Non-Negative Matrix Factorization and Term Structure of Interest Rates

Hellinton H. Takada[a, b] and Julio M. Stern[b]

[a]Quantitative Research, Itaú Asset Management, São Paulo, Brazil
[b]Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

Abstract. Non-Negative Matrix Factorization (NNMF) is a technique for dimensionality reduction with a wide variety of applications from text mining to identification of concentrations in chemistry. NNMF deals with non-negative data and results in non-negative factors and factor loadings. Consequently, it is a natural choice when studying the term structure of interest rates. In this paper, NNMF is applied to obtain factors from the term structure of interest rates and the procedure is compared with other very popular techniques: principal component analysis and Nelson-Siegel model. The NNMF approximation for the term structure of interest rates is better in terms of fitting. From a practitioner point of view, the NNMF factors and factor loadings obtained possess straightforward financial interpretations due to their non-negativeness.

Keywords: Information theory, Entropy, Financial markets.
PACS: 89.70.-a, 89.70.Cf, 89.65.Gh

INTRODUCTION

Non-Negative Matrix Factorization (NNMF) is a multivariate data analysis technique aimed to estimate non-negative factors and factor loadings from non-negative data. NNMF was invented by Paatero and Tapper in 1994 under the name Positive Matrix Factorization (PMF) [1] and the name NNMF was established by Lee and Seung in 1999 [2]. There are several applications of NNMF and some examples from the literature are: text mining [3], image processing [4], sound processing [5], identification of concentrations in chemistry [1], recognition of underlying trends in stock market data [6], and so on. There are many algorithms for NNMF with different optimization strategies such as multiplicative update [4], gradient descent [4] or alternating least squares [1].

The term structure of interest rates or the yield curve is the relationship between interest rates or bond yields and different maturities or terms. The yield curve is important in economy and finance because it reflects current expectations of market participants about future changes in the interest rates. There are several factor models for the yield curve: Litterman and Sheinkman (1991) proposed a three factor model based on Principal Component Analysis (PCA) and suggested names for these factors: level, steepness (or slope) and curvature [7]. Since then, these factors became attributes of the yield curve. Independently, Nelson and Siegel (1987) published a parametric model for the yield curve [8] which was rewritten by Diebold and Li (2006) in terms of the yield curve attributes [9]. A plethora of models evolved from these two approaches.

Obviously, the yield curve data is non-negative by nature and, consequently, it is appropriate to use NNMF. Since a model for yield curve has the objective to give

Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2014)
AIP Conf. Proc. 1641, 369-377 (2015); doi: 10.1063/1.4906000

insights about the current expectations about the future for economic and financial analysis, the non-negativity of factors and factor loadings is desired to easy the interpretation. Given a number of factors, PCA does not minimize the approximation error. Additionally, Nelson-Siegel (NS) model presents some fitting problems and a fourth factor was included by Svensson (1994) to improve the original model [10]. Fixing the number of factors, NNMF algorithms reduce the approximation error of the factorization. Consequently, we expect a better fitting for NNMF when compared with PCA or NS. There is a large variety of divergence measures which have been applied as objective functions when minimizing the approximation error of NNMF: Bregman divergences [11], Renyi's information measure [12], Csiszar's divergences [13], Kompass' divergence [14], the α-divergence [15] or Itakura-Saito divergence [16].

The paper is organized as follows: Firstly, the PCA approach proposed by Litterman and Sheinkman (1991) and the NS parametric factor model are presented. Then, the NNMF problem is reviewed with some details. After the theory, some results using real yield curve data are presented and the obtained factors and factor loadings are compared to that from PCA and NS. Finally, the conclusion together with more comments about the results are given at the end.

FACTORS AND TERM STRUCTURE OF INTEREST RATES

Principal Component Analysis

Since the three factor model from Litterman and Sheinkman (1991), PCA has been used extensively to model the term structure of interest rates and it was verified that a large portion of bond return variation (up to 98%) can be explained by the first three principal components or factors: level, steepness and curvature. PCA is a statistical technique to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components using orthogonal transformation [17].

