2 we see that for  $a_{ii} \equiv 1$  the solution to problem (2.3) has exactly one zero eigenvalue when  $w_{\text{opt}} \leq -1/(n-1)$  (i.e.,  $e^T Ae \leq 0$ ), and exactly  $n - 1$  zero eigenvalues when  $w_{\text{opt}} \geq 1$  (i.e.,  $e^T Ae \geq n^2$ ), and otherwise the solution is nonsingular.

A more general version of  $C(w)$  arises when variables in an underlying model are grouped and separate intra- and inter-group correlations are defined [12]. The correlation matrix is now a block  $m \times m$  matrix  $C(\Gamma) = (C_{ij}) \in \mathbb{R}^{n \times n}$ , where  $\Gamma \in \mathbb{R}^{m \times m}$  and

$$(2.5) \quad C_{ij} = \begin{cases} C(\gamma_{ii}) \in \mathbb{R}^{n_i \times n_i}, & i = j, \\ \gamma_{ij} ee^T \in \mathbb{R}^{n_i \times n_j}, & i \neq j, \end{cases}$$

with  $n = \sum_{i=1}^m n_i$ . The objective function is, with  $A = (A_{ij})$  partitioned conformally with  $C$ ,

$$(2.6) \quad f(\Gamma) = \|A - C(\Gamma)\|_F^2 = \sum_{i=1}^m \|A_{ii} - C(\gamma_{ii})\|_F^2 + \sum_{i \neq j} \|A_{ij} - \gamma_{ij} ee^T\|_F^2.$$

The problem is to minimize  $f(\Gamma)$  subject to  $C$  being in the intersection of the set of positive semidefinite matrices and the set  $\mathcal{C}$  of all patterned matrices of the form (2.5). Both these sets are closed convex sets and hence so is their intersection. It follows from standard results in approximation theory (see, for example, [26, p. 69]) that the problem has a unique solution. This solution can be computed by the alternating projections method, by repeatedly projecting onto the two sets in question. To obtain the projection onto the set  $\mathcal{C}$  we simply apply Theorem 2.2 to each term in the first summation in (2.6) and for  $i \neq j$  set  $\gamma_{ij} = \sum_{(p,q) \in S_{ij}} a_{pq} / |S_{ij}|$ , where  $S_{ij}$  is the set of indices of the elements in  $A_{ij}$  and  $|S_{ij}|$  is the number of elements in  $S_{ij}$ . The latter projection can trivially be incorporated into Algorithm 3.3 of [15], replacing the projection onto the unit diagonal matrices therein, without losing the algorithm's guaranteed convergence.

If the inter-group correlations are equal and nonnegative, say  $\gamma_{ij} \equiv \beta \geq 0$ , and additionally all intra-group correlations satisfy  $\gamma_{ii} \geq \beta$ , the matrix  $C(\Gamma)$  can be represented as an  $m + 1$  factor correlation matrix  $C(X)$ , with  $X \in \mathbb{R}^{n \times (m+1)}$  a block

$m \times (m + 1)$  matrix  $X = (X_{ij})$  with  $X_{ij} \in \mathbb{R}^{n_i}$ , where

$$X_{ij} = \begin{cases} \sqrt{\beta}e \in \mathbb{R}^{n_i}, & j = 1, \\ \sqrt{\gamma_{ii} - \beta}e \in \mathbb{R}^{n_i}, & j = i + 1, \\ 0, & \text{otherwise.} \end{cases}$$

To illustrate, we consider a small example where  $m = 2$  and  $n_1 = n_2 = 2$ . Then  $X$  is a block  $2 \times 3$  matrix and

$$XX^T = \left[ \begin{array}{cc|cc} \sqrt{\beta} & \sqrt{\gamma_{11} - \beta} & 0 & 0 \\ \sqrt{\beta} & \sqrt{\gamma_{11} - \beta} & 0 & 0 \\ \hline \sqrt{\beta} & 0 & \sqrt{\gamma_{22} - \beta} & 0 \\ \sqrt{\beta} & 0 & \sqrt{\gamma_{22} - \beta} & 0 \end{array} \right] \left[ \begin{array}{cc|cc} \sqrt{\beta} & \sqrt{\beta} & \sqrt{\beta} & \sqrt{\beta} \\ \hline \sqrt{\gamma_{11} - \beta} & \sqrt{\gamma_{11} - \beta} & 0 & 0 \\ \hline 0 & 0 & \sqrt{\gamma_{22} - \beta} & \sqrt{\gamma_{22} - \beta} \end{array} \right],$$

which simplifies to the desired form

$$\left[ \begin{array}{cc|cc} \gamma_{11} & \gamma_{11} & \beta & \beta \\ \gamma_{11} & \gamma_{11} & \beta & \beta \\ \hline \beta & \beta & \gamma_{22} & \gamma_{22} \\ \beta & \beta & \gamma_{22} & \gamma_{22} \end{array} \right].$$

**3. One factor problem.** We now consider the one factor problem, for which the correlation matrix has the form, taking  $k = 1$  in (1.4),

$$(3.1) \quad C(x) = \text{diag}(1 - x_i^2) + xx^T, \quad x \in \mathbb{R}^n.$$

The off-diagonal part of  $C(x)$  agrees with that of the rank-1 matrix  $xx^T$ , so  $C(x)$  is of the general diagonal plus semiseparable form [41].

We first consider the uniqueness of this representation.

**THEOREM 3.1.** *Let  $C = C(x)$  for some  $x \in \mathbb{R}^n$  with  $p$  nonzero elements ( $0 \leq p \leq n$ ). If  $p = 1$  then  $C = I$  and  $C = C(y)$  for any  $y$  with at least  $n - 1$  zero entries. If  $p = 2$  and  $x_i, x_j$  are the nonzero entries of  $x$  then  $C = C(y)$  for  $y = \theta x_i e_i + \theta^{-1} x_j e_j$  for any  $\theta \neq 0$ . Otherwise,  $C = C(y)$  for exactly two vectors:  $y = \pm x$ .*

*Proof.* Without loss of generality we can assume  $C = \text{diag}(1 - x_i^2) + xx^T$  has been symmetrically permuted so that  $x_i \neq 0$  for  $i = 1:p$  and  $x_i = 0$  for  $i = p+1:n$ . If  $p = 1$  then  $C = I$  and  $x_1$  is arbitrary, which gives the first part. Suppose  $p > 1$ . We can write

$$(3.2) \quad C = \begin{bmatrix} C_1 & 0 \\ 0 & I \end{bmatrix},$$

where  $C_1 \in \mathbb{R}^{p \times p}$  has all nonzero elements. If  $p = 2$  then  $c_{12} = x_1 x_2 = \theta x_1 \cdot \theta^{-1} x_2 \equiv y_1 y_2$  for any  $\theta \neq 0$  and  $C = C(y)$  with  $y_3, \dots, y_n$  necessarily zero. Assume  $p > 2$  and suppose  $C = \text{diag}(1 - y_i^2) + yy^T$ . Then, from (3.2),  $y_i \neq 0$  for  $i = 1:p$  and  $y_i = 0$  for  $i = p+1:n$ . From  $C = \text{diag}(1 - y_i^2) + yy^T$  we have

$$(3.3) \quad \frac{c_{i,i+1}c_{i,i+2}}{c_{i+1,i+2}} = y_i^2, \quad 1 \leq i \leq p-2,$$

which determines the first  $p-2$  components of  $y_i$  up to their signs, and  $y_p$  is determined by  $y_{p-2}y_p = c_{p-2,p}$  and  $y_{p-1}$  by  $y_{p-1}y_p = c_{p-1,p}$ . Finally, the equations  $c_{1j} = y_1 y_j$ ,  $1 \leq j \leq p$  ensure that  $\text{sign}(y_j)$ ,  $2 \leq j \leq p$ , is determined by  $\text{sign}(y_1)$ .  $\square$

Before addressing the nearness problem we develop some properties of  $C(x)$ .

