

# On The Parameterization of the CreditRisk<sup>+</sup> Model for Estimating Credit Portfolio Risk

Antoine Vandendorpe<sup>a</sup>, Ngoc-Diep Ho<sup>b</sup>, Steven Vanduffel<sup>c</sup> and Paul Van Dooren<sup>b</sup>

<sup>a</sup>*Fortis Central Risk Management, Belgium.*

<sup>b</sup>*Université Catholique de Louvain, Department of Mathematical Engineering, Louvain-La-Neuve, 1348, Belgium.*

<sup>c</sup>*Katholieke Universiteit Leuven, Department of Economics and Applied Economics, Leuven, 3000, Belgium.*

---

## Abstract

The Credit Risk<sup>+</sup> model is one of the industry standards for estimating the credit default risk for a portfolio of credit loans. The natural parameterization of this model requires the default probability to be apportioned using a number of (non-negative) factor loadings. However, in practice only default correlations are often available but not the factor loadings. In this paper we investigate how to deduce the factor loadings from a given set of default correlations. This is a novel approach and it requires the non-negative factorization of a positive semi-definite matrix which is by no means trivial. We also present a numerical optimization algorithm to achieve this.

---

## 1 Introduction

The assessment of credit risk for loan portfolios is now widely established within the financial world. The basic components that are needed to model the default risk for each loan are: the Probability of Default (PD), the Exposure At Default (EAD) and the Loss Given Default (LGD). We assume the EAD and the LGD to be deterministic and it follows then that the distribution function (d.f.) for the total credit portfolio loss can be calculated provided

---

*Email addresses:* [antoine.vandendorpe@fortis.com](mailto:antoine.vandendorpe@fortis.com) (Antoine Vandendorpe), [ho@inma.ucl.ac.be](mailto:ho@inma.ucl.ac.be) (Ngoc-Diep Ho), [Steven.Vanduffel@econ.kuleuven.be](mailto:Steven.Vanduffel@econ.kuleuven.be) (Steven Vanduffel), [vdooren@inma.ucl.ac.be](mailto:vdooren@inma.ucl.ac.be) (Paul Van Dooren).

that one also knows all joint default probabilities or, equivalently, the copula. Since defaults are by nature rare events, fitting a copula in a credit context is far from being a trivial task and most financial institutions only explicitly model the individual and pairwise default probabilities from which default correlations can be directly derived. Indeed, it must be noted that despite its conceptual simplicity it remains difficult to estimate *default correlations*.

Ultimately, as no intermediate process is assumed, the use of observed default data is obviously the best source to estimate single and pairwise default frequencies and default correlations can then be derived from this; see Gordy (2000), Frey and McNeil (2003), Jobst and de Servigny (2005) and de Servigny and Renault (2002) for discussions and empirical studies.

However, confronted with a lack of sufficient default data, financial institutions and also software providers often resort to *equity or asset correlations*. These correlations are then transformed into default correlations by means of the so-called *Merton Model* of the firm; we refer the interested reader to Crouhy *et al.* (2000) for further details. Whilst the wider availability of data and the additional sophistication may then provide some feeling of comfort, this goes at the cost of more model risk. There is no indication that more complex approaches to estimate default correlations provide more reliable estimates; see Chernih *et al.* (2006) for a comparative study.

Unfortunately though, since default correlations are related to individual and pairwise but not multiple default probabilities, they are not only difficult to estimate but they also do not provide a full picture of the dependency. Hence, in order to build a model to estimate their portfolio credit risk, financial companies had to complement their estimates for PD, LGD, EAD and default correlations with explicit or implicit assumptions regarding the dependency structure. This has given rise to three widely established credit risk portfolio models in industry:

- (1) the “Merton-based” approach; see J.P. Morgan’s CreditMetrics (JP Morgan & Co. Inc., 1997) or MKMV’s PortfolioManager (Kealhofer, 1995),
- (2) the “Econometric” approach; see McKinsey & Company’s CreditPortfolioView (Wilson, 1997a,b),
- (3) the “Actuarial” approach; see CreditRisk<sup>+</sup> (Credit Suisse Financial Products, 1997).

Koyluoglu and Hickman (1998) showed that these models can be placed in a single mathematical framework and that they merely differ through the choice and parametrization of their copula; see also Crouhy *et al.* (2000). It follows that the results of these three approaches may differ significantly even when using the same marginal distributions and the same set of default correlations. This is because in general, for a given set of marginals and correlations, several

copulas that preserve the correlations will exist and each of these copulas will give rise to one particular probability distribution function for the total credit portfolio loss  $S$ ; we refer to Frey and McNeil (2003) and Embrechts *et al.* (2003) for various eloquent examples illustrating this.

A major advantage of the CreditRisk<sup>+</sup> model is that an analytical expression for the moment generating function of the loss distribution of the portfolio is available and this gives several possibilities to obtain rapidly the distribution function of  $S$ . The original CreditRisk<sup>+</sup> documentation proposed an algorithm known in actuarial circles as Panjer's recursion; see e.g. Panjer and Willmot (1992). This recursion sometimes suffers from numerical instability and in the literature several remedies have been suggested to deal with this issue; Gordy (2002) suggests a saddle-point approximation whereas Reiss (2003) shows how Fast Fourier Transforms can be used to obtain the distribution function of  $S$ .

Surprisingly, we feel that less attention went to the parameterization of the CreditRisk<sup>+</sup> model and in this paper we will investigate this aspect when information on PD's, EAD's, LGD's and default correlations is available. The relevance of such an analysis is underpinned by the fact that it is precisely this information (only) that financial institutions typically have at hand when building their portfolio models.

We point out that the Credit Risk<sup>+</sup> model traditionally uses PD's, EAD's, LGD's but not default correlation as input parameter. In fact, the Standard - and most simple - Credit Risk<sup>+</sup> model assumes that default probabilities are stochastic (with mean equal to PD). The stochasticity depends on a unique systematic factor that is shared by all credit exposures and fully determines the default correlations. Next, the most general CreditRisk<sup>+</sup> model allocates the different credit risks to more than one segment and apportions the default probability using now a number of (non-negative) factor loadings. From this the default correlation matrix can be deduced.

However, in practical situations one often has the default correlations at hand but not the factor loadings and this puts a burden on the parameterization of the Credit Risk<sup>+</sup> model. Indeed, it is not straightforward to derive from a given set of default correlations the positive factor loadings and this will be the topic of this paper.

Increasing the number of factors in the CreditRisk<sup>+</sup> model is always tied to increasing the rank of the approximation for the default correlation matrix. Ultimately, the use of as many factors as credit exposures would allow to reproduce the correlation matrix exactly. In principle, this would lead to the most accurate approximations, taking into account the amount of information at hand, i.e. PD, LGD, EAD and default correlations. However, a proper use of the CreditRisk<sup>+</sup> model requires all different factor loadings to be positive but

there is no guarantee that the best rank  $k$  approximation (obtained through the singular value decomposition) will also be positive, even when all default correlations are positive. Even if the best rank  $k$  approximation was positive, there is no guarantee to find a non-negative symmetric factorization of it. It is therefore interesting to see how well the default correlation matrix can be approximated by a low rank matrix, while constraining the factor loadings to be positive.

In mathematical language the above problem amounts to obtaining a non-negative factorization of a symmetric positive semi-definite matrix (i.e. a matrix with all non-negative eigenvalues). It was shown in Castral *et al.* (2004) that the symmetry condition complicates the problem considerably. Nevertheless, in the present paper, we consider solutions that constrain the approximation to be symmetric. Moreover, we use a modified cost function which includes a weighting of each individual element of the matrix to be approximated.