Singular Value Decomposition (SVD) is technique from linear algebra used to obtain the principal components [18]. Given a matrix of historical yields $Y = [y_{ij}] \in \mathbb{R}^{T\times V}$ with T days and V vertices or maturities, the SVD factorization results:

$$\hat{Y} = USP', \tag{1}$$

where $\hat{Y}$ is obtained mean centering the data matrix Y, $U = [u_{ij}] \in \mathbb{R}^{T\times T}$, $S = [s_{ij}] \in \mathbb{R}^{T\times V}$, $P = [p_{ij}] \in \mathbb{R}^{V\times V}$, $UU' = I_T$, $PP' = I_V$, the columns of U and P are orthonormal eigenvectors of YY' and S is a diagonal matrix containing the square roots of the corresponding eigenvalues from U or P such that $s_{11} \geq s_{22} \geq \cdots \geq s_{VV}$, since usually $T \geq V$.

Given $K \leq \min(T, V)$, the K-factor model using PCA is given by:

$$\hat{Y} \approx U\tilde{S}\tilde{P}', \tag{2}$$

where $\tilde{S} = [s_{ij}] \in \mathbb{R}^{T\times K}$ and $\tilde{P} = [p_{ij}] \in \mathbb{R}^{V\times K}$. The columns of $U\tilde{S}$ are called factors and the columns of $\tilde{P}$ are the corresponding factor loadings. When $K = 3$, the first column of $U\tilde{S}$ is named level, the second is the steepness and the third is the curvature.

Nelson-Siegel Model

The Nelson and Siegel proposed a parametric model in 1987 and it has been used to fit the term structure of interest rates. Using the formulation from [9], the Nelson-Siegel (NS) is given by:

$$y(t,\tau) = \beta_{t,0} + \beta_{t,1}\,[1 - \exp(-\tau/\lambda)]/(\tau/\lambda) + \beta_{t,2}\{[1 - \exp(-\tau/\lambda)]/(\tau/\lambda) - \exp(-\tau/\lambda)\}, \tag{3}$$

where τ is the maturity (usually in years), $y(t,\tau)$ is the curve yield at maturity τ and $\beta_{t,0}$, $\beta_{t,1}$ and $\beta_{t,2}$ are cross-sectional parameters to be determined for each date t. It is straightforward to notice that these parameters can be determined via a least-squares or similar algorithm [9,20]. The λ parameter can be determined to minimize the possible correlation of $\beta_{t,1}$ and $\beta_{t,2}$ over the time to avoid possible identification problems.

There are important interpretations for some of the parameters: $\beta_{t,0}$ represents the long run level of interest rates, $\beta_{t,1}$ is short-term interest rate and $\beta_{t,2}$ is the medium term component. Using the names from Litterman and Sheinkman (1991), $\beta_{t,0}$ represents the level, $\beta_{t,1}$ captures the steepness and $\beta_{t,2}$ is the curvature.

Non-Negative Matrix Factorization

NNMF seems to be suitable to model the factors behind the term structure of interest rates. Since the matrix of historical yields is nonnegative $Y = [y_{ij}] \in \mathbb{R}_{\geq 0}^{T\times V}$ and given an integer $K \leq \min(T,V)$, the NNMF problem is to find the following approximation:

$$Y \approx \tilde{Y} = FB, \tag{4}$$

where $\tilde{Y} = [\tilde{y}_{ij}] \in \mathbb{R}_{\geq 0}^{T\times V}$, $F = [f_{ij}] \in \mathbb{R}_{\geq 0}^{T\times K}$ and $B = [b_{ij}] \in \mathbb{R}_{\geq 0}^{K\times V}$. Clearly, F represents the factors and B the factor loadings.

