Nonnegative Matrix Factorization of a Correlation Matrix

P. Sonneveld[1], J.J.I.M. van Kan[1], X. Huang[1,*], C.W. Oosterlee[2,1,†]

[1]Delft University of Technology, Numerical Analysis Group, Delft, the Netherlands,
[2]CWI, Centrum Wiskunde & Informatica, Amsterdam, the Netherlands.

June 17, 2008

Abstract

We present a dedicated algorithm for the nonnegative factorization of a correlation matrix from an application in financial engineering. We look for a low-rank approximation. The origin of the problem is discussed in some detail. Next to the description of the algorithm, we prove, by means of a counter example, that an exact nonnegative decomposition of a general positive semidefinite matrix is not always available.

1 Introduction

Nonnegative Matrix Factorization (NMF) algorithms aim to find for a matrix $\boldsymbol{V}$ two matrix factors such that $\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}$, where $\boldsymbol{W}$ and $\boldsymbol{H}$ are both nonnegative matrices, i.e., all elements of $\boldsymbol{W}$ and $\boldsymbol{H}$ are equal to or greater than zero. The non-negativity constraint arises often naturally in applications in physics and engineering.
The notion of Nonnegative Matrix Factorization originates from [9], where simple multiplicative update rules were introduced to solve the approximation problem. Since then, different aspects of NMF, such as its analysis or the extension of the algorithms to various applications have been extensively investigated. For a recent review see [4].
In this article we work with the Nonnegative Matrix Factorization of a correlation matrix, originating from a financial problem. The correlation matrix of n random variables $\Phi_i, i = 1, \ldots, n$ is a $n \times n$ matrix whose (i,j)-th entry reads

$$corr(\Phi_i, \Phi_j) = \frac{E(\Phi_i\Phi_j) - E(\Phi_i)E(\Phi_j)}{\sqrt{[E(\Phi_i^2) - E^2(\Phi_i)][E(\Phi_j^2) - E^2(\Phi_j)]}},$$

*This author wants to thank Group Risk Management, Rabobank, the Netherlands for financial support,

†Corresponding author, c.w.oosterlee@cwi.nl.

where $E(\cdot)$ denotes the expectation operator. A correlation matrix, $\boldsymbol{C}$, is a symmetric positive-semidefinite matrix which has a unit main diagonal. We wish to find a decomposition $\boldsymbol{C} \approx \boldsymbol{A}\boldsymbol{A}^T$ such that $\boldsymbol{A}$ is nonnegative. Moreover, $\boldsymbol{A}\boldsymbol{A}^T$ should be a low-rank approximation of $\boldsymbol{C}$.

Viewed from a different perspective our problem of interest is essentially finding the nearest low-rank correlation matrix, but with an additional non-negativity constraint. The nearest low-rank correlation matrix question has drawn broad attention in the financial community. A variety of methods have been proposed. To name a few, the geometric programming approach [3], the Lagrange multiplier method [16, 14], and majorization [11] have been introduced for this purpose. However, none of these approaches accommodate an additional non-negativity constraint.

This paper is organized as follows. In Section 2, we discuss the application from financial engineering in which the nonnegative matrix factorization problem arises. In Section 3 we present a counter example for the strict equality $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$. In Section 4 we propose an efficient dedicated algorithm for obtaining a nonnegative matrix, $\boldsymbol{A}$, which aims at minimizing the Frobenius norm of the matrix $\boldsymbol{C} - \boldsymbol{A}\boldsymbol{A}^T$, subject to $\boldsymbol{A}_{ij} \geq 0$ and $(\boldsymbol{A}\boldsymbol{A}^T)_{ii} = 1$. Numerical experiments are presented in Section 5.

2 Application Background

2.1 A factor model for credit portfolio losses

Consider a credit portfolio consisting of n obligors with exposure w_i, $i = 1, \ldots, n$. Assume that obligor i defaults if its standardized log asset value X_i is less than some default threshold η_i after a fixed time horizon. The event of default can be modeled as a Bernoulli random variable $D_i = 1_{\{X_i < \eta_i\}}$ with known default probability $p_i = P(X_i < \eta_i)$. It follows that the loss L_i due to obligor i is simply $w_i D_i$ and the portfolio loss is given by

$$L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} w_i D_i. \tag{1}$$

The default indicators D_i and D_j are rarely independent and thus the modeling of the default dependence among obligors is essentially the key issue in portfolio credit loss modeling. Direct modeling of the pairwise correlations is impractical since a bank's credit portfolio can easily contain tens of thousands of obligors. Common practice to reduce the computational complexity is to utilize a so-called *latent factor model* of asset correlations. Popular industrial tools for managing credit portfolios such as KMV's Portfolio Manager [8] and JP Morgan/RiskMetrics Group's CreditMetrics [5] are models of this type. For a summary of the models see [2]. We write for obligor i,

$$X_i = \gamma_i^1 \Psi_1 + \cdots + \gamma_i^M \Psi_M + \epsilon_i = \boldsymbol{\gamma}_i \boldsymbol{\Psi} + \epsilon_i. \tag{2}$$

Here

- $\gamma_i^j \geq 0$ for all i, j. The non-negativity of γ_i^j guarantees that larger values of the factors Ψ_i, ceteris paribus, lead to a smaller number of defaults.
- $\Psi_1, \ldots, \Psi_M$ represent sector (industry, geographic region) indices that are correlated with a known correlation matrix $\boldsymbol{C}$. Since $\boldsymbol{C}$ is a correlation matrix, it has the properties of positive semi-definiteness and a unit main diagonal. The matrix $\boldsymbol{C}$ is further assumed to have only nonnegative entries, which is often justified by empirical evidence.
- ϵ_i denotes an idiosyncratic factor that only affects an obligor itself.
- $\boldsymbol{\Psi}$ and ϵ_i are assumed to be independent for all i.

It follows that, conditional on $\boldsymbol{\Psi}$, D_i and D_j are *independent* and $L(\boldsymbol{\Psi}) = \sum w_i D_i(\boldsymbol{\Psi})$ becomes a weighted sum of independent Bernoulli random variables.
Under such a latent factor model (2), the tail probability of the portfolio loss can be formulated as

$$P(L > x) = \int P(L > x \,|\, \boldsymbol{\Psi})\, dP(\boldsymbol{\Psi}), \tag{3}$$

where $P(\cdot)$ denotes the distribution function. The calculation of the above probability is essential for the determination of the portfolio Value at Risk (VaR), which is defined to be the ζ-quantile of the loss distribution of L for some confidence level, ζ, very close to 1, i.e.,

$$\mathrm{VaR}_\zeta = \inf\{x : P(L \leq x) \geq \zeta\}.$$

The VaR amounts to the capital a bank needs to reserve to stay solvent for probability ζ. Note that the non-negativity of correlation matrix $\boldsymbol{C}$ is a conservative argument in the perspective of risk management as it precludes negative linear relationship between the common factors. As all the factors tend to move in the same direction, extremely adverse scenarios leading to huge losses are more likely.
The integrand $P(L > x \,|\, \boldsymbol{\Psi})$ can be approximated with ease. Various approximations exist and prove to work very well, for example, the recursive method due to [1], the normal approximation method as in [10] or the saddlepoint approximation presented in [7]. Consequently the calculation of the tail probability becomes a high-dimensional numerical integration problem (3), which turns out to be non-trivial.