The problem of Weighted non-negative Matrix Factorization was first stated in Paatero and Tapper (1994) and Paatero (1997) where the cost function was the Weighted Euclidean Distance. Several algorithms were used to solve the problem, including Newton-related methods which have a high complexity. Simpler algorithms were introduced by Lee and Seung (Lee and Seung, 1999). These are based on a set of multiplicative updating rules which will be mentioned later in this text. Modifications of these algorithms are used here to find the minimizers for the *unweighted* Euclidean Distance.

Recently, in Guillamet and Vitri (2003), weights were added in the divergence cost function, but the weight matrix was limited to a diagonal scaling of the matrix to be approximated. Our proposed algorithm incorporates the Lee and Seung algorithm with a non-negative weight matrix of size equal to that of the data matrix. We also modify the cost function to be minimized in order to converge to factors that are nearly symmetric.

In Section 2 we provide a short overview of the CreditRisk<sup>+</sup> model whereas Section 3 presents its correlation structure and Section 4 discusses its parameterization. In Sections 5 and 6, the problem of approximating a symmetric matrix with completely positive matrices is introduced and a new algorithm is given. Finally, some numerical results are given in Section 7.

## 2 The CreditRisk<sup>+</sup> Model

We consider a loan portfolio consisting of  $n$  risky loans. Let  $I_i$  be the indicator variable which equals 1 if risk  $i$  leads to failure in the next reference

period (e.g., one year time horizon) and 0 otherwise. The default probability is stochastic and is denoted by  $Q_i$  with mean  $E[Q_i] = q_i$ . It follows that the probability that risk  $i$  leads to a failure in the corresponding period, is given by  $q_i$ :

$$\Pr [I_i = 1] = E[Q_i] = q_i. \quad (1)$$

The probability that risks  $i$  and  $j$  both default, is denoted by  $q_{ij}$ . Furthermore, let  $EAD_i$  denote the ‘‘Exposure At Default’’ or the maximal amount of loss on risk  $i$ , and let  $LGD_i$  denote the ‘‘Loss Given Default’’ or the percentage of the loss on risk  $i$ , given that a default occurs. The ‘‘Credit Portfolio Loss’’ during the reference period, denoted by  $S$ , is then given by:

$$S = \sum_{i=1}^n I_i (EAD)_i (LGD)_i. \quad (2)$$

We will assume that all  $(EAD)_i$  and  $(LGD)_i$  are deterministic. A general discussion about the use (and abuse) of stochastic recovery rates in credit portfolio models is considered in Dhaene et al. (2005) whereas Bürgisser *et al.* (2001) provides a generalization of the CreditRisk<sup>+</sup> model that deals with stochastic dependent  $LGD$ 's.

Note that the random variable (r.v.) of interest can be written as a sum of  $n$  compound Bernoulli random variables

$$S = \sum_{i=1}^n I_i c_i, \quad c_i = (EAD)_i (LGD)_i. \quad (3)$$

In order to compute the distribution function of  $S$  exactly, knowledge of the multivariate distribution function for  $\mathbf{I} = [I_1 \ I_2 \ \dots \ I_n]^T$  is required but this is hard or even impossible to get. Moreover, even if we would know the multivariate distribution function for  $\mathbf{I}$ , it appears that dependent Bernoulli r.v's  $I_i$  are too cumbersome to work with. Therefore, we will replace the Bernoulli r.v's  $I_i$  by other dependent r.v's  $N_i$  that are ‘‘close’’ to the individual  $I_i$  but are more easy to work with; see Credit Suisse Financial Products (1997) and the references therein. Hence instead of  $S$  we will consider the r.v.  $S^*$  defined as:

$$S^* = \sum_{i=1}^n N_i c_i. \quad (4)$$

In order to define the multivariate distribution of  $\mathbf{N} = [N_1 \ N_2 \ \dots \ N_n]^T$  we assume that there exists a random vector  $\mathbf{\Gamma} = [\Gamma_1 \ \dots \ \Gamma_K]^T$  representing the ‘‘state of the economy’’ such that the random variables  $\bar{N}_i$ , conditioned by  $\Gamma_k = \gamma_k$ , ( $k = 1, 2, \dots, K$ ), are mutually independent and Poisson distributed

with intensity:

$$\mathcal{R}_i = q_i \left( w_{0,i} + \sum_{k=1}^K w_{k,i} \gamma_k \right), \quad (5)$$

i.e.

$$\left( N_i \mid \mathbf{\Gamma} = \gamma_1 \dots \gamma_K \right) \stackrel{d}{=} \text{Poisson}(\mathcal{R}_i). \quad (6)$$

The coefficient  $w_{0,i}$  reflects the portion of idiosyncratic risk that can be attributed to the  $i$ -th risk whereas  $w_{k,i}$  reflects its affiliation to the  $k$ -th common factor. It is important to note that in order to have positive intensities in (5), the coefficients  $w_{k,i}$ ,  $k = 0, 1, \dots, K$  will be constrained to be non-negative.

The random variables  $\Gamma_i$  are assumed to be independent Gamma distributed and the covariance matrix associated with the random vector  $\mathbf{\Gamma}$  is given by:

$$\text{cov}[\mathbf{\Gamma}] = \mathbf{\Sigma} := \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_K^2 \end{bmatrix}.$$

Since the r.v.  $a\Gamma_k$  will be distributed like a Gamma r.v. for any  $k$  and  $a > 0$ , we can assume without loss of generality that  $E[\Gamma_i] = 1$ .

In the next section we will discuss how to choose the different parameters  $w_{k,i}$  and  $\sigma_k^2$  such that the information present within  $\mathbf{N} = [N_1 \ N_2 \ \dots \ N_n]^T$  closely resembles the information of  $\mathbf{I}$ . An analytical expression for the moment generating function of the aggregated loss  $S$  is then given by Haaf et al. (2003) :

$$m_S(t) = \exp \left( \sum_{i=1}^n w_{0,i} q_i (e^{tc_i} - 1) - \sum_{k=1}^K \frac{1}{\sigma_k^2} \ln \left[ 1 - \sigma_k^2 \sum_{i=1}^n w_{k,i} q_i (e^{tc_i} - 1) \right] \right). \quad (7)$$

From this last expression, it is not difficult to derive an algorithm based on the Fast Fourier Transform to compute the probability distribution function of  $S$ ; see e.g. Haaf et al. (2003) or Reiss (2003).

In the remainder of the paper we will assume that for every risk we (only) know its default probability  $q_i$ , its Loss Given Default  $LGD_i$ , its Exposure at Default  $EAD_i$ , as well as its default correlation  $\text{corr}[I_i, I_j]$  with other risks. Since we have that:

$$\text{corr}[I_i, I_j] = \frac{q_{ij} - q_i q_j}{\sqrt{q_i(1 - q_i) q_j(1 - q_j)}}, \quad (8)$$

it follows that we assume that all single and pairwise default probabilities are known, but not the multiple ones. The following two sections shows how to set up the CreditRisk<sup>+</sup> model using this information only.