LEMMA 3.2. *The determinant of  $C(x)$  is given by*

$$(3.4) \quad \det(C(x)) = \prod_{i=1}^n (1 - x_i^2) + \sum_{i=1}^n x_i^2 \prod_{\substack{j=1 \\ j \neq i}}^n (1 - x_j^2).$$

*Proof.* Define the vector  $z(\epsilon)$  by  $z_i = x_i + \epsilon$ . For sufficiently small  $\epsilon$ ,  $z(\epsilon)$  has no element equal to 1 and  $D = \text{diag}(1 - z_i^2)$  is nonsingular. Hence  $C(z) = D + zz^T = D(I + D^{-1}z \cdot z^T)$ , from which it follows that

$$\det(C(z)) = \det(D)(1 + z^T D^{-1}z) = \prod_{i=1}^n (1 - z_i^2) \cdot \left(1 + \sum_{i=1}^n \frac{z_i^2}{1 - z_i^2}\right).$$

On multiplying out, the formula takes the form (3.4) with  $x$  replaced by  $z(\epsilon)$ , and letting  $\epsilon \rightarrow 0$  gives the result, since the determinant is a continuous function of the matrix elements.  $\square$

For the case  $x_i \neq 1$  for all  $i$  the formula (3.4) is a special case of a result in [34, Sec. 2.1].

COROLLARY 3.3. *If  $|x| \leq e$  with  $x_i = 1$  for at most one  $i$  then  $C(x)$  is nonsingular.  $C(x)$  is singular if  $x_i = x_j = 1$  for some  $i \neq j$ .*

The matrix  $C(x)$  is not always a correlation matrix because it is not always positive semidefinite. We know from the discussion of the  $k$  factor case in Section 1 that a sufficient condition for  $C(x)$  to be a correlation matrix is that  $|x| \leq e$ . This condition arises in the factor model described in Section 1 and hence is natural in the applications. The two extreme cases are when  $|x| = e$ , in which case  $C = xx^T$  is of rank 1, and when  $x = 0$ , in which case  $C = I$  has rank  $n$ . The next result shows more generally that the rank is determined by the number of elements of  $x$  of modulus 1.

THEOREM 3.4. *For  $C = C(x) \in \mathbb{R}^{n \times n}$  in (3.1) with  $|x| \leq e$  we have  $\text{rank}(C) = \min(p + 1, n)$ , where  $p$  is the number of  $x_i$  for which  $|x_i| < 1$ .*

*Proof.* By a symmetric permutation of  $C$  we can assume, without loss of generality, that  $|x_i| < 1$  for  $i = 1:p$  and  $|x_i| = 1$  for  $i = p + 1:n$ . The result is true for  $p = n$  by Corollary 3.3, so assume  $p \leq n - 1$ . Partition  $x = [y, z]^T$ , where  $y \in \mathbb{R}^p$ ; thus  $|y| < e$  and  $|z| = e$ . Then

$$C = \begin{bmatrix} C_1 & yz^T \\ zy^T & zz^T \end{bmatrix},$$

where  $C_1 \in \mathbb{R}^{p \times p}$  is positive definite. With  $X^T = \begin{bmatrix} I & 0 \\ -zy^T C_1^{-1} & I \end{bmatrix}$  we have

$$X^T C X = \begin{bmatrix} C_1 & 0 \\ 0 & S \end{bmatrix},$$

where

$$S = zz^T - zy^T C_1^{-1} yz^T = zz^T - (y^T C_1^{-1} y)zz^T = (1 - y^T C_1^{-1} y)zz^T.$$

Hence  $\text{rank}(C) = \text{rank}(C_1) + \text{rank}(S) = p + \text{rank}(S)$ . Now  $C_1 = \text{diag}(1 - y_i^2) + yy^T =: D + yy^T$ , where  $D$  is positive definite, and the Sherman–Morrison formula gives

$$C_1^{-1} = D^{-1} - \frac{D^{-1}yy^T D^{-1}}{1 + y^T D^{-1}y}.$$

So

$$y^T C_1^{-1} y = \frac{y^T D^{-1} y}{1 + y^T D^{-1} y} < 1.$$

Since  $y^T C_1^{-1} y \neq 1$  and  $z \neq 0$ ,  $S$  has rank 1 and the result follows.  $\square$

Now we are ready to address the nearness problem. Consider the problem of minimizing

$$(3.5) \quad f(x) = \|A - (\text{diag}(1 - x_i^2) + xx^T)\|_F^2,$$

subject to  $|x| \leq e$ , where  $A \in \mathbb{R}^{n \times n}$  is symmetric and we can assume without loss of generality that  $a_{ii} = 1$  for all  $i$ . For  $n = 2$ ,  $f(x) = 0$  is the global minimum, attained at  $x = [\theta a_{12}, \theta^{-1}]^T$  for any  $\theta \neq 0$ . For  $n = 3$ ,  $f(x) = 0$  is again achieved; if  $a_{ij} \neq 1$  for all  $i$  and  $j$  then there are exactly two minimizers. But for  $n \geq 4$  there are more equations than variables in  $A = \text{diag}(1 - x_i^2) + xx^T$  and so the global minimum is generally positive.

Note that because of Theorem 3.1 we could further restrict one element of  $x$  to  $[0, 1]$ . We could go further and restrict all the elements of  $x$  to  $[0, 1]$  in order to obtain a correlation matrix with nonnegative elements—a constraint that is imposed in [35], [42].

The function  $f$  is clearly twice continuously differentiable and we need to find its gradient  $\nabla f(x)$  and Hessian  $\nabla^2 f(x)$ . Setting  $\hat{A} = A - I$  and  $D = \text{diag}(x_i)$ , noticing that  $\hat{a}_{ii} \equiv 0$ , and using properties of the trace operator, we can rewrite  $f$  as

$$(3.6) \quad \begin{aligned} f(x) &= \langle \hat{A}, \hat{A} \rangle + 2\langle \hat{A}, D^2 \rangle - 2\langle \hat{A}, xx^T \rangle \\ &\quad + \langle xx^T, xx^T \rangle - 2\langle xx^T, D^2 \rangle + \langle D^2, D^2 \rangle \\ &= \langle \hat{A}, \hat{A} \rangle - 2x^T \hat{A} x + (x^T x)^2 - \sum_{i=1}^n x_i^4. \end{aligned}$$

LEMMA 3.5. *For  $f$  in (3.5) we have*

$$(3.7) \quad \nabla f(x) = 4((x^T x)x - \hat{A}x - D^2 x),$$

$$(3.8) \quad \nabla^2 f(x) = 4(2xx^T + (x^T x)I - \hat{A} - 3D^2).$$

*Proof.* We have  $\nabla(x^T \hat{A}x) = 2\hat{A}x$  and  $\nabla^2(x^T \hat{A}x) = 2\hat{A}$ . Similarly,  $\nabla(\sum_{i=1}^n x_i^4) = 4D^2 x$  and  $\nabla^2(\sum_{i=1}^n x_i^4) = 12D^2$ . It is straightforward to show that for  $h(x) = (x^T x)^2$  we have  $\nabla h(x) = 4(x^T x)x$  and  $\nabla^2 h(x) = 8xx^T + 4(x^T x)I$ . The formulae follow by differentiating (3.6) and using these expressions.  $\square$

Notice that at  $x = 0$ ,  $\nabla f(0) = 0$  and  $\nabla^2 f(0) = -4\hat{A}$ . For  $A \neq I$ , since  $\hat{A}$  is symmetric and indefinite (by virtue of its zero diagonal),  $x = 0$  is a saddle point of  $f$ . Another deduction that can be made from the lemma is that if  $a_{ii} = 1$  and  $|a_{ij}| \leq 1$  for all  $i$  and  $j$  then  $x = e$  is a solution if and only if  $A = ee^T$ .