2.2 Call for a NMF

The number of industrial and country indices in (2) can be quite large. KMV's correlation model, for example, according to [15], identifies "more than 40 countries and 61 industries", whereas CreditMetrics covers "152 country-industry indices, 28 country indices, 19 worldwide industry indices, and 6 regional indices". It is evident that the integral to be solved in (3) can be truly a high

dimensional problem. Monte Carlo (MC) simulation and quasi-Monte Carlo (QMC) methods, which do not suffer from the curse of dimensionality, are the prevailing methods used to solve these multi-dimensional integration problems in finance. However, the event $\{L > x\}$ in (3) becomes a rare event for high loss levels x, that are often the most interesting ones in practice. In this regard both MC and QMC methods can be rather inefficient. Furthermore, the indices Ψ_i and Ψ_j are correlated, which also leads to more complexity for the numerical integration. A version of factor model with *orthogonal* indices is therefore preferred, which means that we are led to a *second level factor model* as follows,

$$\Psi_i = a_i^1 Y_1 + \cdots + a_i^m Y_m + \delta_i = \boldsymbol{a}_i \mathbf{Y} + \delta_i, \tag{4}$$

where Y_i and Y_j are independent and $m \leq M$ (preferably $m \ll M$). Such a decomposition is usually achieved by a principal components analysis (PCA). Note that from (4) we have

$$corr(\Psi_i, \Psi_i) = \boldsymbol{a}_i (\boldsymbol{a}_i)^T. \tag{5}$$

Since $corr(\Psi_i, \Psi_i)$ must be one, we have the hard constraint that $\boldsymbol{a}_i(\boldsymbol{a}_i)^T = 1$. This cannot be expected by employing a straightforward PCA. As a result a different matrix decomposition, which takes care of (5), needs to be developed. In addition, non-negativity of the factor loadings a_i^j is highly desirable. This ensures that the two-level factor model, combining (2) and (4),

$$\mathbf{X} = \boldsymbol{\Gamma}\boldsymbol{\Psi} = \boldsymbol{\Gamma}\mathbf{A}\mathbf{Y}$$

also has only nonnegative coefficients.
The non-negativity requirement is beneficial since it facilitates the computation of the tail probability of portfolio loss (3). This can be demonstrated by the following proposition and example.

Proposition 1. *The function*

$$f(y_1, y_2, ..., y_m) = P(L > x | Y_1 = y_1, Y_2 = y_2, ..., Y_m = y_m),$$

is non-increasing in all its variables y_k.

Proof. see [6]. □

Take, as an example, a homogeneous portfolio in a Gaussian one-factor model, meaning $m = 1$, consisting of 1000 obligors with $w_i = 1$, $p_i = 0.0033$ and $a_i^1 = \sqrt{0.2}$, $i = 1, \ldots, 1000$. We truncate the domain of the common factor Y to the interval $[-5, 5]$ so that the probability of Y falling out of this interval is merely 5.7×10^{-7}. The integrand $P(L > 100|Y)$ is presented in Figure 1. It is indeed a non-increasing function of Y. Furthermore, it decreases rapidly from its upper bound, 1, to its lower bound, 0, for Y in a narrow band (between the two dashed lines in Figure 1), much smaller than the whole domain of Y.
Generally the integral can be computed by a quadrature rule. However since in our problem the integrand is *monotone* and *bounded* in $[0, 1]$, significantly fewer

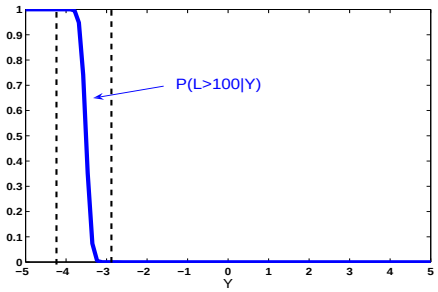


Figure 1: The integrand $P(L > 100|Y)$ as a function of the common factor Y for a portfolio consisting of 1000 obligors with $w_i = 1$, $p_i = 0.0033$ and $a_i^1 = \sqrt{0.2}$, $i = 1, \ldots, 1000$.

evaluations are necessary with an *adaptive integration* algorithm for the same accuracy. For multi-factor models the integration can efficiently be calculated by globally adaptive integration schemes using either a deterministic polynomial interpolation rule or a random (Monte Carlo type) rule. For more details see [6].

Let us restate the problem of interest in the present paper in matrix form. For a given correlation matrix $\boldsymbol{C}_{M\times M}$ with $C_{ij} \geq 0$ and given $m \leq M$, find a matrix $\boldsymbol{A}_{M\times m}$ that minimizes the Frobenius norm of the matrix

$$\boldsymbol{C} - \boldsymbol{A}\boldsymbol{A}^T,$$

subject to $\boldsymbol{A}_{ij} \geq 0$ and $(\boldsymbol{A}\boldsymbol{A}^T)_{ii} = 1$.

Our problem falls in the class of *nonnegative matrix factorization* ($\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H}$) and further requires that $\boldsymbol{W}^T = \boldsymbol{H}$. This is called a *symmetric nonnegative factorization* problem in [13], which discussed another application to portfolio credit risk and in particular, in the framework of CreditRisk$^+$, another popular industrial model. The problem considered in their article is very similar to ours but the diagonal elements of $\boldsymbol{C}$ are not constrained to be 1.

3 Counterexample to $C = AA^T$

In this section we provide a counterexample to the following conjecture

Conjecture 1. *For any matrix $\boldsymbol{C}$ which is nonnegative and positive semidefinite, a nonnegative matrix $\boldsymbol{A}$ exists such that $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$*

Theorem 2. *A counter example for the equality $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T$, with $\boldsymbol{C}$ nonnegative and positive semidefinite, and $\boldsymbol{A}$ nonnegative, is provided by the following matrix $\boldsymbol{C}$:*

$$\boldsymbol{C} = \begin{pmatrix} 2 & \frac{3}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & 2 & \frac{3}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} & 2 & \frac{3}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{3}{2} & 2 & \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \frac{3}{2} & 2 & \frac{3}{2} \\ \frac{3}{2} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{3}{2} & 2 \end{pmatrix} \tag{6}$$

This nonnegative matrix can be scaled so that it has a unit main diagonal. The matrix is generated as $\boldsymbol{C} = \boldsymbol{Z}\boldsymbol{Z}^T$ with the following matrix $\boldsymbol{Z}^T$:

$$\boldsymbol{Z}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & \frac{1}{2}\sqrt{3} & \frac{1}{2}\sqrt{3} & 0 & -\frac{1}{2}\sqrt{3} & -\frac{1}{2}\sqrt{3} \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{7}$$

which represents the edges of a regular six-faced pyramid of height 1. Since $\boldsymbol{C} = \boldsymbol{Z}\boldsymbol{Z}^T$, $\boldsymbol{C}$ is positive semidefinite and therefore satisfies the conditions of the conjecture. A nonnegative matrix, $\boldsymbol{A}$, with the property that $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{C}$ does, however, not exist. Note that the condition $(\boldsymbol{A}\boldsymbol{A}^T)_{ii} = c_{ii}$ is automatically satisfied in the case of the equality.