The NNMF optimization procedures minimizes the approximation error between Y and $\tilde{Y}$. In a generalized way, the Bregman divergence $D_\varphi(Y\|\tilde{Y})$ is used as the objective function to be minimized [11,19]. Considering only separable Bregman divergences,

$$D_\varphi(Y\|\tilde{Y}) = \sum_{ij} D_\varphi(y_{ij}\|\tilde{y}_{ij}) = \sum_{ij}\{\varphi(y_{ij}) - \varphi(\tilde{y}_{ij}) - \nabla\varphi(\tilde{y}_{ij})[\varphi(y_{ij}) - \varphi(\tilde{y}_{ij})]\}, \tag{5}$$

where $\varphi(\cdot)$ is a strictly convex function with a continuous first derivative. Formally, the resulting optimization problems are:

$$\min_{F,B\geq 0} D_{\varphi}(Y\|FB) + J_F(F) + J_B(B) \quad (6)$$

or

$$\min_{F,B\geq 0} D_{\varphi}(FB\|Y) + J_F(F) + J_B(B) \quad (7)$$

where $J_F(\cdot)$ and $J_B(\cdot)$ are penalty functions to enforce certain application-dependent characteristics of the solution, such as sparsity and/or smoothness. It is also important to remember that the Bregman divergences are not symmetric in general. Here, we consider $D_{\varphi}(Y\|FB)$.

Adopting $\varphi(x) = x^2/2$ and $J_F(\cdot) = J_B(\cdot) = 0$, there are some known algorithms to solve the NNMF problem divided in general classes [21]: gradient descent algorithms, multiplicative update algorithms and alternating least squares algorithms (ALS). Actually, when $\varphi(x) = x^2/2$ the Bregman divergence is the squared Frobenius norm. For example, when $\varphi(x) = x\ln(x)$ the Bregman divergence is the Kullback-Leibler divergence.

EMPIRICAL COMPARISON OF FACTORS

In this section, the objective is to present and compare the factors and factor loading obtained from real yield curve data using PCA, NS and NNMF. The data used for the study is the Brazilian term structure of interest rates obtained from future contracts traded at BM&FBovespa. It was used data from 05/13/2003 to 10/09/2013. The vertices chosen are: 3 months, 6 months, 1 year, 2 years and 5 years. To be able to compare the different approaches and having in mind the factors named by Litterman and Sheinkman (1991), the number of factors considered here for the models under comparison is three.

The PCA was implemented using SVD, NS was obtained using an optimization procedure to minimize the least-square error and NNMF was calculated using ALS with multiple starting points. **FIGURE 1** presents the PCA factors and the corresponding factor loading are in **FIGURE 2**. Observing the PCA factor loadings, it becomes clear the origin of the names given by Litterman and Sheinkman (1991) for each factor.

The obtained NS factors and factor loadings are in **FIGURE 3** and **FIGURE 4**, respectively. The NS factor loadings also capture the idea of level, steepness and curvature. As already mentioned, the factors are estimated for each day in the sample. Comparing with the PCA factors, the NS factors over the time are not smooth with several spikes and, consequently, difficult for economic or financial interpretation.