### 3 The Covariance Matrix of the CreditRisk<sup>+</sup> Model

The covariance matrix  $cov[\mathbf{N}]$  associated with the random vector  $\mathbf{N}$  has elements

$$cov[\mathbf{N}]_{ij} = cov[N_i, N_j] \quad (9)$$

It is a well-known result in statistics that for any r.v.'s  $X, Y$  and  $Z$  we have that:

$$cov[X, Y] = E[cov[X, Y|Z]] + cov[E[X|Z], E[Y|Z]]. \quad (10)$$

Furthermore, from the properties of Poisson random variables it follows that

$$var(N_i|\Gamma) = E[N_i|\Gamma] = q_i \left( w_{0,i} + \sum_{k=1}^K w_{k,i} \Gamma_k \right). \quad (11)$$

When  $i \neq j$  ( $i, j = 1, 2, \dots, n$ ) we then have that:

$$\begin{aligned} cov[N_i, N_j] &= E[cov[N_i, N_j|\Gamma]] + cov[E[N_i|\Gamma], E[N_j|\Gamma]] \\ &= 0 + cov \left[ q_i \left( w_{0,i} + \sum_{k=1}^K w_{k,i} \Gamma_k \right), q_j \left( w_{0,j} + \sum_{l=1}^K w_{l,j} \Gamma_l \right) \right] \\ &= q_i q_j \sum_{k=1}^K \sum_{l=1}^K w_{k,i} w_{k,j} cov[\Gamma_k, \Gamma_l] \\ &= q_i q_j \sum_{k=1}^K w_{k,i} w_{k,j} \sigma_k^2. \end{aligned} \quad (12)$$

For the diagonal terms we find that:

$$\begin{aligned} var[N_i] &= E[var[N_i|\Gamma]] + var[E[N_i|\Gamma]] \\ &= q_i \left( \sum_{k=0}^K w_{k,i} \right) + q_i^2 \sum_{k=1}^K w_{k,i}^2 \sigma_k^2. \end{aligned} \quad (13)$$

We find now from (12) that  $(cov[\mathbf{N}])_{ij}$  is given by

$$cov[\mathbf{N}]_{ij} = q_i q_j \sum_{k=1}^K w_{k,i} w_{k,j} \sigma_k^2 + \delta_{ij} q_i \sum_{k=0}^K w_{k,i}. \quad (14)$$

In this expression  $\delta_{ij}$  denotes the Kronecker delta, i.e.  $\delta_{ij} = 1$  if  $i = j$  and  $\delta_{ij} = 0$  if  $i \neq j$ .

For the variance of  $S^*$  it follows now that

$$\text{var}[S^*] = \sum_{i=1}^n \sum_{j=1}^n c_i c_j (\text{cov}[\mathbf{N}])_{ij}. \quad (15)$$

From (4) we obtain an expression for the expected loss of  $S^*$  :

$$E[S^*] = \sum_{i=1}^n c_i q_i \sum_{k=0}^K w_{k,i}. \quad (16)$$

The equations (14), (15) and (16) provide expressions for some key characteristics of our approximate model and hence this should be constructed such that for any  $i, j = 1, 2, \dots, n$ , the following conditions  $C1$  -  $C4$  are (approximately) fulfilled. Note that the notation  $F_X(x)$  is used to denote the d.f. for a r.v.  $X$ .

$$C1: E[S^*] = E[S]$$

$$C2: \text{var}[S^*] = \text{var}[S]$$

$$C3: \text{cov}[\mathbf{N}] = \text{cov}[\mathbf{I}].$$

$$C4: F_{N_i}(x) = F_{I_i}(x) \text{ for all } i = 1, 2, \dots, n. \text{ and } x = 0, 1, 2, \dots, n.$$

We notice that the different conditions are not all independent from each other. In particular, condition  $C4$  implies that  $E[N_i] = E[I_i]$  should hold for  $i = 1, 2, \dots, n$ , and in this case condition  $C1$  will also be fulfilled automatically. Moreover, since the  $N_i$  are conditionally Poisson distributed whereas the  $I_i$  are conditionally Bernoulli distributed, condition  $C4$  can never be met exactly. Hence, our model is subject to some approximation error which will be addressed further in the paper.

In the remainder of the paper we will always assume that  $E[N_i] = E[I_i]$  but this means that for any  $i = 1, 2, \dots, n$  we must have that :

$$\sum_{k=0}^K w_{k,i} = 1. \quad (17)$$

In the next section we will discuss how to choose the coefficients  $w_{i,j}$  and  $\sigma_i$  such that the available information from the individual *true* model is preserved by our approximate model.

## 4 Parameterization of the model

### 4.1 The Standard CreditRisk<sup>+</sup> model

The standard CreditRisk<sup>+</sup> model assumes that there is no idiosyncratic risk and only one common factor, *i.e.*  $w_{1,i} > 0$ ,  $\sigma_1^2 > 0$  whereas all other parameters vanish. It remains to determine  $w_{1,i}$  and  $\sigma_1^2$  such that the different conditions  $C1$  to  $C4$  are (approximately) met.

From (17) it follows immediately that  $w_{1,i} = 1$ . In order to make sure that also condition  $C2$  is fulfilled, we find now from (15) that

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j q_i q_j \sigma_1^2 + \sum_{i=1}^n c_i^2 q_i = \text{var}[S] \quad (18)$$

should hold and from this it follows that  $\sigma_1^2$  is given by:

$$\sigma_1^2 = \frac{\text{var}(S) - (\sum_{i=1}^n c_i^2 q_i)}{(\sum_{i=1}^n c_i q_i)^2}. \quad (19)$$

Note that in order to guarantee that  $\sigma_1^2 > 0$  we must have

$$\text{var}(S) > \left( \sum_{i=1}^n c_i^2 q_i \right) > \sum_{i=1}^n c_i^2 q_i (1 - q_i) = \text{var}(S_{ind}) \quad (20)$$

where  $S_{ind}$  denotes the aggregate claims under the individual model assuming that all risks are independent. Hence, the approximate model only makes sense in the case that the different correlations are (on average) positive. Since different loans are subject to some common economic factors this condition will in practice always be fulfilled.

We conclude that this parameterization guarantees that conditions  $C1$  and  $C2$  are met. The question is now to which extent conditions  $C3$  and  $C4$  will be fulfilled.

In order to determine the distribution function of  $N_i$ , we determine its moment generation function. We find:

$$\begin{aligned}
E[\exp(tN_i)] &= E[E[\exp(tN_i) \mid \Gamma]] \\
&= E[\exp(q_i \Gamma_1 (\exp(t) - 1))] \\
&= m_{\Gamma_1}(q_i (\exp(t) - 1)) \\
&= \left( \frac{\beta_1}{\beta_1 - q_i (\exp(t) - 1)} \right)^{\beta_1} \\
&= \left( \frac{\frac{\beta_1}{\beta_1 + q_i}}{1 - \left(1 - \frac{\beta_1}{\beta_1 + q_i}\right) \exp(t)} \right)^{\beta_1}, \tag{21}
\end{aligned}$$

with  $\beta_1$  defined as

$$\beta_1 = \frac{1}{\sigma_1^2}. \tag{22}$$

Inspection of (21) tells us that

$$N_i \stackrel{d}{=} NB\left(\beta_1, \frac{\beta_1}{\beta_1 + q_i}\right), \tag{23}$$

where  $NB$  denotes the Negative Binomial distribution. We see now that the distributions of  $I_i$  and  $N_i$  will be “close” to each other, and hence that condition  $C4$  is approximately met, provided that  $\frac{q_i}{\beta}$  is small enough such that higher order terms can be neglected. It appears that in practical situations this condition is often fulfilled; see Crédit Suisse Financial Products (1997). Indeed, we have that

$$\Pr[N_i = 0] = \left( \frac{\beta_1}{\beta_1 + q_i} \right)^{\beta_1} \approx 1 - q_i = \Pr[I_i = 0], \tag{24}$$

while

$$\Pr[N_i = 1] = \beta_1 \left( \frac{\beta_1}{\beta_1 + q_i} \right)^{\beta_1} \left( 1 - \frac{\beta_1}{\beta_1 + q_i} \right) \approx q_i = \Pr[I_i = 1]. \tag{25}$$

Note that  $\text{var}[I_i] = q_i(1 - q_i) \leq \text{var}[N_i] = q_i\left(1 + \frac{q_i}{\beta}\right)$ .

In order to assess condition  $C3$  we note that for  $i \neq j$ , the pairwise correlations in the approximated model are given by

$$\text{cor}[N_i, N_j] = \frac{\sqrt{q_i q_j}}{\beta_1} \frac{1}{\sqrt{1 + \frac{q_i}{\beta}} \sqrt{1 + \frac{q_j}{\beta}}} \approx \frac{\sqrt{q_i q_j}}{\beta_1}. \tag{26}$$

Hence, the approximation will perform best if the “exact” correlations  $\text{corr}[I_i, I_j]$  are approximately equal to  $\frac{\sqrt{q_i q_j}}{\beta}$ .