Proof. Let's suppose that the matrix $\boldsymbol{A}$ exists, consisting of rows $\mathbf{a}^0, \mathbf{a}^1, \cdots, \mathbf{a}^5$ of a fixed, but arbitrary length. *In this section the index k should be set modulo 6 when appropriate.* It follows from (6) that the inner product $(\mathbf{a}^k, \mathbf{a}^{k+3}) = 0$, implying that all vectors must have a set of elements equal to zero, since all vectors, $\mathbf{a}$, must be nonnegative. Let us define $S_k := \{i | a_i^k \neq 0\}$. Apparently $S_k \cap S_{k+3} = \emptyset$, as otherwise $(\mathbf{a}^k, \mathbf{a}^{k+3}) \neq 0$.
We have to satisfy

1. $(\mathbf{a}^k, \mathbf{a}^k) = 2$,
2. $(\mathbf{a}^k, \mathbf{a}^{k+1}) = \frac{3}{2}$,
3. $(\mathbf{a}^k, \mathbf{a}^{k-2}) = \frac{1}{2}$,

for $k = 0, 1, \cdots, 5$, but since a part of those vectors is equal to zero we can formulate this differently:

$$\sum_{i \in S_k} (a_i^k)^2 = 2, \tag{8}$$

$$\sum_{i \in S_k \cap S_{k+1}} a_i^k a_i^{k+1} = \frac{3}{2}, \tag{9}$$

$$\sum_{i \in S_k \cap S_{k-2}} a_i^{k-2} a_i^k = \frac{1}{2}, \tag{10}$$

We note, that the sets $S_k \cap S_{k+1}$ and $S_k \cap S_{k-2}$ are disjoint, since $S_{k+1} \cap S_{k-2} = \emptyset$. From (10) we find

$$\frac{1}{2} = \sum_{i \in S_k \cap S_{k-2}} a_i^{k-2} a_i^k \leq \frac{1}{2} \sum_{i \in S_k \cap S_{k-2}} (a_i^{k-2})^2 + (a_i^k)^2, \tag{11}$$

with the equality sign only if $a_i^k = a_i^{k-2}, i \in S_k \cap S_{k-2}$. Summing over k, we obtain

$$\sum_{k=0}^{5} \left(\frac{1}{2} \sum_{i \in S_k \cap S_{k-2}} (a_i^k)^2 + \frac{1}{2} \sum_{i \in S_k \cap S_{k+2}} (a_i^k)^2 \right) \geq 3. \tag{12}$$

From (9) we find in the same way

$$\frac{3}{2} = \sum_{i \in S_k \cap S_{k+1}} a_i^k a_i^{k+1} \leq \frac{1}{2} \sum_{i \in S_k \cap S_{k+1}} (a_i^k)^2 + (a_i^{k+1})^2, \tag{13}$$

with the equality sign only if $a_i^k = a_i^{k+1}, i \in S_k \cap S_{k+1}$. Summing again over k, we obtain

$$\sum_{k=0}^{5} \left(\frac{1}{2} \sum_{i \in S_k \cap S_{k+1}} (a_i^k)^2 + \frac{1}{2} \sum_{i \in S_k \cap S_{k-1}} (a_i^k)^2 \right) \geq 9. \tag{14}$$

From (12) and (14) we obtain, because $S_k \cap (S_{k+2} \cup S_{k-1}) \subset S_k$ and similarly $S_k \cap (S_{k+1} \cup S_{k-2}) \subset S_k$,

$$\sum_{k=0}^{5} \sum_{i \in S_k} (a_i^k)^2 \geq 12. \tag{15}$$

In combination with (8) this implies, that inequality (15) is in fact an equality. By the observations made with inequalities (11) and (13) we conclude that, in order for this to be possible, we should have

$$\sum_{i \in S_k \cap S_{k-2}} (a_i^k)^2 = \sum_{i \in S_k \cap S_{k-2}} (a_i^{k-2})^2 = \frac{1}{2}, \tag{16}$$

$$\sum_{i \in S_k \cap S_{k+1}} (a_i^k)^2 = \sum_{i \in S_k \cap S_{k+1}} (a_i^{k+1})^2 = \frac{3}{2}, \tag{17}$$

for all k.

Next, we observe that for all $\mathbf{a}^k$ the set of indices of nonzero elements, S_k can be split into two disjoint subsets: $S_k \cap S_{k-2}$ and $S_k \cap S_{k+1}$. These subsets are disjoint because $S_{k-2} \cap S_{k+1} = \emptyset$ but they also completely cover S_k because $\sum_{i \in S_k} (a_i^k)^2 = 2$, $\sum_{i \in S_k \cap S_{k+1}} (a_i^k)^2 = \frac{3}{2}$ and $\sum_{i \in S_k \cap S_{k-2}} (a_i^k)^2 = \frac{1}{2}$. In the same way, S_k can be partitioned in two other disjoint sets, $S_k \cap S_{k-1}$ and $S_k \cap S_{k+2}$. So, in fact, each S_k consists of four disjoint subsets:

1. $\Sigma_k = S_k \cap S_{k-1} \cap S_{k+1}$,

2. $\Sigma_{k-1} = S_k \cap S_{k-2} \cap S_{k-1}$,

3. $\Sigma_{k+1} = S_k \cap S_{k+1} \cap S_{k+2}$,

4. $T_k = S_k \cap S_{k-2} \cap S_{k+2}$.

In view of the equalities (9), (10), (16) and (17) the vector $\mathbf{a}^k$ should be identical to $\mathbf{a}^{k+1}$ on $S_k \cap S_{k+1}$ and hence also on Σ_k since that is a smaller index set. By the same reasoning we find that $\mathbf{a}_k$ is equal to $\mathbf{a}_{k-1}$ on Σ_k. Applying the same reasoning to all six $\mathbf{a}_k$-vectors we find six disjoint Σ-sets $\Sigma_0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ and Σ_5. On Σ_k the vectors $\mathbf{a}^{k-1}, \mathbf{a}^k$ and $\mathbf{a}^{k+1}$ are identical and the other three vanish, as $(\mathbf{a}^k, \mathbf{a}^{k+3}) = 0$ and $a_i^k > 0, i \in S_k$, by the definition of S_k. Applying an analogous reasoning to T_k we find that $\mathbf{a}^k$, $\mathbf{a}^{k+2}$ and $\mathbf{a}^{k-2}$ are identical on T_k and that the other three vectors vanish on T_k. Evidently, T_0, T_2 and T_4 coincide, as do T_1, T_3 and T_5. So, there are exactly two different T-sets, say T_0 and T_1. Furthermore, index set S_k is completely covered by Σ_{k-1}, Σ_k, Σ_{k+1} and the appropriate T. On an index set Σ_j, all $\mathbf{a}^k$ are either zero or equal to a vector $\mathbf{v}^j$ *which only depends on* Σ_j, and not on k. Equally, on an index set T_j all $\mathbf{a}^k$ are either zero or equal to a vector $\mathbf{w}^j$ which only depends on j, and not on k. This implies that inner products of the form $(\mathbf{a}^k, \mathbf{a}^m)$, restricted to Σ_j are either 0 or $\|\mathbf{v}^j\|^2$, and those restricted to T_j are either 0 or $\|\mathbf{w}^j\|^2$. Therefore, we may represent each vector, $\mathbf{a}^k$, by an eight component vector, ξ^k, defined in the following way:

$$\xi_k^k := \|\mathbf{v}^k\| = \mu_k, \;\; \xi_{k-1}^k := \|\mathbf{v}^{k-1}\| = \mu_{k-1}, \;\; \xi_{k+1}^k := \|\mathbf{v}^{k+1}\| = \mu_{k+1},$$