Denote the global minimizer of  $f$  by  $\bar{x}$ . If  $f(\bar{x}) = 0$  then  $A = \text{diag}(1 - \bar{x}_i^2) + \bar{x}\bar{x}^T$  is precisely of the sought structure and we call  $A$  reproducible. We ignore the constraint  $|x| \leq e$  for the rest of this section. We now examine the properties of the Hessian matrix at  $\bar{x}$  for reproducible  $A$  and will later draw conclusions about the



non-reproducible case. Note that (3.8) simplifies to  $\nabla^2 f(\bar{x}) = 4((\bar{x}^T \bar{x})I + \bar{x} \bar{x}^T - 2\bar{D}^2)$ , where  $\bar{D} = \text{diag}(\bar{x}_i)$ . Therefore we consider the matrix

$$(3.9) \quad H_n = H_n(x) = (x^T x)I + x x^T - 2D^2, \quad x \in \mathbb{R}^n.$$

For example,

$$H_4 = \begin{bmatrix} x_2^2 + x_3^2 + x_4^2 & x_1 x_2 & x_1 x_3 & x_1 x_4 \\ x_2 x_1 & x_1^2 + x_3^2 + x_4^2 & x_2 x_3 & x_2 x_4 \\ x_3 x_1 & x_3 x_2 & x_1^2 + x_2^2 + x_4^2 & x_3 x_4 \\ x_4 x_1 & x_4 x_2 & x_4 x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix}.$$

We want to determine the definiteness and nonsingularity properties of  $H_n$ . Without loss of generality we can suppose that

$$(3.10) \quad x_1 \geq x_2 \geq \cdots \geq x_p > x_{p+1} = \cdots = x_n = 0,$$

with  $p \geq 1$ . If  $n = 4$  and  $p = 3$  then  $H_4$  has the form

$$\begin{bmatrix} x_2^2 + x_3^2 & x_1 x_2 & x_1 x_3 & 0 \\ x_2 x_1 & x_1^2 + x_3^2 & x_2 x_3 & 0 \\ x_3 x_1 & x_3 x_2 & x_1^2 + x_2^2 & 0 \\ 0 & 0 & 0 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix} = \text{diag}(H_3, x_1^2 + x_2^2 + x_3^2).$$

In general,

$$H_n = \text{diag}(H_p, D_p), \quad D_p = (x_1^2 + x_2^2 + \cdots + x_p^2)I_{n-p}.$$

$D_p$  has positive diagonal entries and hence the definiteness properties of  $H_n$  are determined by those of  $H_p$ . So the problem has been reduced to the case of positive  $x_i$ .

**THEOREM 3.6.**  *$H_n$  is positive semidefinite. Moreover,  $H_n$  is nonsingular if and only if at least three of  $x_1, x_2, \dots, x_n$  are nonzero.*

*Proof.* From the foregoing analysis we can restrict attention to  $H_p$  and assume that (3.10) holds. Let  $W = \text{diag}(x_1, x_2, \dots, x_p)$ . Then  $\tilde{H}_p = W^T H_p W$  has the form illustrated for  $p = 4$  by

$$\tilde{H}_4 = \begin{bmatrix} x_1^2(x_2^2 + x_3^2 + x_4^2) & x_1^2 x_2^2 & x_1^2 x_3^2 & x_1^2 x_4^2 \\ x_2^2 x_1^2 & x_2^2(x_1^2 + x_3^2 + x_4^2) & x_2^2 x_3^2 & x_2^2 x_4^2 \\ x_3^2 x_1^2 & x_3^2 x_2^2 & x_3^2(x_1^2 + x_2^2 + x_4^2) & x_3^2 x_4^2 \\ x_4^2 x_1^2 & x_4^2 x_2^2 & x_4^2 x_3^2 & x_4^2(x_1^2 + x_2^2 + x_3^2) \end{bmatrix}.$$

Thus  $\tilde{H}_p$  is diagonally dominant with nonnegative diagonal elements and with equality in the diagonal dominance conditions for every row (or column); it is therefore positive semidefinite by Gershgorin's theorem. Suppose  $\tilde{H}_p$  is singular. Then  $\lambda = 0$  is an eigenvalue lying on the boundary of the set of Gershgorin discs (in fact it is on the boundary of every Gershgorin disc). Hence by [18, Thm. 6.2.5], since  $\tilde{H}_p$  has all nonzero entries any null vector  $z$  of  $\tilde{H}_p$  has the property that  $|z_i|$  is the same for all  $i$ . Hence any null vector can be taken to have elements  $z_i = \pm 1$ . But it is easy to see that no such vector can be a null vector of  $\tilde{H}_p$  for  $p > 2$ . Hence  $\tilde{H}_p$  is nonsingular for  $p > 2$ . Since  $H_p$  is congruent to  $\tilde{H}_p$ ,  $H_p$  is positive definite for  $p > 2$ . For  $p = 1, 2$ ,  $H_p$  is singular. The result follows.  $\square$

Since  $\bar{x}$  is, by definition, a global minimizer, and is usually one of exactly two distinct global minimizers  $\pm\bar{x}$ , by Theorem 3.1, Theorem 3.6 does not provide any significant new information about  $\bar{x}$ . However, it does tell us something about the non-reproducible case. For general  $A$ ,  $\widehat{H}_n = \frac{1}{4}\nabla^2 f(x)$  can be written, using (3.8),

$$\widehat{H}_n = ((x^T x)I + xx^T - 2D^2) + (xx^T - \widehat{A} - D^2) = H_n + E_n,$$

where  $H_n$ , defined in (3.9), is positive semidefinite by Theorem 3.6 and moreover positive definite if at least three components of  $x$  are nonzero. Now  $E_n$  has zero diagonal and in general is indefinite. Furthermore,  $E_n$  is singular at a stationary point  $x$  since  $E_n x = 0$  by (3.7). We can conclude that at a stationary point  $x$  having at least three nonzero components the Hessian  $\nabla^2 f(x) = 4\widehat{H}_n$  will be positive definite if  $\|E_n\|$  is sufficiently small, that is, if  $|(E_n)_{ij}| = |x_i x_j - a_{ij}|$  is sufficiently small for all  $i$  and  $j$ . In this case  $x$  is a local minimizer of  $f$ .

**4.  $k$  factor problem.** Now we consider the general  $k$  factor problem, for which  $C(X) = D + \sum_{j=1}^k x_j x_j^T$  as in (1.4). We require that (1.3) holds, so that  $C(X)$  is positive semidefinite and hence is a correlation matrix.

As noted by Lawley and Maxwell [24], the representation (1.4) is far from unique as we can replace  $X$  by  $XQ$  for any orthogonal matrix  $Q \in \mathbb{R}^{k \times k}$  without changing  $C(X)$ . This corresponds to a rotation of the factors in the terminology of factor analysis. Some approaches to determine a unique representation are described in [20], [24]. Probably the most popular one is the varimax method of Kaiser [21]. Given an  $X$  defining a matrix  $C(X)$  with  $k$  factor structure this method maximizes the function

$$V(P) = \left\| \left( I_n - \frac{1}{n} ee^T \right) (XP \circ XP) \right\|_F$$

over all orthogonal  $P$  and then uses the representation  $C(\widehat{X}P)$ . Here the symbol “ $\circ$ ” denotes the Hadamard product ( $A \circ B = (a_{ij}b_{ij})$ ). The method rotates and reflects the rows of  $X$  so that the elements of each column differ maximally from their mean value, which explains the name varimax.

The nearness problem for our  $k$  factor representation is to minimize

$$(4.1) \quad f(X) = \|A - (I + XX^T - \text{diag}(XX^T))\|_F^2$$

over all  $X \in \mathbb{R}^{n \times k}$  satisfying the constraints (1.3). As before,  $A \in \mathbb{R}^{n \times n}$  is symmetric with unit diagonal and we set  $\widehat{A} = A - I$ . We now obtain the first and second derivatives of  $f$ .