## 4.2 The Generalized CreditRisk<sup>+</sup> model

From (26) it follows that in the Standard CreditRisk<sup>+</sup> model risky loans with the same default probability can be grouped into a so-called cluster. A cluster is defined to contain risks with the same stochastic default probability. It follows that the risks that are present within a cluster (intra cluster) are all equally correlated with each other and also have the same correlation with the different risks from any other given cluster (inter cluster).

In order to parameterize a Generalized CreditRisk<sup>+</sup> model that (approximately) satisfies conditions  $C1-C4$  it is useful to rewrite (14) into a matrix equation.

Let us define the following vectors and matrices:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_n \end{bmatrix}, \quad \mathbf{w}_i = \begin{bmatrix} w_{1,i} & \dots & w_{K,i} \end{bmatrix}^T.$$

Furthermore we introduce the matrices  $\mathbf{Q}$  and  $\mathbf{c}$  as follows:

$$\mathbf{Q} := \begin{bmatrix} q_1 & & \\ & \ddots & \\ & & q_n \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} c_1 & \dots & c_n \end{bmatrix}^T. \quad (27)$$

Finally, we consider the following unit-vector,

$$\mathbf{1} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}^T. \quad (28)$$

We find now from (12) that the final expression for the covariance structure of  $\mathbf{N}$  is given by:

$$\text{cov}[\mathbf{N}] = \mathbf{Q}\mathbf{W}^T\mathbf{\Sigma}\mathbf{W}\mathbf{Q} + \mathbf{Q}. \quad (29)$$

For the variance of  $S^*$  we find from (29) the following expression:

$$\text{var}[S^*] = \mathbf{c}^T (\mathbf{Q}\mathbf{W}^T\mathbf{\Sigma}\mathbf{W}\mathbf{Q} + \mathbf{Q}) \mathbf{c}. \quad (30)$$

The objective of this section is to find for each of the  $i$  risks ( $i = 1, 2, \dots, n$ ) the non-negative factor loadings  $w_{i,j}$  and the  $K$  volatilities  $\sigma_k$ , ( $j = 0, 1, \dots, K$ ;  $k = 1, 2, \dots, K$ ), i.e. the matrices  $\mathbf{W}$  and  $\mathbf{\Sigma}$  of a CreditRisk<sup>+</sup> model such that the conditions  $C1 - C4$  are (approximately) met. From our definition of a cluster it follows that all the risks belonging to the same cluster will be assigned the same parameters  $w_{i,j}$ .

Let  $N$  denote the number of clusters. For each  $1 \leq i \leq N$ , denote the  $i^{\text{th}}$  column of the identity matrix  $I_N$  by the vector  $\mathbf{e}_i \in \mathbb{R}^N$ , i.e.

$$\mathbf{e}_1 := \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T, \quad \dots, \quad \mathbf{e}_N := \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix}^T.$$

Denote by  $\mathbf{Cl} \in \mathbb{N}^n$  the vector such that  $\mathbf{Cl}_i \in \{1, \dots, N\}$  is the cluster number to which risk  $i$  belongs. Define the following matrix  $\mathbf{E} \in \mathbb{R}^{N \times n}$ :

$$\mathbf{E} := \begin{bmatrix} \mathbf{e}_{\mathbf{Cl}_1} & \mathbf{e}_{\mathbf{Cl}_2} & \dots & \mathbf{e}_{\mathbf{Cl}_n} \end{bmatrix}. \quad (31)$$

Note that  $\mathbf{E}^T \mathbf{E} = \mathbf{I}_n$ . Define the matrix  $\mathbf{C} \in \mathbb{R}^{N \times N}$  such that  $\mathbf{C}_{i,j}$  is equal to the (given) correlation between risks of cluster  $k$  and risks of cluster  $l$ . We will assume that  $\mathbf{C}$  is non-negative and positive semi-definite. The non-negativity assumption  $\mathbf{C}_{ij} \geq 0$  reflects that different credit risks are, at least to some extent, subject to some common economic factors and these drive the correlations to be non-negative. As we will see it will also make sure that the matrix  $\mathbf{W}$  of factor loadings will be non-negative, which is strictly needed in order to have non-negative intensities  $\mathcal{R}_i$  in (6) necessary to construct any CreditRisk<sup>+</sup> model. The semi-definiteness assumption  $\mathbf{v}^T \mathbf{C} \mathbf{v} \geq 0, \forall \mathbf{v} \neq 0$  ensures that the variance of any portfolio is non-negative.

For any pair  $i, j$  with  $i \neq j$  we have that

$$\text{corr}(I_i, I_j) = \mathbf{C}_{\mathbf{Cl}_i, \mathbf{Cl}_j} = \mathbf{e}_{\mathbf{Cl}_i}^T \mathbf{C} \mathbf{e}_{\mathbf{Cl}_j}. \quad (32)$$

Note that the diagonal elements of  $\mathbf{C}$  are not equal to 1 because they represent the average correlation between two different risks which are in the same cluster, and not the correlation between a risk and itself. As a consequence,  $\mathbf{C}$  is strictly speaking not a correlation matrix. By taking this into account, (32) can be rewritten as follows:

$$\text{corr}[\mathbf{I}] = \mathbf{E}^T \mathbf{C} \mathbf{E} + \mathbf{I}_n - \text{diag}[\mathbf{E}^T \mathbf{C} \mathbf{E}]. \quad (33)$$

It follows that:

$$\begin{aligned} \text{cov}[\mathbf{I}] &= [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} \mathbf{E}^T \mathbf{C} \mathbf{E} [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} \\ &\quad + [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} (\mathbf{I}_n - \text{diag}[\mathbf{E}^T \mathbf{C} \mathbf{E}]) [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2}. \end{aligned} \quad (34)$$

Because default probabilities and default correlations are small,

$$[\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} (\mathbf{I}_n - \text{diag}[\mathbf{E}^T \mathbf{C} \mathbf{E}]) [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} \approx \mathbf{Q}. \quad (35)$$

Note that this approximation only affects the diagonal part of  $\text{cov}[\mathbf{I}]$ . By using this approximation and by equating (29) and (34) we obtain the following

equation:

$$[\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} \mathbf{E}^T \mathbf{C} \mathbf{E} [\mathbf{Q}(\mathbf{I} - \mathbf{Q})]^{1/2} = \mathbf{Q} \mathbf{W}^T \mathbf{\Sigma} \mathbf{W} \mathbf{Q} \quad (36)$$

and it remains to determine  $\mathbf{W}$  and  $\mathbf{\Sigma}$ . Let us denote by  $\hat{\mathbf{Q}} \in \mathbb{R}^{N \times N}$  the diagonal matrix that contains the (expected) default probability per cluster:

$$\hat{\mathbf{Q}}_{i,i} = q_j, \quad \forall j \text{ such that } \mathbf{C} \mathbf{l}_j = i.$$

In order to find  $\mathbf{W}$  and  $\mathbf{\Sigma}$  that satisfy (36), let us assume that there exists a matrix  $\mathbf{W}_0 \in \mathbb{R}^{N \times K}$  with only non-negative elements such that

$$\mathbf{C} = \mathbf{W}_0 \mathbf{W}_0^T. \quad (37)$$

Obtaining such a factorization, which requires  $\mathbf{C}$  to be non-negative, is by no way trivial and will be developed in Section 5. One obtains now from (36) that

$$\begin{aligned} \mathbf{W}^T \mathbf{\Sigma} \mathbf{W} &= \mathbf{Q}^{-1/2} (\mathbf{I} - \mathbf{Q})^{1/2} \mathbf{E}^T \mathbf{C} \mathbf{E} \mathbf{Q}^{-1/2} (\mathbf{I} - \mathbf{Q})^{1/2} \\ &= \mathbf{E}^T \hat{\mathbf{Q}}^{-1/2} (\mathbf{I} - \hat{\mathbf{Q}})^{1/2} \mathbf{W}_0 \mathbf{W}_0^T \hat{\mathbf{Q}}^{-1/2} (\mathbf{I} - \hat{\mathbf{Q}})^{1/2} \mathbf{E} \\ &= \mathbf{W}_1^T \mathbf{W}_1 \end{aligned} \quad (38)$$

with  $\mathbf{W}_1 \in \mathbb{R}^{K \times n}$  a matrix with only non-negative elements given by

$$\mathbf{W}_1 = \mathbf{W}_0^T (\mathbf{I} - \hat{\mathbf{Q}})^{1/2} \hat{\mathbf{Q}}^{-1/2} \mathbf{E}. \quad (39)$$