(indices taken modulo 6). If k is even then $\xi_6^k = \|\mathbf{w}^0\|$. Otherwise, $\xi_7^k = \|\mathbf{w}^1\|$. The other four components of ξ^k are zero, and

$$(\mathbf{a}^k, \mathbf{a}^{k-1}) = \xi_{k-1}^k \xi_{k-1}^{k-1} + \xi_k^k \xi_k^{k-1} = \mu_{k-1}^2 + {\mu_k}^2 = \frac{3}{2}, \quad k = 0, \ldots, 5, \tag{18}$$

hence $\mu_k^2 = \frac{3}{4} + \nu(-1)^k, \nu \in [-\frac{3}{4}, \frac{3}{4}]$. But since we should satisfy

$$2 = (\mathbf{a}_k, \mathbf{a}_k) \geq \mu_{k-1}^2 + \mu_k^2 + \mu_{k+1}^2 = \frac{9}{4} + \nu(-1)^k, \tag{19}$$

we have a contradiction for positive ν, if k is even and for negative ν, if k is odd. □

4 An algorithm for nonnegative factorization

Now we return to the problem of finding the best nonnegative factor with respect to the Frobenius norm. So, with $\boldsymbol{E} = \boldsymbol{C} - \boldsymbol{A}\boldsymbol{A}^T$, we need to minimize

$$\|\boldsymbol{E}\|_{\text{frob}}^2 = \sum_{k,l} E_{k,l}^2 \text{ with } \text{diag}(\boldsymbol{E}) = \boldsymbol{O}.$$

So, we aim to produce an approximated nonnegative factorization of $\boldsymbol{C}$. Attempts with standard constrained optimization programs, such as Rosen's *gradient projection method* [12] did not converge at all, most likely because of the non-convexity of the problem. Therefore, we approach the problem here by a relaxation technique.
Suppose we have an approximation for $\boldsymbol{A}$, satisfying the constraints, then we aim at improving the rows of $\boldsymbol{A}$, one by one. Let $\boldsymbol{c}^j$ denote the j-th column of $\boldsymbol{C}$, then we solve:

$$\text{Minimize } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|_2, \quad \text{with} \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \|\boldsymbol{x}\| = 1. \tag{20}$$

Vector $\boldsymbol{x}$ is meant to be an improved version of $\boldsymbol{a}^j$. Replacement of the row $\boldsymbol{a}^j$ by $\boldsymbol{x}$, should result in an improved approximation.
Also for this subproblem, the application of standard constrained optimization methods did not show any success (although we did not experiment too extensively), as we expect a fundamental reason behind the failure of these techniques. The requirement $\boldsymbol{x} \geq \boldsymbol{0}$ typically gives rise to a linear programming problem, whereas the constraint $\|\boldsymbol{x}\| = 1$ asks for an analytic approach, such as a procedure with Lagrange multipliers. These two techniques appear to be not on speaking terms, so we split the treatment of the two constraints, by setting up a novel approach, based on relatively basic numerical tools.

4.1 The non-negativity conditions

Let the object function, Θ, and its gradient, $\boldsymbol{g}$, be defined by:

$$\Theta(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|^2, \ \boldsymbol{g} = \frac{\partial \Theta}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j). \tag{21}$$

The non-negativity condition usually leads to an approximate solution for which the elements x_k can be divided in a so-called feasible set, $\mathcal{F}$, and its complement, the non-feasible set. The solution satisfies

$$\begin{aligned} x_k > 0, \ g_k = 0, \qquad & k \in \mathcal{F}, \\ x_k = 0, \ g_k \geq 0, \qquad & k \notin \mathcal{F}. \end{aligned}$$

If $\boldsymbol{x}$ and $\boldsymbol{g}$ satisfy these complementary relations, we obviously have reached a *local* minimum. By a convexity argument, it is easily proved that this minimum is in fact an absolute minimum.
In Matlab the routine **lsqnonneg** [1] has been developed especially for the non-negative least-squares problem (i.e. (20) without the unit norm constraint). This routine failed rather often, uttering protesting remarks about tolerances that were too severe, or not severe enough.
In cases where the official Matlab routine worked well, it was rather slow compared to the use of the procedure within an iterative process, in which the **nnls** routine is a heavily used elementary step.

[1] Actually we used a routine named '**nnls**', which is present in an older Matlab version. This routine is mathematically equivalent with **lsqnonneg** .

In the Matlab routine **lsqnonneg** the feasible set is determined in a trial-and-error procedure similar to the classical Simplex method. This requires the least squares solution of (restricted variants of) the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}$. For reasons of numerical stability, these systems are solved using the pseudo-inverse of the restricted matrix. Because our problems have moderate size ($N \approx 50$), this rather expensive way of operating does not seem necessary. Therefore we replaced the pseudo-inverse approach in **lsqnonneg** by a basic direct solution, using Matlab's formal solution method `x=A\c`.
Also we searched for a shorter path to the solution, by choosing a significantly more efficient strategy for updating the feasible set. The most successful variant, called **pmnnlsq** (poor man's nonnegative least squares), works as follows:

Suppose we have a temporary feasible set $\mathcal{T}$.

1. We solve the system restricted to $\mathcal{T}$, and we set $x_k = 0$ outside this set.
2. We remove the indices $k \in \mathcal{T}$ for which x_k is negative, and we add indices $l \notin \mathcal{T}$ to set $\mathcal{T}$ if $g_l < 0$, because the corresponding x_l might grow positively with a decreasing object function.
3. Repeat steps 1 and 2 on the updated $\mathcal{T}$, until no further updates can be made.

When this process terminates, it has found the nonnegative least squares solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{c}$. Unfortunately, **pmnnlsq** is not always finite, but often it is, and then it is substantially faster than **lsqnonneg** .
In a future version of our routine, we will make use of **pmnnlsq** , and replace it by **lsqnonneg** for the few cases that **pmnnlsq** appears to hang.

4.2 Unit norm condition

Let $\boldsymbol{e}^1, \boldsymbol{e}^2, \ldots, \boldsymbol{e}^p$ be the Cartesian unit vectors in $\mathbb{R}^p$, and let $\boldsymbol{A}$ be a real nonnegative $M \times p$ matrix. Suppose we have found a solution $\boldsymbol{x}$ to the nonnegative least squares problem of minimizing $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$ under condition $\boldsymbol{x} \geq \boldsymbol{0}$. To this solution corresponds a feasible subset $\mathcal{F}$ and a corresponding 'feasible subspace' $\mathcal{E} = \text{span}(\cup_{k\in\mathcal{F}}\{\boldsymbol{e}^k\})$.
Let $\widetilde{\boldsymbol{A}}$ be the matrix consisting of the columns of $\boldsymbol{A}$ with indices in $\mathcal{F}$, and let $\widetilde{\boldsymbol{x}}$ be the feasible, i.e. the nonzero, part of $\boldsymbol{x}$, then $\widetilde{\boldsymbol{x}}$ is the ordinary least squares solution to $\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{x}} = \boldsymbol{b}$.
The intersection of the unit sphere in $\mathbb{R}^p$ with the subspace $\mathcal{E}$ is the unit sphere in $\mathcal{E}$. If we succeed in finding a solution $\widehat{\boldsymbol{x}} \in \mathcal{E}$, subject to the condition $\|\widehat{\boldsymbol{x}}\| = 1$, the extension of $\widehat{\boldsymbol{x}}$ to $\mathbb{R}^p$ (by choosing $x_k = 0$ for the non-feasible components) will approximate a solution of problem (20) well. Only a few issues may hamper the convergence of this algorithm:

1) A feasible variation could exist in which Θ decreases.