Since  $\widehat{A}$  has zero diagonal we have  $\langle \widehat{A}, \text{diag}(XX^T) \rangle = 0$  and also  $\langle \text{diag}(XX^T) - XX^T, \text{diag}(XX^T) \rangle = 0$ . The function  $f$  can therefore be written

$$(4.2) \quad f(X) = \langle \widehat{A}, \widehat{A} \rangle - 2\langle \widehat{A}X, X \rangle + \langle XX^T, XX^T \rangle - \langle XX^T, \text{diag}(XX^T) \rangle.$$

The next result gives a formula for the gradient, which is now most conveniently expressed as the matrix  $\nabla f(x) = (\partial f(x)/\partial x_{ij}) \in \mathbb{R}^{n \times k}$ .

LEMMA 4.1. *For  $f$  in (4.1) we have*

$$(4.3) \quad \nabla f(X) = 4(X(X^T X) - \widehat{A}X - \text{diag}(XX^T)X).$$

*Proof.* It is straightforward to show that  $\nabla\langle\widehat{A}X, X\rangle = 2\widehat{A}X$ . Next, consider the term  $h_1(x) = \langle XX^T, XX^T\rangle$ . Consider the auxiliary function  $g_1 : \mathbb{R} \rightarrow \mathbb{R}$ , given by  $g_1(t) = h_1(X + tZ)$ , for arbitrary  $Z \in \mathbb{R}^{n \times k}$ . Clearly,  $g_1'(0) = \langle \nabla h_1(X), Z \rangle$ . After some algebraic manipulations we find that

$$g_1'(0) = 2\langle X^T X, X^T Z \rangle + 2\langle X^T X, Z^T X \rangle = 4\langle X(X^T X), Z \rangle.$$

Therefore,  $\nabla h_1(X) = 4X(X^T X)$ . Similarly, we find that the gradient of  $h_2(x) = \langle XX^T, \text{diag}(XX^T)\rangle$  is  $\nabla h_2(X) = 4 \text{diag}(XX^T)X$ . The result follows.  $\square$

Notice that when  $k = 1$ , (4.3) reduces to (3.7).

The Hessian of  $f$  is an  $nk \times nk$  matrix that is most conveniently viewed as a matrix representation of the Fréchet derivative of  $\nabla f$ . Recall that the Fréchet derivative of a function  $g : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$  is a linear operator  $L_g$  such that  $g(X + E) = g(X) + L_g(X, E) + o(\|E\|)$  for all  $E \in \mathbb{R}^{m \times n}$  [16, Sec. 3.1]. We can determine the Fréchet derivative of  $\nabla f$  from this definition by finding the linear part of the expansion. For example, to find the derivative of the first term in (4.3) we set  $f_1(X) = X(X^T X)$  and consider

$$f_1(X + E) = f_1(X) + X(X^T E) + X(E^T X) + E(X^T X) + O(\|E\|^2).$$

Hence  $L_{f_1}(X, E) = X(X^T E) + X(E^T X) + E(X^T X)$ . For the third term,  $f_3$ , we have, similarly,  $L_{f_3}(X, E) = \text{diag}(XE^T)X + \text{diag}(EX^T)X + \text{diag}(XX^T)E$ .

LEMMA 4.2. *The Hessian of  $f$  in (4.1) is given by*

$$(4.4) \quad \begin{aligned} L_{\nabla f}(X, E) = & 4(X(X^T E) + X(E^T X) + E(X^T X) - \widehat{A}E \\ & - (\text{diag}(XE^T)X + \text{diag}(EX^T)X + \text{diag}(XX^T)E)). \end{aligned}$$

**5. Numerical methods.** The problem of interest is

$$(5.1a) \quad \text{minimize } f(X) = \|A - (I + XX^T - \text{diag}(XX^T))\|_F^2,$$

$$(5.1b) \quad \text{subject to } X \in \Omega := \{ X \in \mathbb{R}^{n \times k} : \sum_{j=1}^k x_{ij}^2 \leq 1, i = 1:n \},$$

where  $A \in \mathbb{R}^{n \times n}$  is a given symmetric matrix. The set  $\Omega$  is convex. However, since the objective function  $f$  in (5.1a) is nonconvex we can only expect to find a local minimum, though if we achieve  $f(X) = 0$  we know that  $X$  is a global minimizer.

We consider several different numerical methods for solving the problem. We first consider how to start the iterations. Let  $\lambda$  be the largest eigenvalue of  $A$ , which is at least 1 if  $A$  has unit diagonal, which can be assumed without loss of generality. We take for the starting matrix  $X^{(0)}$  a matrix  $\alpha ve^T$  whose columns are all the same multiple of the eigenvector  $v$  corresponding to  $\lambda$ . The scalar  $\alpha$  is chosen to minimize  $f(\alpha ve^T)$  subject to  $\alpha ve^T$  staying in the feasible set  $\Omega$ . Straightforward computations show that the optimal  $\alpha$  is

$$\alpha_{\text{opt}} = \min \left\{ \left( \frac{(\lambda - 1)\|v\|_2^2}{k\|v\|_2^4 - k\sum_i v_i^4} \right)^{1/2}, \frac{1}{k^{1/2} \max_i |v_i|} \right\}.$$

This  $X^{(0)}$  can be inexpensively computed by using the power method or the Lanczos method to obtain  $\lambda$  and  $v$ . Moreover, it is guaranteed to yield a smaller value of  $f$

than the zero matrix if  $\lambda > 1$  since, from (4.2),

$$\begin{aligned}
f(\alpha_{\text{opt}}ve^T) &= \langle \widehat{A}, \widehat{A} \rangle - 2\alpha_{\text{opt}}^2 k(\lambda - 1)\|v\|_2^2 + \alpha_{\text{opt}}^4 k^2 \|v\|_2^4 - \alpha_{\text{opt}}^4 k^2 \sum_i v_i^4 \\
&= \langle \widehat{A}, \widehat{A} \rangle - \alpha_{\text{opt}}^2 k \left( 2(\lambda - 1)\|v\|_2^2 - \alpha_{\text{opt}}^2 \left( k\|v\|_2^4 - k \sum_i v_i^4 \right) \right) \\
&\leq \langle \widehat{A}, \widehat{A} \rangle - \alpha_{\text{opt}}^2 k \left( 2(\lambda - 1)\|v\|_2^2 - (\lambda - 1)\|v\|_2^2 \right) \\
&= \langle \widehat{A}, \widehat{A} \rangle - \alpha_{\text{opt}}^2 k(\lambda - 1)\|v\|_2^2 < f(0).
\end{aligned}$$

As noted by Anderson, Sidenius, and Basu [2], and as we will see later for some problem types, minimizing  $f$  without the constraint  $X \in \Omega$  may yield a solution of the constrained problem (5.1). This motivates us to consider first methods that ignore or only partly incorporate the constraint. The first method is the alternating directions (or coordinate search) method. Regarding  $f$  as a function of just  $x_{ij}$  we have

$$f(x_{ij}) = \text{const.} + 2 \sum_{q \neq i} \left( a_{iq} - \sum_{s=1}^k x_{is}x_{qs} \right)^2,$$

so

$$\begin{aligned}
f'(x_{ij}) &= 4 \sum_{q \neq i} (-x_{qj}) \left( a_{iq} - \sum_{s=1}^k x_{is}x_{qs} \right)^2 \\
&= 4 \left( - \sum_{q \neq i} x_{qj} a_{iq} + \sum_{q \neq i} x_{qj} x_{ij} x_{qj} + x_{qj} \sum_{s \neq j} x_{is}x_{qs} \right) \\
&= 4 \left( x_{ij} \sum_{q \neq i} x_{qj}^2 + \sum_{q \neq i} x_{qj} \left( \sum_{s \neq j} x_{is}x_{qs} - a_{iq} \right) \right).
\end{aligned}$$

Hence  $f'(x_{ij}) = 0$  for

$$(5.2) \quad x_{ij} = \frac{\sum_{q \neq i} x_{qj} \left( a_{iq} - \sum_{s \neq j} x_{is}x_{qs} \right)}{\sum_{q \neq i} x_{qj}^2}.$$

We can therefore repeatedly minimize over each  $x_{ij}$  in turn using (5.2). If the new  $x_{ij}$  is not in the interval  $[-1, 1]$  we project it back onto the interval by reducing  $|x_{ij}|$  appropriately, since  $x_{ij}$  must lie in this interval if it is in  $\Omega$ . Convergence of this method to a stationary point of  $f$  can be proved under suitable conditions [22, Sec. 8.1], [39]. After the projection step  $x$  may nevertheless lie outside  $\Omega$  if  $k > 1$ , but we do not project onto  $\Omega$  because this may cause the method not to converge.