Let us now determine  $\mathbf{W}$  and  $\mathbf{\Sigma}$ . Remember that the  $\mathbf{W}$  matrix must only have non-negative elements smaller than 1. In order to correct this in  $\mathbf{W}_1$ , we propose the following solution. Define

$$\sigma := \max_{i,j} \mathbf{W}_1(i,j), \quad \mathbf{W} := \frac{1}{\sigma} \mathbf{W}_1, \quad (40)$$

$$\mathbf{\Sigma} := \sigma^2 \mathbf{I}_K, \quad w_0 := \mathbf{1}_n - \mathbf{W}^T \mathbf{1}_K \quad (41)$$

where  $\mathbf{I}_K$  denotes the identity matrix of dimension  $K$  and  $\mathbf{1}_n$  is the column vector of dimension  $n$  with all its elements equal to 1. It is clear that

$$\mathbf{W}^T \mathbf{\Sigma} \mathbf{W} = \mathbf{W}_1^T \mathbf{W}_1. \quad (42)$$

Moreover, from (40), (39) and (37) it follows that the matrix  $\mathbf{W}$  is non-negative, as required, and (17) is satisfied as well. From these matrices, the coefficients of the Multifactor CreditRisk<sup>+</sup> model with  $K$  idiosyncratic factors are the following:

- (1) The Gamma variables  $\Gamma_i$  ( $i = 1, 2, \dots, K$ ) have the same variance  $\sigma^2$  and expectation 1.

- (2) Each coefficient for the idiosyncratic risk is equal to the corresponding coefficient in the vector  $w_0$ :

$$w_{0,i} := w_0(i).$$

- (3) For any  $i \in \{1, \dots, n\}$  and  $k \in \{1, \dots, K\}$ ,

$$w_{k,i} := \mathbf{W}(k, i).$$

From the preceding discussion, it follows that with this choice conditions  $C1$  -  $C3$  are approximately fulfilled and a similar reasoning as in Section 4.1 shows that also condition  $C4$  is approximately met.

**Remark 4.1** *Note that the computations have to be done only once. Indeed, once the matrix*

$$\hat{\mathbf{W}} := \mathbf{W}_0^T (\mathbf{I} - \hat{\mathbf{Q}})^{1/2} \hat{\mathbf{Q}}^{-1/2}$$

*has been computed, the  $\mathbf{W}$ -matrix for a given portfolio is just a copy of columns of  $\hat{\mathbf{W}}$  that correspond to the clusters of each risk of the portfolio.*

## 5 Non-Negative Matrix Factorization (NNMF)

It follows from (37) that a key step of the proposed model consists in factorizing the matrix  $\mathbf{C}$  into a product of two non-negative matrices and this will be discussed in this Section.

We will present the *Non-Negative Matrix Factorization (NNMF)* problem. In fact, this is a more general problem than we actually have but when adding a symmetry constraint it reduces to our original problem. The symmetry issue will be considered in section 6.

In this section, we will need the following notation. The Hadamard product  $\mathbf{A} \circ \mathbf{B}$  and Hadamard division  $\frac{[\mathbf{A}]}{[\mathbf{B}]}$  of the  $m \times n$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ , are the  $m \times n$  matrices with elements  $\mathbf{A}_{i,j} \mathbf{B}_{i,j}$  and  $\mathbf{A}_{i,j} / \mathbf{B}_{i,j}$ , respectively.

The NNMF problem imposes only non-negativity conditions on the factors (i.e.  $\mathbf{C} \approx \sum_{i=1}^k u_i v_i^T$ ,  $u_i, v_i \geq 0$ ) and can be stated as follows:

**NNMF:** *Given a non-negative ( $m \times n$ ) matrix  $\mathbf{C}$ , find two non-negative matrices  $\mathbf{U}(m \times p)$  and  $\mathbf{V}(p \times n)$  that minimize  $F(\mathbf{C}, \mathbf{UV})$ , where  $F$  is a cost function defining the “nearness” between two matrices.*

The choice of  $F$  will impact on how to find the minimizers, and hence, the solution of the minimization problem. The most popular choice is the Frobenius norm (or the *Euclidean Distance*)

$$F(\mathbf{C}, \mathbf{UV}) = \frac{1}{2} \|\mathbf{C} - \mathbf{UV}\|^2 = \frac{1}{2} \sum_{ij} (\mathbf{C}_{ij} - [\mathbf{UV}]_{ij})^2 = \frac{1}{2} \sum_{ij} [\mathbf{C} - \mathbf{UV}]_{ij}^2.$$

Lee and Seung (1999) propose a simple algorithm for NNMF which is based on multiplicative updating rules related to the cost function above. We also derive a similar algorithm for the problem of *Weighted Non-Negative Matrix Factorization (WNNMF)* which minimizes the *Weighted Euclidean Distance*

$$F_{\mathbf{K}}(\mathbf{C}, \mathbf{UV}) = \|\mathbf{C} - \mathbf{UV}\|_{\mathbf{K}}^2 = \frac{1}{2} \sum_{ij} [\mathbf{K} \circ (\mathbf{C} - \mathbf{UV}) \circ (\mathbf{C} - \mathbf{UV})]_{ij}. \quad (43)$$

Note that for solving the NNMF, there is also another popular choice of cost function which is the *Kullback-Leibler Divergence*; see Lee and Seung (1999, 2000).

We first review some known information and results about NNMF.

### 5.1 Gradient Information

Although  $\|\mathbf{C} - \mathbf{UV}\|^2$  is convex in  $\mathbf{U}$  and in  $\mathbf{V}$ , it is not convex in the two at the same time. One can show that there are many local minimizers. A simple strategy is to alternately minimize one of the matrices ( $\mathbf{U}$  or  $\mathbf{V}$ ) while keeping the other one fixed. Upon convergence of such an iteration, both gradients versus  $\mathbf{U}$  and  $\mathbf{V}$  are zero, and hence a stationary point is reached. Typically, this will be a local minimizer. One can explicitly obtain the gradients:

$$\nabla_{\mathbf{U}} \|\mathbf{C} - \mathbf{UV}\|^2 = -(\mathbf{C} - \mathbf{UV})\mathbf{V}^T, \quad \nabla_{\mathbf{V}} \|\mathbf{C} - \mathbf{UV}\|^2 = -\mathbf{U}^T(\mathbf{C} - \mathbf{UV}).$$

The Kuhn-Tucker optimality conditions for the constrained optimization problem are:

$$\mathbf{U} \circ (\mathbf{C}\mathbf{V}^T - \mathbf{UV}\mathbf{V}^T) = 0, \quad \mathbf{V} \circ (\mathbf{U}^T\mathbf{C} - \mathbf{U}^T\mathbf{UV}) = 0.$$

Additional properties of the stationary points are mentioned in Castronovo *et al.* (2004).