2) The solution $\widetilde{\boldsymbol{x}}$ in $\mathcal{E}$ may have one or more negative components. This may happen if the hyper-ellipsoids $\Theta(\widetilde{\boldsymbol{x}}) = \text{const} = \overline{C}$ have very different axes, and very skew orientations.
A robust remedy for these issues requires some more attention in a future variant. For now we neglect these.

4.2.1 Analysis of the unit-norm problem.

We now concentrate on the problem of minimizing $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|$ under the condition $\|\boldsymbol{x}\| = 1$. In this analysis the nonnegativity does not play a role, since everything happens in a feasible subspace $\mathcal{E}$ that is already obtained. For convenience, we drop the tilde signs etc.
In a more familiar setting, in which the second constraint reads $\|\boldsymbol{x}\| \leq 1$, the domain is clearly convex. In our case, however, we have to satisfy $\|\boldsymbol{x}\| = 1$ and this domain is not convex. Many local minima and even saddlepoint-solutions could be found.
Let $\widehat{\boldsymbol{x}}$ be the solution of the unconstrained least squares problem. Define for convenience $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A}$. The object function, Θ, can then be defined by:

$$\begin{aligned} \Theta(\boldsymbol{x}) &= \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \|\boldsymbol{b}\|^2 \\ &= (\boldsymbol{x} - \widehat{\boldsymbol{x}})^T\boldsymbol{B}(\boldsymbol{x} - \widehat{\boldsymbol{x}}) + \Theta(\widehat{\boldsymbol{x}}). \end{aligned} \tag{22}$$

The usual least squares solution is determined by the normal equations $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$, which is equivalent to putting the gradient of Θ to zero:

$$\boldsymbol{g} = \frac{\partial\Theta}{\partial\boldsymbol{x}} = 2(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{b}) = \boldsymbol{0}$$

The hyper-surface $\Theta(\boldsymbol{x}) = \overline{C}$ is an ellipsoid in a p-dimensional space, centered around $\widehat{\boldsymbol{x}}$. If $\overline{C}$ is very small, this surface is either completely in the interior of the unit sphere, or completely outside. If we let $\overline{C}$ grow, then eventually the surface will have contact to the sphere $\|\boldsymbol{x}\| = 1$, and this contact will be a one point contact [2]. Since both surfaces are smooth, the normal directions have to be the same in the inside situation, or opposite in the outside case. Therefore the gradient of Θ should be proportional to $\boldsymbol{x}$: So, we need to have

$$\boldsymbol{g} = 2(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{b}) = 2\alpha\boldsymbol{x}, \tag{23}$$

for some scalar α. Solution of (23) for a given α formally yields

$$\boldsymbol{x}(\alpha) = (\boldsymbol{B} - \alpha\boldsymbol{I})^{-1}\boldsymbol{c},$$

with $\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{b}$. Then the requirement $\|\boldsymbol{x}\| = 1$ reads

$$\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-2}\boldsymbol{c} = 1, \tag{24}$$

[2] Although this may not completely be true, we drop this for the moment, because it is a highly exceptional situation.

which is a nonlinear equation in α. This equation can be solved by several techniques, provided a good initial estimate is available.
Probably several solutions can be found, some of which will obviously not make sense, but certainly a choice should be made. First we obtain insight in the proper choice of solution from a geometric point of view.

4.2.2 Geometric consideration

Let's suppose that $\boldsymbol{x}$ is the point of first contact between $\Theta(\boldsymbol{x}) = \overline{C}$ and $\|\boldsymbol{x}\| = 1$, then we must have $\Theta(\widetilde{\boldsymbol{x}}) \geq \Theta(\boldsymbol{x})$, for all $\widetilde{\boldsymbol{x}}$ with $\|\widetilde{\boldsymbol{x}}\| = 1$. Now, let $\widetilde{\boldsymbol{x}} = \tau\boldsymbol{x} + \sigma\boldsymbol{t}$ be a vector with $\|\widetilde{\boldsymbol{x}}\| = 1$, with

$$\tau = \cos(\varphi),\ \sigma = \sin(\varphi),\ \boldsymbol{t}^T\boldsymbol{x} = 0,\ \|\boldsymbol{t}\| = 1. \tag{25}$$

Then, we should have

$$\begin{aligned}&\Theta(\boldsymbol{x} + (\tau - 1)\boldsymbol{x} + \sigma\boldsymbol{t}) - \Theta(\boldsymbol{x})\\ &\quad = \boldsymbol{g}^T((\tau - 1)\boldsymbol{x} + \sigma\boldsymbol{t}) + \big((\tau - 1)^2\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} + 2(\tau - 1)\sigma\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{x} + \sigma^2\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t}\big) \geq 0,\end{aligned}$$

for all $\boldsymbol{t} \perp \boldsymbol{x}$, with $|\tau|$ and $|\sigma|$ sufficiently small.
Now for small σ, we have $\tau - 1 = -\frac{1}{2}\sigma^2 + O(\sigma^4)$ from (25), and using $\boldsymbol{g}^T\boldsymbol{t} = 2\alpha\boldsymbol{x}^T\boldsymbol{t} = 0$, we get

$$-\frac{1}{2}\sigma^2\boldsymbol{g}^T\boldsymbol{x} + \sigma^2\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t} + O(\sigma^3) \geq 0.$$

Dividing by σ^2, letting $\sigma \to 0$, and using $\boldsymbol{g}^T\boldsymbol{x} = 2\alpha\|\boldsymbol{x}\|^2 = 2\alpha$, we finally obtain

$$-\alpha + \boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t} \geq 0$$

for all $\boldsymbol{t} \perp \boldsymbol{x}$, with $\|\boldsymbol{t}\| = 1$. This is equivalent to $\alpha \leq \mathcal{R}(\boldsymbol{B}, \boldsymbol{y})$, for each $\boldsymbol{y} \perp \boldsymbol{x}$, where $\mathcal{R}$ denotes the Rayleigh quotient function corresponding to $\boldsymbol{B}$.
Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of $\boldsymbol{B}$, then according to a theorem by Rayleigh, a $\theta \in (0, 1)$ exists such that:

$$\min_{\boldsymbol{t} \perp \boldsymbol{x}} \frac{\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{t}}{\boldsymbol{t}^T\boldsymbol{t}} = \lambda_1 + \theta(\lambda_2 - \lambda_1).$$

Hence $\alpha < \lambda_1 + \theta(\lambda_2 - \lambda_1)$ for some $\theta \in (0, 1)$. It follows that for our first contact, $\alpha < \lambda_2$ in any case. Next, it will be shown that the optimal value for α is the *smallest solution* of equation (24), and satisfies $\alpha < \lambda_1$.