Anderson, Sidenius, and Basu [2] propose another method to solve the  $k$  factor problem. For  $F(X) = I - \text{diag}(XX^T)$  it iteratively generates a sequence  $\{X_i\}_{i \geq 0}$  with

$$(5.3) \quad X_i = \underset{X \in \mathbb{R}^{n \times k}}{\text{argmin}} \|A - F(X_{i-1}) - XX^T\|_F.$$

The minimizer of (5.3) is found by principal component analysis. Let  $P^T \Lambda P$  be a spectral decomposition of  $A - F(X_{i-1})$ , with  $P$  orthogonal and  $\Lambda$  diagonal with diagonal elements in nonincreasing order. Then the minimizer is (in MATLAB notation)

$X_i = P(:, 1:k)\tilde{\Lambda}^{1/2}$ , where  $\tilde{\Lambda} = \text{diag}(\max(\lambda_1, 0), \dots, \max(\lambda_k, 0))$ . Thus just the  $k$  largest eigenvalues and corresponding eigenvectors of  $A - F(X_{i-1})$  are needed, and these can be inexpensively computed by the Lanczos iteration or by orthogonally reducing the matrix to tridiagonal form and applying the bisection method followed by inverse iteration [40, pp. 227 ff.]. This method is also known as the principal factors method [10, Sec. 10.4].

We note that this method is equivalent to the alternating projections method that generates a sequence  $\{Z_i\}_{i \geq 0}$  with  $Z_i = P_S(P_U(Z_{i-1}))$ , where  $P_S$  and  $P_U$  are projection operators onto the sets

$$\begin{aligned} U &:= \{W \in \mathbb{R}^{n \times n} : w_{ij} = a_{ij} \text{ for } i \neq j\}, \\ S &:= \{W \in \mathbb{R}^{n \times n} : W = XX^T \text{ for some } X \in \mathbb{R}^{n \times k}\}. \end{aligned}$$

The projection  $P_S(Z)$  is formed by the construction described in the previous paragraph. With  $X_0 = Z_0$ , the equivalence between the  $\{X_k\}$  and the  $\{Z_k\}$  is given by  $Z_i \equiv X_i X_i^T$ .

Although this method has been successfully used [2], [19] it is not guaranteed to converge. The standard convergence theory [7] for the alternating projections method is not applicable since the set  $S$  is not convex for  $k < n$  and the sets  $U$  and  $S$  do not have a point in common unless the objective function  $f$  is zero at the global minimum.

Since there is no guarantee that the final iterates of the alternating directions and principal factors methods lie in the feasible set  $\Omega$ , we project onto this set after the computation. To project an  $n \times k$  matrix  $Y$  with rows  $y_i^T$  onto  $\Omega$  we simply replace any row  $y_i^T$  such that  $\|y_i\|_2 > 1$  by  $y_i^T / \|y_i\|_2$ . We denote this projection by  $P(Y)$ .

The next method solves the full, constrained problem (5.1) and generates a sequence of matrices that is guaranteed to converge  $r$ -linearly to a stationary point of (5.1). Introduced by Birgin, Martínez, and Raydan [3], [4], the spectral projected gradient method (SPGM) aims to minimize a continuously differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  on a nonempty closed convex set. The method has the form  $x_{k+1} = x_k + \alpha_k d_k$  where  $d_k$  is chosen to be  $P(x_k - t_k \nabla f(x_k)) - x_k$ , with  $t_k > 0$  a pre-computed scalar. The direction  $d_k$  is guaranteed to be descent direction [3, Lem. 2.1] and the scalar  $\alpha_k$  is selected by a nonmonotone line search strategy. The cost per iteration is low for our problem because the projection  $P$  is inexpensive to compute. An R implementation of the method is available [43].

Our analysis in the previous sections suggests applying a Newton method to our problem since the gradient and the Hessian are explicitly known and can be computed in a reasonable time. As the constraints defining  $\Omega$  in (5.1b) are nonlinear for  $k > 1$  we distinguish here between the one factor case and the  $k$  factor case.

For  $k = 1$  we use the routine `e041b` of the NAG Toolbox for MATLAB [30], which implements a globally convergent modified Newton method for minimizing a nonlinear function subject to upper and lower bounds on the variables; these bounds allow us to enforce the constraint (5.1b). This method uses the first derivative and the Hessian matrix.

For  $k > 1$  we apply the routine `e04wd` of the NAG Toolbox for MATLAB, which implements an SQP method. This routine deals with the nonlinear constraints (5.1b) but does not use the Hessian. In order to have an unconstrained optimization method that we can compare with the principal factors method, we apply the function `fminunc` of the MATLAB Optimization Toolbox [27], which implements a subspace trust region method based on the interior-reflective Newton method. This algorithm uses the first

derivative and the Hessian. As for the principal factors method, if necessary we project the final iterate onto the feasible set  $\Omega$  to satisfy the constraints.

**6. Computational experiments.** Our experiments were performed in MATLAB R2007a using the NAG Toolbox for MATLAB Mark 22.0 on an Intel Pentium 4 (3.20 GHz). Each method was used with the same starting matrix, determined as described in Section 5. In order to define the stopping criterion used in all the algorithms we first introduce an easy to compute measurement of stationarity. We define the function  $q : \mathbb{R}^{n \times k} \mapsto \mathbb{R}^{n \times k}$  by

$$q(X) = P(X - \nabla f(X)) - X.$$

It can be shown that a point  $X^* \in \Omega$  is a stationary point of our problem (5.1) if and only if  $q(X^*) = 0$  [8, (2.6)]. The stopping criterion is

$$(6.1) \quad \|q(X)\|_F \leq \text{tol},$$

where  $\text{tol}$  will be specified for the individual tests below. We use the same notation and criterion when no constraints are imposed, in which case  $P$  is the identity and  $q(x)$  reduces to the gradient  $-\nabla f(X)$ .

Since the final iterates of these methods may not be in the feasible set  $\Omega$ , prior to our enforced projection onto it, we introduce a measurement of constraint violation at a point  $X$ , given by the function  $v : \mathbb{R}^{n \times k} \rightarrow \mathbb{R}$  with

$$(6.2) \quad v(X) = \sum_{i=1}^n \max(\|y_i\|_2^2 - 1, 0), \quad X^T = [y_1, \dots, y_n].$$

Our test matrices are chosen from four classes.

- **corrand**: a random correlation matrix generated by `gallery('randcorr', n)`.
- **corkfac**: a random correlation matrix generated by  $A = \text{diag}(I - XX^T) + XX^T$  where  $X \in \mathbb{R}^{n \times k}$  is a random matrix with elements from the uniform distribution on  $[-1, 1]$  that is then projected onto  $\Omega$ . Here the objective function  $f$  is zero at the global minimum.
- **randneig**: A symmetric matrix generated by  $A = \frac{1}{2}(B + B^T) + \text{diag}(I - B)$  where  $B$  is the first matrix out of a sequence of random matrices with elements from the uniform distribution on  $[-1, 1]$  such that  $A$  has a negative eigenvalue.
- **cor1399**: A symmetric, unit-diagonal matrix constructed from stock data provided by a fund management company. It has dimension  $n = 1399$  and is highly rank-deficient but not positive semidefinite. This matrix was also used in [5], [15].

**6.1. Test results for  $k = 1$ .** We first consider the one factor case. In Tables 6.1 and 6.2 we report results averaged over 10 instances of each of the three classes of random matrices for  $n = 100$  and  $n = 2000$  with tolerance  $\text{tol} = 10^{-3}$  and  $\text{tol} = 10^{-6}$ , respectively. Table 6.3 gives the results for the matrix `cor1399` with tolerance  $\text{tol} = 10^{-3}$ . The abbreviations are

- AD: alternating directions method.
- PFM: principal factors method.
- `fminu`: `fminunc`.
- SQP: `e04wd`.
- t: mean computational time (seconds).