## 5.2 Multiplicative updating rules

In Lee and Seung (1999), algorithms for NNMF are proposed that are based on a number of multiplicative updating rules described in the following theorems:

**Theorem 1** *The Euclidean distance  $\|\mathbf{C} - \mathbf{UV}\|^2$  is non-increasing under the updating rules:*

$$\mathbf{V} \leftarrow \mathbf{V} \circ \frac{[\mathbf{U}^T \mathbf{C}]}{[\mathbf{U}^T \mathbf{UV}]}, \quad \mathbf{U} \leftarrow \mathbf{U} \circ \frac{[\mathbf{C} \mathbf{V}^T]}{[\mathbf{UV} \mathbf{V}^T]}. \quad (44)$$

The proof of this theorem can be found in Lee and Seung (2000) and will be extended for the weighted cases in Theorem 3. Note also that the above updating rules are the same as in Lee and Seung (1999, 2000) but are rewritten into matrix form using the Hadamard product and Hadamard division, in order to allow an easy comparison with the updating rules for the weighted cases.

In order to derive properties of these updating rules we need a simple lemma.

**Lemma 2** *Let  $\mathbf{A}$  be a symmetric non-negative matrix and  $\mathbf{v}$  be a positive vector. Let  $\mathbf{D}_x = \text{diag}(\mathbf{x})$  denote a diagonal matrix with the entries of  $\mathbf{x}$  on diagonal. Then the matrix  $\hat{\mathbf{A}} = \text{diag}\left(\frac{[\mathbf{A}\mathbf{v}]}{[\mathbf{v}]}\right) - \mathbf{A}$  is positive semi-definite.*

**Proof 1** *It is easy to see that  $\text{diag}\left(\frac{[\mathbf{A}\mathbf{v}]}{[\mathbf{v}]}\right) = \mathbf{D}_v^{-1} \mathbf{D}_{\mathbf{A}\mathbf{v}}$ . The scaled version of  $\hat{\mathbf{A}}$ :  $\hat{\mathbf{A}}_s = \mathbf{D}_v \hat{\mathbf{A}} \mathbf{D}_v = \mathbf{D}_{\mathbf{A}\mathbf{v}} \mathbf{D}_v - \mathbf{D}_v \mathbf{A} \mathbf{D}_v$  is then easily seen to be diagonally dominant, i.e. each diagonal element is positive and larger than or equal to the sum of the absolute values of the off-diagonal elements in the same row. It is well known that such matrices are also positive semi-definite; see Horn and Johnson (1999). Hence,  $\hat{\mathbf{A}}$  is also positive semi-definite.*

The following theorem then generalizes Theorem 1 to the weighted case:

**Theorem 3** *The weighted Euclidean distance  $\|\mathbf{C} - \mathbf{UV}\|_{\mathbf{K}}^2$  is non-increasing under the updating rules:*

$$\mathbf{V} \leftarrow \mathbf{V} \circ \frac{[\mathbf{U}^T (\mathbf{K} \circ \mathbf{C})]}{[\mathbf{U}^T (\mathbf{K} \circ (\mathbf{UV}))]}, \quad \mathbf{U} \leftarrow \mathbf{U} \circ \frac{[(\mathbf{K} \circ \mathbf{C}) \mathbf{V}^T]}{[(\mathbf{K} \circ (\mathbf{UV})) \mathbf{V}^T]}. \quad (45)$$

**Proof 2** *We only treat the updating rule for  $\mathbf{V}$ , the other can be proven similarly.*

*First, we split the cost function column by column and consider the partial cost function of each column of  $\mathbf{C}$ ,  $\mathbf{V}$  and  $\mathbf{K}$ , which we denoted by  $\mathbf{c}$ ,  $\mathbf{v}$  and*

$\mathbf{k}$  respectively:

$$F(\mathbf{v}) = F_q(\mathbf{c}, \mathbf{U}\mathbf{v}) = \frac{1}{2} \sum_i (\mathbf{k}_i (\mathbf{c}_i - [\mathbf{U}\mathbf{v}]_i)^2) = \frac{1}{2} (\mathbf{c} - \mathbf{U}\mathbf{v})^T \mathbf{D}_k (\mathbf{c} - \mathbf{U}\mathbf{v}) \quad (46)$$

where  $\mathbf{D}_k = \text{diag}(\mathbf{k})$ . Let  $\mathbf{v}^t$  be the current approximation of the minimizer of  $F(\mathbf{v})$  then one can rewrite  $F$  in the following quadratic form:

$$F(\mathbf{v}) = F(\mathbf{v}^t) + (\mathbf{v} - \mathbf{v}^t)^T \nabla F(\mathbf{v}^t) + \frac{1}{2} (\mathbf{v} - \mathbf{v}^t)^T \mathbf{U}^T \mathbf{D}_k \mathbf{U} (\mathbf{v} - \mathbf{v}^t) \quad (47)$$

where the gradient  $\nabla F(\mathbf{v}^t)$  can be explicitly written as

$$\nabla F(\mathbf{v}^t) = -\mathbf{U}^T \mathbf{D}_k (\mathbf{c} - \mathbf{U}\mathbf{v}^t). \quad (48)$$

Next, we approximate  $F$  by a simpler quadratic model (see Figure 2):

Fig. 1. Error minimization using auxiliary function

$$G(\mathbf{v}, \mathbf{v}^t) = F(\mathbf{v}^t) + (\mathbf{v} - \mathbf{v}^t)^T \nabla F(\mathbf{v}^t) + \frac{1}{2} (\mathbf{v} - \mathbf{v}^t)^T \mathbf{D}(\mathbf{v}^t) (\mathbf{v} - \mathbf{v}^t) \quad (49)$$

where  $G(\mathbf{v}^t, \mathbf{v}^t) = F(\mathbf{v}^t)$  and  $\mathbf{D}(\mathbf{v}^t)$  is a diagonal matrix chosen to make  $\mathbf{D}(\mathbf{v}^t) - \mathbf{U}^T \mathbf{D}_k \mathbf{U}$  positive definite implying that  $G(\mathbf{v}, \mathbf{v}^t) - F(\mathbf{v}) \geq 0, \forall \mathbf{v}$ . The choice for  $\mathbf{D}(\mathbf{v}^t)$  is similar to that of the Lee and Seung algorithm:

$$\mathbf{D}(\mathbf{v}^t) = \text{diag} \left( \frac{[\mathbf{U}^T \mathbf{D}_k \mathbf{U} \mathbf{v}^t]}{[\mathbf{v}^t]} \right). \quad (50)$$

Lemma 2 assures the positive semi-definiteness of  $\mathbf{D}(\mathbf{v}^t) - \mathbf{U}^T \mathbf{D}_k \mathbf{U}$ . As a result, we have

$$F(\mathbf{v}^t) = G(\mathbf{v}^t, \mathbf{v}^t) \geq \min_{\mathbf{v}} G(\mathbf{v}, \mathbf{v}^t) = G(\mathbf{v}^{t+1}, \mathbf{v}^t) \geq F(\mathbf{v}^{t+1}) \quad (51)$$

where  $\mathbf{v}^{t+1}$  is found by solving  $\frac{\partial G(\mathbf{v}, \mathbf{v}^t)}{\partial \mathbf{v}} = 0$ :

$$\mathbf{v}^{t+1} = \mathbf{v}^t - \mathbf{D}(\mathbf{v}^t)^{-1} \nabla F(\mathbf{v}^t) \quad (52)$$

$$= \mathbf{v}^t + \text{diag} \left( \frac{[\mathbf{v}^t]}{[\mathbf{U}^T \mathbf{D}_k \mathbf{U} \mathbf{v}^t]} \right) \mathbf{U}^T \mathbf{D}_k (\mathbf{c} - \mathbf{U} \mathbf{v}^t) \quad (53)$$

$$= \mathbf{v}^t + \mathbf{v}^t \circ \frac{[\mathbf{U}^T \mathbf{D}_k (\mathbf{c} - \mathbf{U} \mathbf{v}^t)]}{[\mathbf{U}^T \mathbf{D}_k \mathbf{U} \mathbf{v}^t]} \quad (54)$$

$$= \mathbf{v}^t \circ \frac{[\mathbf{U}^T \mathbf{D}_k \mathbf{c}]}{[\mathbf{U}^T \mathbf{D}_k \mathbf{U} \mathbf{v}^t]} \quad (55)$$

$$= \mathbf{v}^t \circ \frac{[\mathbf{U}^T (\mathbf{k} \circ \mathbf{c})]}{[\mathbf{U}^T (\mathbf{k} \circ (\mathbf{U} \mathbf{v}^t))]} \quad (56)$$

Putting together the updating rules for all the columns of  $\mathbf{V}$  gives the desired updating rule for the whole matrix  $\mathbf{V}$  as in (45). The relation (51) shows that the weighted Euclidean distance is non-increasing under the update rule for  $\mathbf{V}$ .