4.2.3 Algebraic consideration

Let $\boldsymbol{x}(\alpha)$ be the solution of (23), for a given α. We should find α such that $\|\boldsymbol{x}(\alpha)\|^2 = 1$. From all values of α for which $\|\boldsymbol{x}(\alpha)\|^2 = 1$, we should select the value for which $\Theta(\boldsymbol{x}(\alpha))$ is minimal. Define $\boldsymbol{x}(\alpha)$, $F(\alpha)$ and $N(\alpha)$ by

$$\boldsymbol{B}\boldsymbol{x}(\alpha) = \alpha\boldsymbol{x}(\alpha) + \boldsymbol{c}, \tag{26}$$

$$N(\alpha) = \boldsymbol{x}(\alpha)^T\boldsymbol{x}(\alpha), \tag{27}$$

$$F(\alpha) = \boldsymbol{x}(\alpha)^T\boldsymbol{B}\boldsymbol{x}(\alpha) - 2\boldsymbol{x}(\alpha)^T\boldsymbol{c} + \boldsymbol{b}^T\boldsymbol{b}. \tag{28}$$

then we must determine

$$\min_{N(\alpha)=1} F(\alpha).$$

The following theorem provides us with a simple choice for the α-value for which F is minimal.

Theorem 3. *Let $\boldsymbol{B}$ be a symmetric positive definite matrix. Let $\boldsymbol{x}(\alpha)$, $F(\alpha)$ and $N(\alpha)$ be defined by (26), (27), (28), then, for each pair α and β with $\alpha \neq \beta$ and $N(\alpha) = N(\beta)$, the following inequality holds:*

$$\frac{F(\alpha) - F(\beta)}{\alpha - \beta} > 0. \tag{29}$$

Proof. The expression for $\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}$ can be simplified by left multiplication of (26) by $\boldsymbol{x}(\alpha)^T$:

$$\boldsymbol{x}(\alpha)^T\boldsymbol{B}\boldsymbol{x}(\alpha) = \alpha\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\alpha) + \boldsymbol{x}(\alpha)^T\boldsymbol{c} = \alpha N(\alpha) + \boldsymbol{x}(\alpha)^T\boldsymbol{c},$$

and therefore $F(\alpha)$ can be written as $F(\alpha) = \alpha N(\alpha) - \boldsymbol{x}(\alpha)^T\boldsymbol{c} + \boldsymbol{b}^T\boldsymbol{b}$. Then, we may write

$$F(\alpha) - F(\beta) = \alpha N(\alpha) - \beta N(\beta) - [\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}.$$

If $N(\alpha) = N(\beta)$ this is equivalent to

$$F(\alpha) - F(\beta) = (\alpha - \beta)\left[\frac{1}{2}N(\alpha) + \frac{1}{2}N(\beta)\right] - [\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}.$$

An expression for $[\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c}$ can be obtained by left multiplication of (26) by $\boldsymbol{x}(\beta)$, yielding

$$\boldsymbol{x}(\beta)^T\boldsymbol{B}\boldsymbol{x}(\alpha) = \alpha\boldsymbol{x}(\beta)^T\boldsymbol{x}(\alpha) + \boldsymbol{x}(\beta)^T\boldsymbol{c}. \tag{30}$$

Interchanging α and β in this equation leaves the inner product $\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)$ unchanged. Also the left-hand side does not change, since $\boldsymbol{B}$ is symmetric. By subtracting the interchanged variant of (30) from (30) itself, we therefore get:

$$0 = (\alpha - \beta)\boldsymbol{x}(\beta)^T\boldsymbol{x}(\alpha) + [\boldsymbol{x}(\beta) - \boldsymbol{x}(\alpha)]^T\boldsymbol{c},$$

from which it follows that

$$[\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)]^T\boldsymbol{c} = (\alpha - \beta)\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta).$$

So, the *F-difference* can be written as

$$F(\alpha) - F(\beta) = (\alpha - \beta)\left[\frac{1}{2}N(\alpha) + \frac{1}{2}N(\beta) - \boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)^T\right].$$

Substituting $\|\boldsymbol{x}(\alpha)\|^2$ for $N(\alpha)$, etc, we finally arrive at

$$\begin{aligned} F(\alpha) - F(\beta) &= \frac{1}{2}(\alpha - \beta)[\|\boldsymbol{x}(\alpha)\|^2 + \|\boldsymbol{x}(\beta)\|^2 - 2\boldsymbol{x}(\alpha)^T\boldsymbol{x}(\beta)] \\ &= \frac{1}{2}(\alpha - \beta)\|\boldsymbol{x}(\alpha) - \boldsymbol{x}(\beta)\|^2, \end{aligned}$$

which implies (29). □

According to this theorem, we should search for the smallest solution of the equation $N(\alpha) = 1$.
Now consider the explicit formula (24) for $\|\boldsymbol{x}(\alpha)\|^2$,

$$N(\alpha) = \|\boldsymbol{x}(\alpha)\|^2 = \boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-2}\boldsymbol{c}.$$

Differentiating $N(\alpha)$ with respect to α gives $N'(\alpha) = 2\boldsymbol{c}^T(\boldsymbol{B} - \alpha\boldsymbol{I})^{-3}\boldsymbol{c}$. If $\alpha < \lambda_1$, $\boldsymbol{B} - \alpha\boldsymbol{I}$ is a positive definite matrix, and so is $(\boldsymbol{B} - \alpha\boldsymbol{I})^{-3}$. Therefore N is monotonically increasing for $\alpha < \lambda_1$. Now,

$$\begin{array}{lll} N(\alpha) \to 0 & \text{as} & \alpha \to -\infty, \\ N(\alpha) \to +\infty & \text{as} & \alpha \uparrow \lambda_1. \end{array}$$

Hence, there is precisely one value $\alpha < \lambda_1$ for which $N(\alpha) = 1$, and this value is the smallest solution of equation (24).
This makes it relatively easy to determine an initial guess for the iterative solution procedure.

4.3 Algorithm

Here is a global description of the algorithm for the nonnegative matrix factorization of $\boldsymbol{C}$ into $\boldsymbol{A}$ and $\boldsymbol{A}^T$ subject to $a_i a_i^T = 1$:

- If $\boldsymbol{A}$ is an approximate nonnegative factor, with $\text{diag}(\boldsymbol{A}\boldsymbol{A}^T) = \boldsymbol{I}$, then the main step in the algorithm is the replacement of all rows of $\boldsymbol{A}$ by improved versions.

- The *replacement of a row*, $\boldsymbol{a}^j$, by $\boldsymbol{x}$ requires the solution of the following expression:

$$\text{Minimize } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{c}^j\|^2, \text{ subject to: } \boldsymbol{x} \geq \boldsymbol{0}, \ \|\boldsymbol{x}\| = 1.$$

  The solution of this norm-restricted, nonnegative least squares problem for row replacement is done in two steps:

  1. Find a nonnegative least squares solution, $\boldsymbol{x}_c$, with **pmnnlsq** from Section 4.1. If this process is not finite (and thus periodic), choose the Matlab procedure **lsqnonneg** instead.

     The indices of the nonzero entries of $\boldsymbol{x}_c$ build the feasible set $\mathcal{F}$. Suppose that this set has p elements.
  2. Let $\widehat{\boldsymbol{B}}$ the $p \times p$ sub-matrix of $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A}$, obtained by removing the rows and columns with indices that are not in $\mathcal{F}$. Let $\widehat{\boldsymbol{c}}$ be the $\mathcal{F}$-*restriction* of $\boldsymbol{A}^T\boldsymbol{c}^j$. Determine by means of Newton's procedure the minimal solution of the expression $\|(\widehat{\boldsymbol{B}} - \alpha\boldsymbol{I})^{-1}\widehat{\boldsymbol{c}}\|^2$. If the restricted solution is not feasible, as it gives rise to negative entries, remove the incorrect indices from $\mathcal{F}$, and repeat the procedure. If the solution is not minimal if observed in the complete space, accept this sub-optimal solution. Continue by returning to step 1 with the next row.