TABLE 6.1  
Results for the random one factor problems with  $\text{tol} = 10^{-3}$ .

	t	it	it <sub>sd</sub>	dist	nq	v	t	it	it <sub>sd</sub>	dist	nq	v
	$n = 100$						$n = 2000$					
	corrاند, dist <sub>0</sub> =5.6646, nq <sub>0</sub> =8e-2						corrاند, dist <sub>0</sub> =26.006, nq <sub>0</sub> =5e-3					
AD	0.22	110	78	5.6642	9e-4	0	3.3	5.2	1.5	26.006	9e-4	0
PFM	0.09	10	5.4	5.6642	8e-4	0	68	1.1	0.2	26.006	2e-4	0
e041b	0.02	4.7	2.4	5.6643	3e-4	0	23	1.8	0.4	26.006	6e-4	0
SPGM	0.11	57	29	5.6642	6e-4	0	9.8	5.2	0.8	26.006	8e-4	0
	corkfac, dist <sub>0</sub> =0.3697, nq <sub>0</sub> =6e0						corkfac, dist <sub>0</sub> =0.3718, nq <sub>0</sub> =3e1					
AD	0.01	5.0	0.6	2.25e-5	4e-4	0	3.1	5.2	0.6	5.06e-6	4e-4	0
PFM	0.03	3.0	0	4.03e-5	6e-4	0	15	2.2	0.3	1.56e-6	1e-4	0
e041b	0.01	2.0	0	1.45e-7	3e-6	0	16	2.0	0	1.5e-11	1e-9	0
SPGM	0.02	6.0	1.2	2.67e-5	3e-4	0	11	4.6	0.9	7.72e-6	4e-4	0
	randneig, dist <sub>0</sub> =43.606, nq <sub>0</sub> =6e2						randneig, dist <sub>0</sub> =824.13, nq <sub>0</sub> =2e4					
AD	0.01	5.9	0.3	40.398	3e-4	0	3.8	7.2	1.3	815.79	5e-4	0
PFM	0.03	3	0.2	40.418	6e-4	3	22	3.0	0	815.81	2e-6	15
e041b	0.16	61.9	5.2	40.398	1e-4	0	4167	1222	22	815.79	2e-6	0
SPGM	0.02	6.0	0.0	40.398	5e-4	0	9.4	7.2	0.4	815.79	2e-4	0

TABLE 6.2  
Results for the random one factor problems with  $\text{tol} = 10^{-6}$ .

	t	it	it <sub>sd</sub>	dist	nq	v	t	it	it <sub>sd</sub>	dist	nq	v
	$n = 100$						$n = 2000$					
	corrاند, dist <sub>0</sub> =5.6646, nq <sub>0</sub> =8e-2						corrاند, dist <sub>0</sub> =26.006, nq <sub>0</sub> =5e-3					
AD	0.72	393	188	5.6642	9e-7	0	3938	7282	1653	26.006	9e-7	0
PFM	0.32	31	13	5.6642	8e-7	0	827	18	5.4	26.006	8e-7	0
e041b	0.02	7.2	2.5	5.6643	2e-8	0	36	5.0	1.6	26.006	6e-7	0
SPGM	0.22	128	44	5.6642	6e-7	0	638	760	546	26.006	8e-7	0
	corkfac, dist <sub>0</sub> =0.3632, nq <sub>0</sub> =6e0						corkfac, dist <sub>0</sub> =0.3718, nq <sub>0</sub> =3e1					
AD	0.02	9.8	0.5	2.73e-8	4e-7	0	6.1	9.2	2.4	8.73e-9	7e-7	0
PFM	0.06	5.6	0.5	3.19e-8	4e-7	0	21	3.2	0.4	3.91e-9	3e-7	0
e041b	0.01	3.0	0	1.8e-14	4e-13	0	15	2.0	0	1.5e-11	1e-9	0
SPGM	0.03	9.9	2.0	1.97e-8	2e-7	0	13	8.2	2.4	6.88e-9	3e-7	0
	randneig, dist <sub>0</sub> =43.606, nq <sub>0</sub> =6e2						randneig, dist <sub>0</sub> =824.13, nq <sub>0</sub> =2e4					
AD	0.02	8.6	0.5	40.398	4e-7	0	3.4	10.0	0	815.79	3e-7	0
PFM	0.06	5.0	0	40.418	2e-7	3	19.0	4.0	0	815.81	1e-9	15
e041b	0.09	61	5.7	40.398	1e-7	0	4171	1222	22	815.79	2e-6	0
SPGM	0.02	9.0	0	40.398	1e-7	0	11	9.6	0.5	815.79	2e-7	0

- it: mean number of iterations.
- it<sub>sd</sub>: standard deviation of the number of iterations.
- dist<sub>0</sub>: mean initial value of  $f(X)^{1/2}$ .
- dist: mean final value of  $f(X)^{1/2}$  after the final projection onto the feasible set.
- nq<sub>0</sub>: mean initial value of  $\|q(X)\|_F$ .
- nq: mean final value of  $\|q(X)\|_F$  before the final projection onto the feasible set.
- v: mean final value of  $v(X)$  before the final projection onto the feasible set.

For the method AD one iteration is defined to be a sweep over which the objective function  $f$  is minimized over each coordinate direction in turn.

Several comments can be made on Tables 6.1–6.3.

1. The values of  $v$  in (6.2) are all zero except for PFM on the randneig matrices, where the final projection onto  $\Omega$  causes dist for the accepted  $X$  to exceed that for the other methods. Except in these cases the mean function values of the final iterates

TABLE 6.3  
*Results for the one factor problem for cor1399, with tol = 10<sup>-3</sup> and tol = 10<sup>-6</sup>.*

	t	it	dist	nq	v	t	it	dist	nq	v
	tol = 10 <sup>-3</sup>					tol = 10 <sup>-6</sup>				
	cor1399, dist <sub>0</sub> =118.7753, nq <sub>0</sub> =9e0									
AD	1.08	6.0	118.7752	2e-4	0	1.80	10.0	118.7752	5e-7	0
PFM	0.96	2.0	118.7752	6e-5	0	1.31	3.0	118.7752	2e-7	0
e041b	8.16	2.0	118.7752	5e-10	0	8.16	2.0	118.7752	5e-10	0
SPGM	4.83	7.0	118.7752	2e-5	0	5.67	10.0	118.7752	9e-7	0

of the methods do not differ significantly. In particular, for the corkfac matrices the sequences appear to approach the global minimum. Except for the randneig problems all the constraints are inactive at the computed final iterates, so by Theorem 3.4 the matrices  $C(X)$  have full rank. For these problems about half of the constraints are inactive and this number is slightly bigger for the matrix returned by PFM than for the other methods.

2. None of the methods always outperforms the others in computational time. The relative performance of the individual methods depends on the tolerance, the problem size and the problem type. AD performs very well for  $\text{tol} = 10^{-3}$  but is the least efficient method for the corrand matrices with  $\text{tol} = 10^{-6}$ . Turning to the problem size, for the small tolerance an increased  $n$  gives a bigger time advantage of AD over the other two methods, which is due to the remarkably low number of approximately  $4n^2$  operations taken by AD for each iteration, compared with the Newton method e041b, which requires  $O(n^3)$  operations. Finally, the efficiency of the methods depends on the matrix type as can be seen for  $n = 2000$  in Table 6.2, where in execution time the first three methods rank exactly in the reverse order for the corrand matrices compared with the randneig matrices. For the latter matrices, many steps appear to be required to approach the region of quadratic convergence for the Newton method.