**Remark 1:** The unweighted case can be considered as a special case of the weighted case, where all weights are set to 1. Therefore, the above proof also holds for the unweighted case.

**Remark 2:** The best  $\text{rank} - k$  approximation is not necessarily a sub-matrix of the best  $\text{rank} - (k + 1)$  approximation.

**Remark 3:** The non-negativity constraint on the matrices  $\mathbf{U}$  and  $\mathbf{V}$  is automatically satisfied by these updating rules if the starting matrices  $U_0$  and  $V_0$  are non-negative.

**Remark 4:** In order to prevent *zero-division* exceptions during the execution of the algorithm, we replace in practice the above updating rules by the following ones:

$$\mathbf{V} \leftarrow \mathbf{V} \circ \frac{[\mathbf{U}^T(\mathbf{K} \circ \mathbf{C})]}{[\mathbf{U}^T(\mathbf{K} \circ (\mathbf{U}\mathbf{V}))] + \epsilon \mathbf{1}_{k \times n}}, \quad \mathbf{U} \leftarrow \mathbf{U} \circ \frac{[(\mathbf{K} \circ \mathbf{C})\mathbf{V}^T]}{[(\mathbf{K} \circ (\mathbf{U}\mathbf{V}))\mathbf{V}^T] + \epsilon \mathbf{1}_{m \times k}}, \quad (57)$$

where  $\epsilon$  is a small positive constant.

**Remark 5:** The fixed point obtained from the above algorithm may not be a stationary point of the cost function but a simple modification of the above algorithm can fix this, as pointed out in Lin (2005). Since the problem is not convex, the approximation error plays a more important role than the type of fixed point. Therefore, it is useful to restart the algorithm several times to have the best possible approximation.

## 6 Symmetric Non-Negative Matrix Factorization (SNNMF)

In the previous section, we have introduced a method for the NNMF problem. In this section, we will modify this method for symmetric non-negative matrices to produce the desired factorization. Formally, the problem we are dealing with is the following:

$$(SNNMF) \quad \text{minimize } \|\mathbf{C} - \mathbf{U}\mathbf{U}^T\|_{\mathbf{K}}^2 \quad \text{subject to } \mathbf{U} \in \mathbb{R}_+^{n \times k},$$

where  $\mathbf{C}$  and  $\mathbf{K}$  are symmetric non-negative matrices.

We call this a *Symmetric Non-Negative Matrix Factorization (SNNMF)*, and it is obviously a special case of NNMF that approximates the matrix  $\mathbf{C}$  by a *completely positive (cp)* matrix  $\mathbf{U}\mathbf{U}^T$  (Barioli and Berman, 2003). We cite here some facts about *completely positive* matrices:

- A real non-negative matrix  $\mathbf{C}$  which can be decomposed as  $\mathbf{U}\mathbf{U}^T$  ( $\mathbf{U} \geq 0$ ) is a *completely positive* matrix.
- The smallest number of columns of  $\mathbf{U} \geq 0$  such that  $\mathbf{C} = \mathbf{U}\mathbf{U}^T$  is called the *completely positive rank (cp-rank)* of  $\mathbf{C}$ .
- No polynomial time algorithms are known yet to address the following questions. *Is a given matrix completely positive?* and *What is the cp-rank of a given matrix?*

Therefore, it is not realistic to solve the true SNNMF problem but one needs to consider a relaxed version of it. Our approach is to use the NNMF problem to produce acceptable results for the SNNMF problem.

First we will see if the fixed points of NNMF for symmetric matrices might

be symmetric as well. Then we show how to use the NNMF algorithm for obtaining symmetric approximations. We treat these separately in the following paragraphs.

### 6.1 Fixed point of NNMF for symmetric matrices

It appears that there are no efficient methods yet to solve the (SNNMF) problem. Even if one uses the NNMF algorithm to look for an  $\mathbf{UV}$  approximation of a symmetric matrix  $\mathbf{C}$ , nothing guarantees that  $\mathbf{UV}$  is also symmetric. Some interesting results are to be found in Catral et al. (2004) and are related to symmetry of the fixed points of NNMF. But these results deal only with the simple cases where the fixed points of NNMF have zero gradients. For other cases, there is still no guarantee for symmetry. We should therefore modify the NNMF algorithm in order to favor symmetric fixed points. In the next subsection, we derive one such particular way to modify the NNMF algorithm.

### 6.2 Fixed point symmetrization

Our strategy is to *drive* the NNMF algorithm to converge to a fixed point of SNNMF *if possible*. We start the NNMF algorithm with an asymmetric starting point (i.e.  $\mathbf{U}_0 \mathbf{V}_0$ ), and then force the two factors (i.e.  $\mathbf{U}_k$  and  $\mathbf{V}_k$ ) to be equal by minimizing their normed difference. One way to do it is to use the following modified cost function for a symmetric matrix  $\mathbf{C}$ :

$$F_s(\mathbf{C}, \mathbf{UV}) = \|\mathbf{C} - \mathbf{UV}\|_{\mathbf{K}}^2 + \alpha \|\mathbf{U}^T - \mathbf{V}\|^2, \quad (58)$$

where the earlier part ( $\|\mathbf{C} - \mathbf{UV}\|_{\mathbf{K}}^2$ ) is called the approximation error and the latter part ( $\|\mathbf{U}^T - \mathbf{V}\|^2$ ) is called the asymmetry penalty.

Incorporating additional linear or quadratic costs with the Euclidean Distance does not change significantly the updating rules. Using the same reasoning as presented in the theorems 1 and 3, one can obtain the following updating rules for the modified cost function:

$$\mathbf{V} \leftarrow \max \left( \mathbf{V} \circ \frac{[\mathbf{U}^T(\mathbf{K} \circ \mathbf{C}) + \alpha(\mathbf{V} - \mathbf{U}^T)]}{[\mathbf{U}^T(\mathbf{K} \circ (\mathbf{UV})) + \epsilon \mathbf{1}_{k \times n}]}, 0 \right), \quad (59)$$

$$\mathbf{U} \leftarrow \max \left( \mathbf{U} \circ \frac{[(\mathbf{K} \circ \mathbf{C})\mathbf{V}^T + \alpha(\mathbf{U} - \mathbf{V}^T)]}{[(\mathbf{K} \circ (\mathbf{UV}))\mathbf{V}^T + \epsilon \mathbf{1}_{m \times k}]}, 0 \right), \quad (60)$$

where the  $\max$  function is introduced to keep the factors non-negative. Note that to improve the search, we have to apply a balancing step after each update of  $\mathbf{U}$  or  $\mathbf{V}$ . The step consists in balancing the norm of each column of  $\mathbf{U}$  and that of the corresponding row of  $\mathbf{V}$

$$\mathbf{D}_{ii} = \frac{\sqrt{\|\mathbf{V}_{i,:}\|}}{\sqrt{\|\mathbf{U}_{:,i}\|}} \quad \mathbf{U} = \mathbf{U}\mathbf{D} \quad \mathbf{V} = \mathbf{D}^{-1}\mathbf{V}. \quad (61)$$

The parameter  $\alpha$  appears as the trade-off between the approximation error and the symmetry of the fixed point. The right choices of  $\alpha$  can lead the search direction towards the symmetric fixed points. Overestimating  $\alpha$  could lead to a symmetric factorization without taking into account the approximation error.