The iteration process is stopped by an Aitken error estimate: Let $r_n = \|\boldsymbol{E}_n\|_{\text{frob}}$, and suppose $r_n \to r$, then the Aitken estimate of $r_n - r$ reads

$$r_n - r \approx \frac{(r_n - r_{n-1})^2}{r_n - 2r_{n-1} + r_{n-2}}, \tag{31}$$

which is based on the hypothesis of linear convergence of the sequence $\{r_n\}$:

$$r_n \approx r + \text{ const } \rho^n. \tag{32}$$

This criterion appears to be useful, also in cases where the hypothesis (32) is not valid.

5 Numerical results

In this section we present some factorization results obtained with the algorithm presented in the previous section. All experiments are performed in Matlab, version 7.4, on an Intel(R) Core (TM) 2 6700 2.66 GHz processor.
The first example is taken from [16]. The correlation matrix $\boldsymbol{C}$ is given as follows,

$$\begin{pmatrix}
1.0000 & 0.8415 & 0.6246 & 0.6231 & 0.5330 & 0.4287 & 0.3274 & 0.4463 & 0.2439 & 0.3326 & 0.2625 \\
0.8415 & 1.0000 & 0.7903 & 0.7844 & 0.7320 & 0.6346 & 0.4521 & 0.5812 & 0.3439 & 0.4533 & 0.3661 \\
0.6246 & 0.7903 & 1.0000 & 0.9967 & 0.8108 & 0.7239 & 0.5429 & 0.6121 & 0.4426 & 0.5189 & 0.4251 \\
0.6231 & 0.7844 & 0.9967 & 1.0000 & 0.8149 & 0.7286 & 0.5384 & 0.6169 & 0.4464 & 0.5233 & 0.4299 \\
0.5330 & 0.7320 & 0.8108 & 0.8149 & 1.0000 & 0.9756 & 0.5676 & 0.6860 & 0.4969 & 0.5734 & 0.4771 \\
0.4287 & 0.6346 & 0.7239 & 0.7286 & 0.9756 & 1.0000 & 0.5457 & 0.6583 & 0.4921 & 0.5510 & 0.4581 \\
0.3274 & 0.4521 & 0.5429 & 0.5384 & 0.5676 & 0.5457 & 1.0000 & 0.5942 & 0.6078 & 0.6751 & 0.6017 \\
0.4463 & 0.5812 & 0.6121 & 0.6169 & 0.6860 & 0.6583 & 0.5942 & 1.0000 & 0.4845 & 0.6452 & 0.5673 \\
0.2439 & 0.3439 & 0.4426 & 0.4464 & 0.4969 & 0.4921 & 0.6078 & 0.4845 & 1.0000 & 0.6015 & 0.5200 \\
0.3326 & 0.4533 & 0.5189 & 0.5233 & 0.5734 & 0.5510 & 0.6751 & 0.6452 & 0.6015 & 1.0000 & 0.9889 \\
0.2625 & 0.3661 & 0.4251 & 0.4299 & 0.4771 & 0.4581 & 0.6017 & 0.5673 & 0.5200 & 0.9889 & 1.0000
\end{pmatrix}$$

The correlation matrix $\boldsymbol{C}$ is illustrated graphically on a two-dimensional grid in Figure 2(a), where the colored surface is formed by the c_{ij} in matrix $\boldsymbol{C}$. Figure 2(b)-(d) correspond to the nonnegative low-rank approximations AA^T, with rank m =2, 3 and 6, respectively. The results obtained are comparable to those in [16], where the non-negativity constraint was not imposed.
The convergence in the Frobenius norm with increasing rank is presented in Figure 3(a). With regard to the computation time, an approximation can be found within a second for all $m \leq 11$. Figure 3(b) illustrates the 6 largest eigenvalues of $\boldsymbol{C}$ and its nonnegative low-rank approximation with rank $m = 6$.

In our second example $\boldsymbol{C}$ is a 50×50 correlation matrix with its entries defined to be

$$c_{ij} = LongCorr + (1 - LongCorr)e^{\kappa|t_i - t_j|},$$

$$\kappa = d_1 - d_2 \max(t_i, t_j),$$

where $LongCorr = 0.3$, $d_1 = -0.12$, $d_2 = 0.005$.
Figure 4(a) shows the convergence in the Frobenius norm with increasing rank. Figure 4(b) compares the 5 largest eigenvalues of $\boldsymbol{C}$ to its nonnegative low-rank approximations with rank $m = 5, 10$. As the size of the matrix is significantly bigger than the one in previous example, this computation is more costly. The CPU time for finding the nonnegative low-rank approximations for $m = 1, \cdots, 30$ is on average around 10 seconds. Moreover, Figure 5 indicates that as the rank of $\boldsymbol{A}$ increases, the CPU time for finding the nonnegative low-rank approximations also tends to increase.

6 Conclusion

We have presented a dedicated algorithm for the nonnegative factorization of a correlation matrix, as it appears in an application from financial engineering. The algorithm is based on a two-step procedure. First the non-negativity constraint is dealt with, by means of nonnegative least-squares routines, available in Matlab. Secondly, the unit norm condition is taken into account. The algorithm comes with a detailed explanation of all its steps. The methods works well, as is confirmed by some numerical experiments.
Finally, we have also presented a counter example, showing that it is not always possible to find an exact nonnegative matrix factorization, with factors of the same rank as the original correlation matrix.

References

[1] L. Andersen, J. Sidenius, and S. Basu. All your hedges in one basket. *RISK*, (November):67–72, 2003.

[2] M. Crouhy, D. Galai, and R. Mark. A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1):59–117, 2000.

[3] M.W. Berry, M. Brown, A.N. Langville, V.P. Pauca, and R.J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1):155–173, 2007.

[4] P. Groenen and R. Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra and its Applications*, 422(2-3):629–653, 2007.

[5] G. Gupton, C. Finger, and M Bhatia. *CreditMetrics – Technical document.* J.P.Morgan New York, 1997.

[6] X. Huang and C. W. Oosterlee. *Adaptive integration for multi-factor portfolio credit loss models.* Technical Report 07-13, Department of Applied Mathematical Analysis, Delft University of Technology, October 2007.

[7] X. Huang, C. W. Oosterlee, and J. A. M. van der Weide. Higher order saddlepoint approximations in the Vasicek portfolio credit loss model. *Journal of Computational Finance*, 11(1):93–113, 2007.

[8] S. Kealhofer. *Portfolio Management of Default Risk.* KMV Corporation, May 2001.

[9] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[10] R Martin. *Credit portfolio modeling handbook*. Credit Suisse Financial products, 2004.

[11] R. Pietersz and P. Groenen. Rank reduction of correlation matrices by majorization. *Quantitative Finance*, 4(6):649–662, 2004.