3. Interestingly, PFM, for which we do not have a convergence guarantee, shows robust behavior in terms of the required number of iterations and is clearly the best method on the cor1399 matrix. It satisfies the stopping criterion in these tests in a few iterations for every problem instance. However, we found that for small problem sizes PFM can show very poor convergence, as illustrated by the matrix

$$(6.3) \quad A = \begin{bmatrix} 1.0000 & 1.0669 & -1.0604 & 0.4903 & 0.9747 \\ 1.0669 & 1.0000 & 3.2777 & 0.3914 & 1.0883 \\ -1.0604 & 3.2777 & 1.0000 & 1.1075 & 0.8823 \\ 0.4903 & 0.3914 & 1.1075 & 1.0000 & 1.0431 \\ 0.9747 & 1.0883 & 0.8823 & 1.0431 & 1.0000 \end{bmatrix}.$$

For the corresponding two factor problem PFM requires 11,415,465 iterations to satisfy the stopping criterion (6.1) with  $\text{tol} = 10^{-3}$ . This matrix was found after just 22 function evaluations using the implementation `mdsmax` [14] of the multidirectional search method of Torczon [38] to maximize the number of iterations required by PFM. This is in contrast to maximizing the iterations taken by SPGM for a two factor problem with the same problem size, yielding after 2000 function evaluations in `mdsmax` a matrix requiring only 118 iterations.

**6.2. Test results for  $k > 1$ .** We now repeat the same tests for the  $k$  factor case. We averaged the results over 10 instances of each of the three classes of random



TABLE 6.4  
Results for the random  $k$  factor problems with  $tol = 10^{-3}$ .

	t	it	it <sub>sd</sub>	dist	nq	v	t	it	it <sub>sd</sub>	dist	nq	v
	$k = 2$						$k = 6$					
	<b>corrاند</b> , dist <sub>0</sub> =18.52, nq <sub>0</sub> =9.572e-3											
AD	148	624	311	18.49	9e-4	0	704	836	463	18.38	9e-4	0
PFM	9.6	2.4	0.5	18.49	7e-4	0	10	4.1	1.4	18.38	6e-4	0
fmnu	76	34	8.6	18.49	5e-4	0	167	52	16	18.38	5e-4	0
SPGM	19	188	62	18.49	8e-4	0	24	235	62.2	18.38	8e-4	0
SQP	9e3	270	145	18.49	5e-4	0	2e5	485	316	18.38	6e-4	0
	<b>corkfac</b> , dist <sub>0</sub> =161.9, nq <sub>0</sub> =7.76						<b>corkfac</b> , dist <sub>0</sub> =120.3, nq <sub>0</sub> =2.83					
AD	3.1	13.4	0.5	1.3e-5	5e-4	0	11	13.0	0.5	2.1e-5	4e-4	0
PFM	1.1	3.3	0.5	1.9e-5	6e-4	0	1.7	4.0	0.0	3.6e-6	7e-5	0
fmnu	5.0	8.0	0.7	1.1e-5	4e-4	0	10	14.4	1.3	1.1e-5	2e-4	0
SPGM	1.5	13.1	1.0	1.6e-5	5e-4	0	2.8	22.3	2.1	1.6e-5	3e-4	0
SQP	6e3	69	52	1.2e-5	4e-4	0	5e4	222	224	2.4e-5	3e-4	0
	<b>randneig</b> , dist <sub>0</sub> =415, nq <sub>0</sub> =1.31											
AD	1e4	4e4	2e4	408.6	9e-4	30	8e3	9816	3e3	421.0	9e-4	127
PFM	7.3	5.2	0.4	408.7	4e-4	31	10	6.0	1.0	421.0	5e-4	127
fmnu	75	27.8	5.8	408.7	4e-4	31	146	40.8	6.6	421.0	4e-4	127
SPGM	350	2447	1175	407.5	8e-4	0	122	1263	383	407.2	9e-4	0
SQP	3e3	1801	726	407.5	4e-4	1e-11	6e3	262	47	407.2	6e-4	1e-6

TABLE 6.5  
Results for the random  $k$  factor problems with  $tol = 10^{-6}$ .

	t	it	it <sub>sd</sub>	dist	nq	v	t	it	it <sub>sd</sub>	dist	nq	v
	$k = 2$						$k = 6$					
	<b>corrاند</b> , dist <sub>0</sub> =18.525, nq <sub>0</sub> =9.5725e-3											
AD	1e3	4212	4e3	18.4	9e-7	0	5e3	5955	4e3	18.3	9e-7	0
PFM	63	17.8	15.6	18.4	7e-7	0	95	28.1	14	18.3	8e-7	0
fmnu	124	47.3	11.9	18.4	5e-7	0	280	68.2	28	18.3	3e-7	0
SPGM	74	646	627	18.4	7e-7	0	108	892	574	18.3	8e-7	0
SQP	2e4	594	519	18.4	5e-7	0	3e5	641	305	18.4	4e-7	0
	<b>corkfac</b> , dist <sub>0</sub> =161.9, nq <sub>0</sub> =7.76						<b>corkfac</b> , dist <sub>0</sub> =120.3, nq <sub>0</sub> =2.83					
AD	4.2	17.8	0.6	1.6e-8	5e-7	0	15	17.8	0.6	2.2e-8	4e-7	0
PFM	1.6	5.0	0.0	2.3e-9	8e-8	0	2.4	6.0	0.0	1.9e-9	3e-8	0
fmnu	6.2	10	0.7	9.5e-9	3e-7	0	12	16.7	1.2	1.2e-8	2e-7	0
SPGM	1.9	17.4	1.4	1.1e-8	4e-7	0	3.1	26.8	2.5	1.8e-8	4e-7	0
SQP	7e3	75	52	1.8e-8	3e-7	0	5e4	275	278	2.1e-8	0	0
	<b>randneig</b> , dist <sub>0</sub> =415, nq <sub>0</sub> =1.31											
AD	2e4	7e4	4e4	408.6	9e-7	30	1e4	1.2e4	8e3	421	9e-7	127
PFM	9.4	10	2.2	408.7	5e-7	31	11	10	2.1	421	9e-7	127
fmnu	74	46	6.2	408.7	4e-7	31	109	56	3.5	421	1e-7	127
SPGM	477	3081	1e3	407.5	8e-7	0	276	2925	721	407.2	9e-7	0
SQP	5e3	2373	877	407.5	5e-7	1e-11	8e3	447	90	407.2	3e-7	1e-11

matrices for  $n = 1000$  and  $k = 2, 6$  and report the results in Table 6.4 for  $tol = 10^{-3}$  and Table 6.5 for  $tol = 10^{-6}$ . We make some comments on the results.

1. As for  $k = 1$ , the values of  $v$  in (6.2) are all zero except for the randneig problems where these values for the methods disregarding the constraints (1.3) are significantly greater than the convergence tolerance. The values of  $v$  for randneig indicate that forcing the values of  $x$  to lie in the interval  $[-1, 1]$  in AD does not remedy the infeasibility. Moreover, the methods AD, PFM, and fmnu, which all disregard the constraints (1.3), return a final iterate for  $k = 6$  and randneig for which the mean function value is noticeably greater than the mean *initial* function value, caused by

the projection onto the feasible set  $\Omega$  at the end of the computation. This represents a serious failure of the minimization and shows the importance of properly treating the constraints within the method for the randneig problems.

2. In most cases for an increased value of  $k$  more iterations and more computational time are required to compute a solution, presumably due to the increased number of variables. However, interestingly this is not always the case, as shown by the results for AD, PFM, and SPGM on the randneig problems.

3. Only two methods correctly solve the problem in all cases: SPGM and SQP. Since SQP is very slow to converge in all cases, SPGM is clearly the preferred method in terms of efficiency combined with reliability.

**7. Conclusions.** We have obtained new theoretical understanding of the factor-structured nearest correlation matrix problem, particularly through explicit results for the one parameter and one factor cases. Our original motivation for studying this problem came from the credit basket securities application in [2] and the knowledge that the principal factors method has been used in the finance industry, despite the fact that it ignores the nonlinear problem constraints (5.1b). Our experiments have shown that this method, along with alternating directions and `fminunc`, often performs surprisingly well—partly because the constraints are often not active at the solution. However, all three methods can fail to solve the problem, as the randneig matrices show. Moreover, the principal factors method is not supported by any convergence theory. Our conclusion is that the spectral projected gradient method is the method of choice. It has guaranteed convergence, benefits from the ease with which iterates can be projected onto the convex constraint set, and because of the nonmonotone line search strategy can avoid narrow valleys at the beginning of the convergence process.