An appropriate value of  $\alpha$  is believed to depend on the best possible approximation error, but this error is not known in general. Experiments show that the range of possible values of  $\alpha$  is quite large, so that, one can restart the algorithm with a better value of  $\alpha$  determined from the previous results of the algorithm.

It is also useful to note that, after all, the results might not be in the desired form in which  $\mathbf{U} = \mathbf{V}^T$ . But by having forced the symmetry, one can expect that the error  $\|\mathbf{C} - \mathbf{U}\mathbf{U}^T\|^2$  and/or  $\|\mathbf{C} - \mathbf{V}^T\mathbf{V}\|^2$  are not too far from the real approximation error  $\|\mathbf{C} - \mathbf{U}\mathbf{V}\|^2$ .

## 7 Numerical Results

The proposed CreditRisk<sup>+</sup> model has been applied to real data. First, the cluster correlation matrix  $\mathbf{C}$  has been computed from historical asset data. In order to obtain a low-rank non-negative approximation of the correlation matrix, the SNNMF algorithm has been used for increasing values of the number  $K$  of systematic risks. This number  $K$  is also equal to the rank of the non-negative factorization of  $\mathbf{C}$ . The error between the approximated non-negative factorization and  $\mathbf{C}$  is compared with the error obtained without imposing the non-negativity of the factorization (i.e. by using the SVD decomposition).

Fig. 2. Low Rank Approximation of the Correlation Matrix

As we can see, the SNNMF algorithm performs very well in practice. With

this approach, using 10 systematic risks instead of using a single factor approach allowed to reduce the error on the correlations by a factor 3, while still preserving the numerical tractability of the model. We believe that this is a substantial improvement.

Note that the SNNMF algorithm can be used in any problem where the entries of the matrix  $\mathbf{C}$  or any product of the type  $\mathbf{v}^T \mathbf{C} \mathbf{v}$  have to be evaluated frequently, since the complexity for this is substantially reduced when using the low rank approximation, while the accuracy is still very good.

## 8 Concluding Remarks

The Credit Risk<sup>+</sup> model is one of the industry standards to estimate the risk of a portfolio of credit exposures. In practical situations one often only has the individual credit risk characteristics and the default correlations available and this makes the parameterization a difficult exercise. In this paper a solution has been proposed that enables the parameterization whilst preserving the characteristics of the portfolio.

## Acknowledgements

This paper presents research supported by the Concerted Research Action (ARC) "Large Graphs and Networks" of the French Community of Belgium, and by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the authors. A research fellowship from FRIA is also gratefully acknowledged by the second author Ngoc-Diep Ho. Steven Vanduffel acknowledges the financial support of Fortis Central Risk Management through the Fortis Chair in Financial and Actuarial Risk Management (2006-2011) as well as the financial support by the Onderzoeksfonds K.U.Leuven (GOA/02: Actuariële, financiële en statistische aspecten van afhankelijkheden in verzekerings- en financiële portefeuilles).

The authors also thank Bruno De Cleen, Ivan Goethals, Andrew Chernih, Luc Hoegaerts and Luc Henrard for interesting discussions that have led to this paper, and an anonymous referee for helpful suggestions that helped improve the presentation of the different results.

## References

- F. Barioli and A. Berman (2003). The maximal cp-rank of rank  $k$  completely positive matrices. *Linear Algebra and its Applications*, 363, 17-33.
- V. Blondel, N. D. Ho and P. Van Dooren 2005, Non-negative Matrix Factorization - Extensions and Applications, *Internal report 005-35, Cesame, Université catholique de Louvain*.
- P. Bürgisser, A. Kurth, A. Wagner (2001). Incorporating severity variations into credit risk. *Journal of Risk*, 3(4), 5-31.
- M. Catral, L. Han, M. Neumann, and R. Plemmons (2004). On reduced rank nonnegative matrix factorizations for symmetric matrices. *Linear Algebra and its Applications*, 393, 107-126.
- A. Chernih, S. Vanduffel and L. Henrard (2006). Asset Correlations: A Literature Review and analysis of the impact of dependent loss given defaults, available online at [www.defaultrisk.com](http://www.defaultrisk.com).
- Credit Suisse Financial Products (1997). "CreditRisk<sup>+</sup>, A Credit Risk Management Framework", Technical document, available online at [www.csfb.com/creditrisk](http://www.csfb.com/creditrisk)
- M. Crouhy, D. Galai and R. Mark (2000). "A comparative analysis of current credit risk models". *Journal of Banking and Finance*, 24, 57-117.
- J. Dhaene, M. Goovaerts, S. Vanduffel, R. Koch, R. Olieslagers and O. Romijn (2006), Consistent Assumptions for Modeling Credit Loss Correlations. *Journal of Actuarial Practice*, 13, 165-174.
- Embrechts P, McNeil AJ and Straumann D (2003). Correlation and dependence in risk management: properties and pitfalls . *Risk management: value at risk and beyond*, edited by Dempster M, published by Cambridge University Press, Cambridge.
- R. Frey and A. McNeil (2003). Dependent Defaults in Models of Portfolio Credit Risk. *Journal of Risk* 6(1), 59-92.
- M. Gordy (2000). A Comparative Anatomy of Credit Risk Models. *Journal of Banking and Finance*, 24(1-2), 119-149.
- M. Gordy (2002), Saddlepoint approximation of CreditRisk<sup>+</sup>. *Journal of Banking and Finance*, 26, 1335-1353;
- D. Guillaumet, J. Vitrià, and B. Schiele (2003). Introducing a weighted nonnegative matrix factorization for image classification. *Pattern Recognit. Lett.*, 24(14), 2447-2454.
- H. Haaf, O. Reiss, and J. Schoenmakers (2003). Numerically stable computation of CreditRisk<sup>+</sup>. *Technical report, Weierstrass-Institut*.
- R. A. Horn and C. R. Johnson (1999), Matrix Analysis. *Cambridge University Press*.
- N. J. Jobst & A. de Servigny (2005). An Empirical Analysis of Equity Default Swaps II: Multivariate Insights. *Working Paper, Standard & Poor's*.
- S. Kealhofer (1995) "Managing default risk in derivative portfolios". *Derivative Credit Risk: Advances in Measurement and Management. Risk Publications, London*.

- U. Koyluoglu and A. Hickman (1998) “Reconcilable differences”. *Risk* 11(10), 56-62.
- D. D. Lee and H. S. Seung (1999). Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401, 788-791.
- Daniel D. Lee and H. Sebastian Seung (2000). Algorithms for nonnegative matrix factorization. *Neural Information Processing Systems (NIPS)*, 556-562.
- C.-J. Lin (2005). On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *Information and Support Services Technical Report*, Department of Computer Science, National Taiwan University.
- JP Morgan & Co. Inc.(1997) “Credit Metrics”, *Technical Document*, available online at [www.creditmetrics.com](http://www.creditmetrics.com).
- P. Paatero (1997). A weighted nonnegative least squares algorithm for three-way ‘parafac’ factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 38, 223-242.
- P. Paatero and U. Tapper (1994). Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5, 111-126.
- H. H. Panjer and G. E. Willmot (1992). *Insurance Risk Models*. Society of Actuaries, Schaumburg, II.
- O. Reiss (2003). Fourier inversion algorithms for Generalized CreditRisk<sup>+</sup> models and an extension to incorporate market risk. *WIAS-Preprint* No. 817.
- A. de Servigny and O. Renault (2002). Default correlation: empirical evidence. *Standard & Poor’s*.
- T. Wilson (1997a) “Portfolio credit risk: Part I”, *Risk* (September), 111-117.
- T. Wilson (1997b) “Portfolio credit risk: Part II”, *Risk* (October), 56-61.