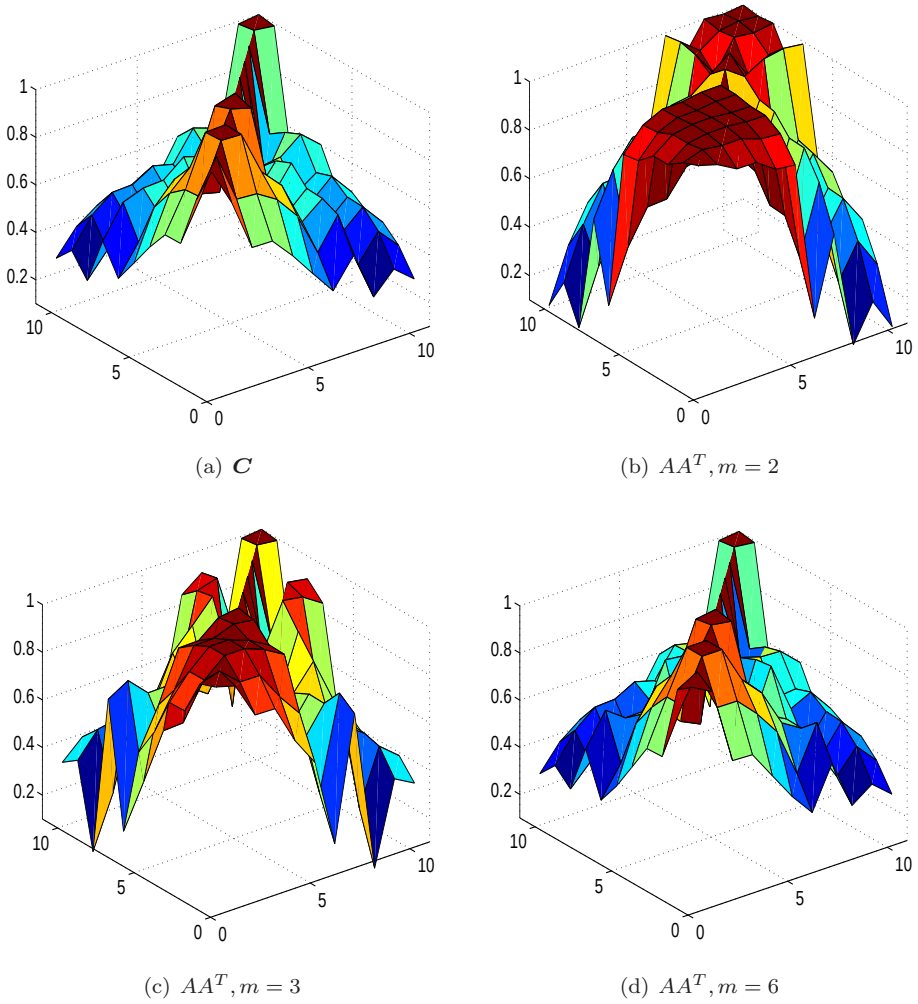
[12] J.B. Rosen. The gradient projection method for nonlinear programming, Part 1: Linear constraints. *SIAM J. Appl. Math.,* 8: 181-217, 1960.

[13] A. Vandendorpe, N.-D. Ho, S. Vanduffel, and P. Van Dooren. On the parameterization of the CreditRisk$^+$ model for estimating credit portfolio risk. *Insurance: Mathematics and Economics*, 42(2):736–745, 2008.

[14] L. Wu. Fast at-the-money calibration of the Libor market model using Lagrange multipliers. *Journal of Computational Finance*, 6(2):39–77, 2003.

[15] B. Zeng and J. Zhang. *An Empirical Assessment of Asset Correlation Models*. KMV Corporation, November 2001.

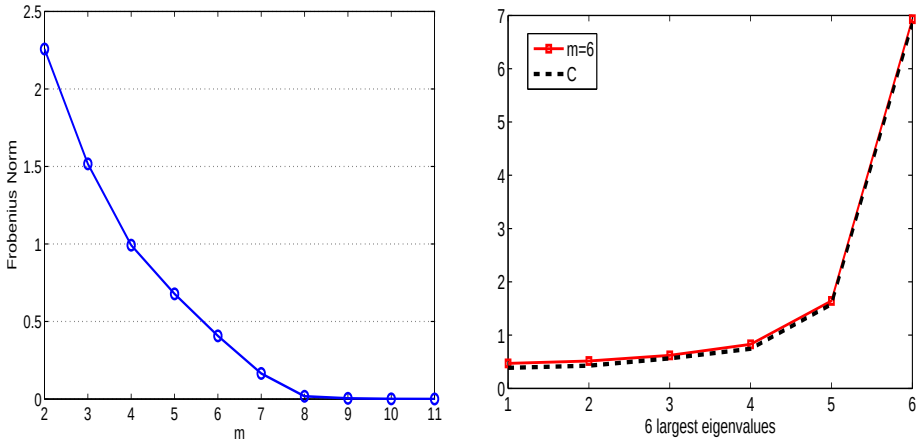
[16] Z. Zhang and L. Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra and its Applications*, 364:161–187, 2003.



(a) $\boldsymbol{C}$ (b) $AA^T, m = 2$

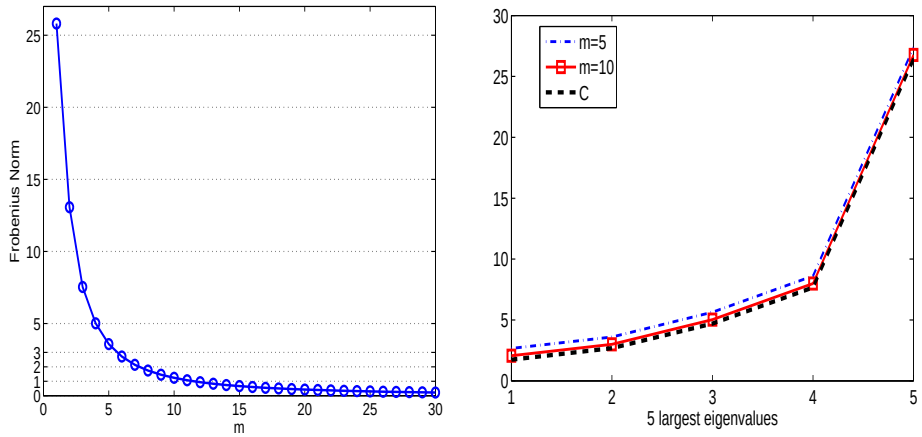
(c) $AA^T, m = 3$ (d) $AA^T, m = 6$

Figure 2: (a) The correlation structure of $\boldsymbol{C}$ displayed on a 2D grid, (b)-(d) The correlation structure of the nonnegative low-rank approximations to $\boldsymbol{C}$. The ranks of matrices shown in (b)-(d) are m =2, 3 and 6 respectively.



(a) (b)

Figure 3: (a) Convergence in the Frobenius norm with increasing rank m. (b) The 6 largest eigenvalues of matrix $\boldsymbol{C}$ and its nonnegative low-rank approximations ($m = 6$).



(a) (b)

Figure 4: (a) Convergence in the Frobenius norm with increasing rank m. (b) The 5 largest eigenvalues of matrix $\boldsymbol{C}$ and its nonnegative low-rank approximations ($m = 5, 10$).

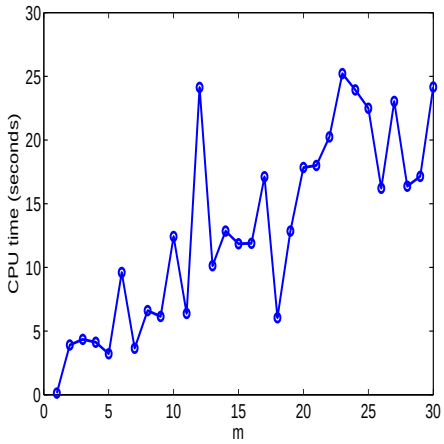


Figure 5: CPU time with increasing rank m.