#### REFERENCES

- [1] Hansjörg Albrecher, Sophie Ladoucette, and Wim Schoutens. A generic one-factor Lévy model for pricing synthetic CDOs. In *Advances in Mathematical Finance*, Michael C. Fu, Robert A. Jarrow, Ju-Yi J. Yen, and Robert J. Elliott, editors, Applied and Numerical Harmonic Analysis, Birkhäuser, Boston, MA, USA, 2007, pages 259–277.
- [2] Leif Anderson, Jakob Sidenius, and Susanta Basu. All your hedges in one basket. *Risk*, pages 67–72, November 2003. [www.risk.net](http://www.risk.net).
- [3] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10(4):1196–1211, 2000.
- [4] Ernesto G. Birgin, José Mario Martínez, and Marcos Raydan. Algorithm 813: SPG—Software for convex-constrained optimization. *ACM Trans. Math. Software*, 27(3):340–349, 2001.
- [5] Rüdiger Borsdorf and Nicholas J. Higham. A preconditioned Newton algorithm for the nearest correlation matrix. MIMS EPrint 2008.50, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, April 2008. 13 pp. Revised November 2008. To appear in *IMA J. Numer. Anal.* Advance Access published on June 8, 2009, DOI 10.1093/imanum/drn085.
- [6] Michel Crouhy, Dan Galai, and Robert Mark. A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24:59–117, 2000.
- [7] Frank Deutsch. *Best Approximation in Inner Product Spaces*. Springer-Verlag, New York, 2001. xv+338 pp. ISBN 0-387-95156-3.
- [8] J. C. Dunn. Global and asymptotic convergence rate estimates for a class of projected gradient processes. *SIAM J. Control Optim.*, 19(3):368–400, 1981.
- [9] Christopher C. Finger. A methodology to stress correlations. *RiskMetrics Monitor*, Fourth Quarter:3–11, 1997.
- [10] James E. Gentle. *Elements of Computational Statistics*. Springer-Verlag, New York, 2002. xv+420 pp. ISBN 0-387-95489-9.

- [11] Paul Glasserman and Sira Suchintabandit. Correlation expansions for CDO pricing. *Journal of Banking & Finance*, 31:1375–1398, 2007.
- [12] Jon Gregory and Jean-Paul Laurent. In the core of correlation. *Risk Magazine*, 17(10):87–91, 2004.
- [13] Igor Grubišić and Raoul Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.*, 422:629–653, 2007.
- [14] Nicholas J. Higham. The Matrix Computation Toolbox. <http://www.ma.man.ac.uk/~higham/mctoolbox>.
- [15] Nicholas J. Higham. Computing the nearest correlation matrix—A problem from finance. *IMA J. Numer. Anal.*, 22(3):329–343, 2002.
- [16] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008. xx+425 pp. ISBN 978-0-898716-46-7.
- [17] Jimmy E. Hilliard and Susan D. Jordan. Measuring risk in fixed payment securities: An empirical test of the structured full rank covariance matrix. *Journal of Financial and Quantitative Analysis*, 26(3):345–362, 1991.
- [18] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985. xiii+561 pp. ISBN 0-521-30586-1.
- [19] Peter Jäckel. Splitting the core. Working Paper, ABN AMRO, London, 2005.
- [20] I. T. Jolliffe. *Principal Component Analysis*. Second edition, Springer-Verlag, New York, 2002. xxix+487 pp. ISBN 0-387-95442-2.
- [21] Henry F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.
- [22] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45(3):385–482, September 2003.
- [23] Samuel Kotz, W. L. Pearn, and Dean W. Wichern. Eigenvalue-eigenvector analysis for a class of patterned correlation matrices with an application. *Statistics and Probability Letters*, 2: 119–125, 1984.
- [24] D. N. Lawley and A. E. Maxwell. Factor analysis as a statistical method. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 12(3):209–229, 1962.
- [25] F. Lillo and R. N. Mantegna. Spectral density of the correlation matrix of factor models: A random matrix theory approach. *Phy. Rev. E*, page 016219, 2005.
- [26] David G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969. xvii+326 pp. ISBN 0-471-55359-X.
- [27] *Optimization Toolbox 4 User's Guide*. The MathWorks, Inc., Natick, MA, USA, 2009. Online version.
- [28] A. E. Maxwell. Factor analysis. In *Encyclopedia of Statistical Sciences*, Samuel Kotz, Campbell B. Read, N. Balakrishnan, and Brani Vidakovic, editors, Wiley, New York, 2006. Electronic.
- [29] Massimo Morini and Nick Webber. An EZI method to reduce the rank of a correlation matrix in financial modelling. *Applied Mathematical Finance*, 13(4):309–331, 2006.
- [30] *NAG Toolbox for MATLAB*. NAG Ltd., Oxford. <http://www.nag.co.uk/>.
- [31] Hou-Duo Qi and Defeng Sun. A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.*, 28(2):360–385, 2006.
- [32] Hou-Duo Qi and Defeng Sun. Correlation stress testing for Value-at-Risk: An unconstrained convex optimization approach. *Computational Optimization and Applications*, 2009. Online First published on January 13, 2009, DOI 10.1007/s10589-008-9231-4.
- [33] Hou-Duo Qi, Zhonghang Xia, and Guangming Xing. An application of the nearest correlation matrix on web document classification. *Journal of Industrial and Management Optimization*, 3(4):701–713, 2007.
- [34] S. N. Roy, B. G. Greenberg, and A. E. Sarhan. Evaluation of determinants, characteristic equations and their roots for a class of patterned matrices. *J. Roy. Statist. Soc. Ser. B*, 22(2):348–359, 1960.
- [35] P. Sonneveld, J. J. I. M. van Kan, X. Huang, and C. W. Oosterlee. Nonnegative matrix factorization of a correlation matrix. *Linear Algebra Appl.*, 431:334–349, 2009.
- [36] Ted H. Szatrowski. Patterned covariances. In *Encyclopedia of Statistical Sciences*, Samuel Kotz, Campbell B. Read, N. Balakrishnan, and Brani Vidakovic, editors, Wiley, New York, 2006. Electronic.
- [37] Alexander Tchernitser and Dmitri H. Rubisov. Robust estimation of historical volatility and correlations in risk management. *Quantitative Finance*, 9:43–54, 2009.
- [38] Virginia J. Torczon. On the convergence of the multidirectional search algorithm. *SIAM J.*

- Optim.*, 1(1):123–145, 1991.
- [39] Virginia J. Torczon. On the convergence of pattern search algorithms. *SIAM J. Optim.*, 7(1): 1–25, 1997.
- [40] Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997. xii+361 pp. ISBN 0-89871-361-7.
- [41] R. Vandebril, M. Van Barel, G. H. Golub, and N. Mastronardi. A bibliography on semiseparable matrices. *Calcolo*, 42:249–70, 2005.
- [42] Antoine Vandendorpe, Ngoc-Diep Ho, Steven Vanduffel, and Paul Van Dooren. On the parameterization of the CreditRisk<sup>+</sup> model for estimating credit portfolio risk. *Insurance: Mathematics and Economics*, 42(2):736–745, 2008.
- [43] Ravi Varadhan and Paul D. Gilbert. BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software*, 32(4):1–26, 2009.
- [44] Q. J. Wang, D. E. Robertson, and F. H. S. Chiew. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, 45:W05407, 2009.
- [45] Tom Wansbeek. Eigenvalue-eigenvector analysis for a class of patterned correlation matrices with an application: A comment. *Statistics and Probability Letters*, 3:95–96, 1985.
- [46] Zhenyue Zhang and Lixin Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra Appl.*, 364:161–187, 2003.