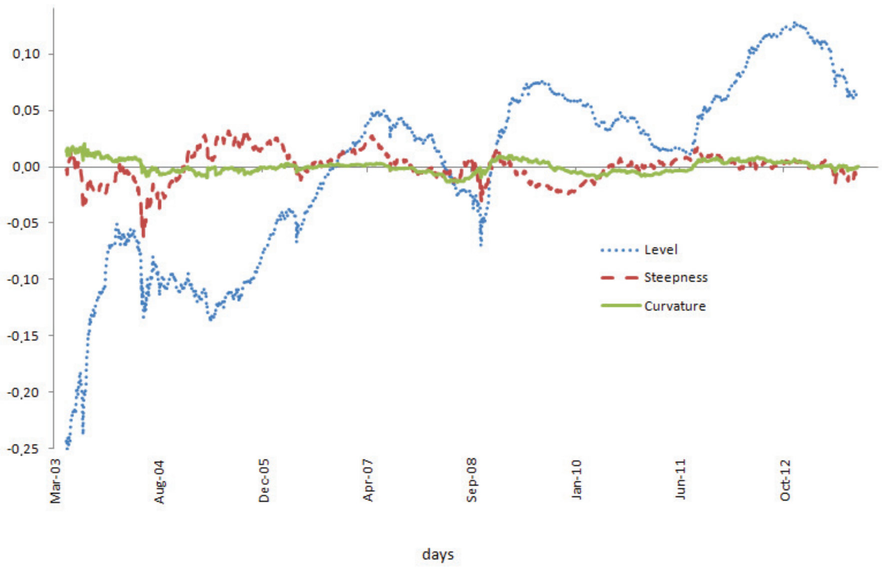

FIGURE 1. Factors for yield curve using PCA.

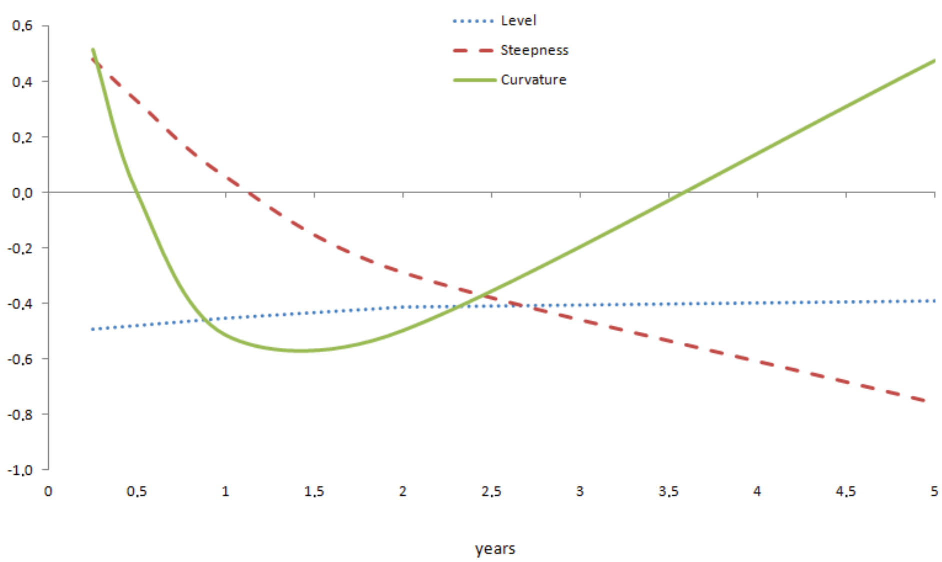

FIGURE 2. Factor loadings for yield curve using PCA.

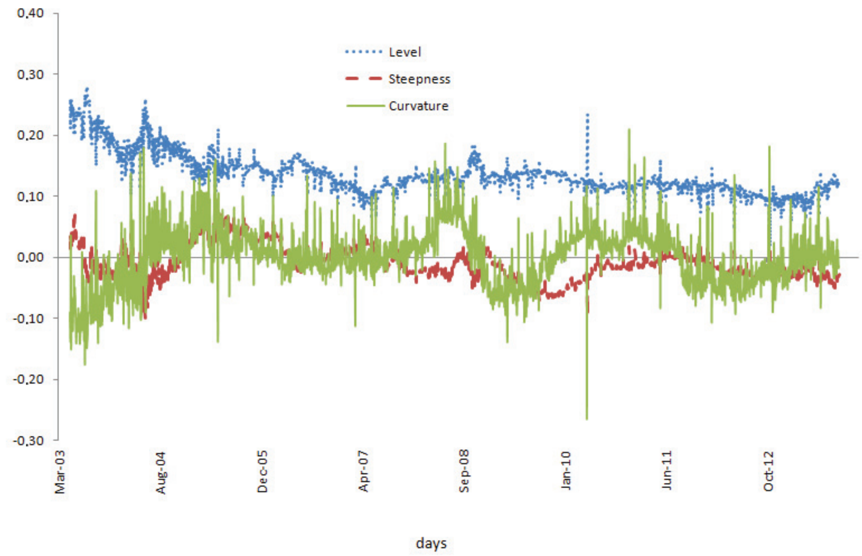

FIGURE 3. Factors for yield using NS.

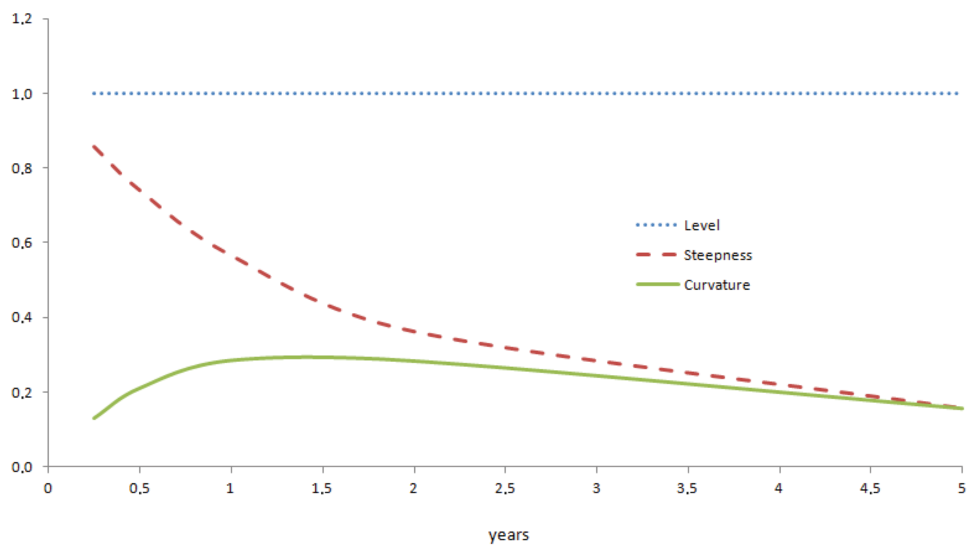

FIGURE 4. Factor loadings for yield curve using NS.

The factors and factor loadings from NNMF are in **FIGURE 5** and **FIGURE 6**, respectively. The obtained factor loadings are not the same ones pointed by Litterman and Sheinkman (1991). Actually, we identify a steepness factor and two different curvatures. The inclusion of a second curvature to model the yield curve is not new in the literature. Actually, a fourth factor is usually included in the NS framework under the name Svensson's factor to improve the data adjustment [10]. The Nelson-Siegel-Svensson model is given by:

$$y(t,\tau) = \beta_{t,0} + \beta_{t,1}\,[1-\exp(-\tau/\lambda_1)]/(\tau/\lambda_1) \\ +\beta_{t,2}\{[1-\exp(-\tau/\lambda_1)]/(\tau/\lambda_1) - \exp(-\tau/\lambda_1)\} \\ +\beta_{t,3}\{[1-\exp(-\tau/\lambda_2)]/(\tau/\lambda_2) - \exp(-\tau/\lambda_2)\} \qquad (8)$$

where $\beta_{t,3}$ is the fourth factor representing a second curvature, λ_1 and λ_2 represent the two curvatures and the remaining parameters are the same from (3). Clearly, the second curvature from NNMF brings the idea of the Svensson's factor.

The NNMF obtained factors are smooth and, consequently, it is natural to use them for investment strategies or economic/financial analysis about the future behavior of the term structure of interest rates related to its steepness and two curvatures. The non-negativeness of the factors and factor loadings implies in a easy interpretation: when a factor increases, the yield curve explanation by the corresponding factor loading also increases and *vice versa*.

Finally, the approximation error for each factor model using the Frobenius norm is in **TABLE 1**. Since the NNMF minimizes the approximation error, the data fitting is better. Obviously, the purpose of PCA is not to improve the fitting given a number a factors and the NS is not being able to adjust perfectly its parametric form to the data.

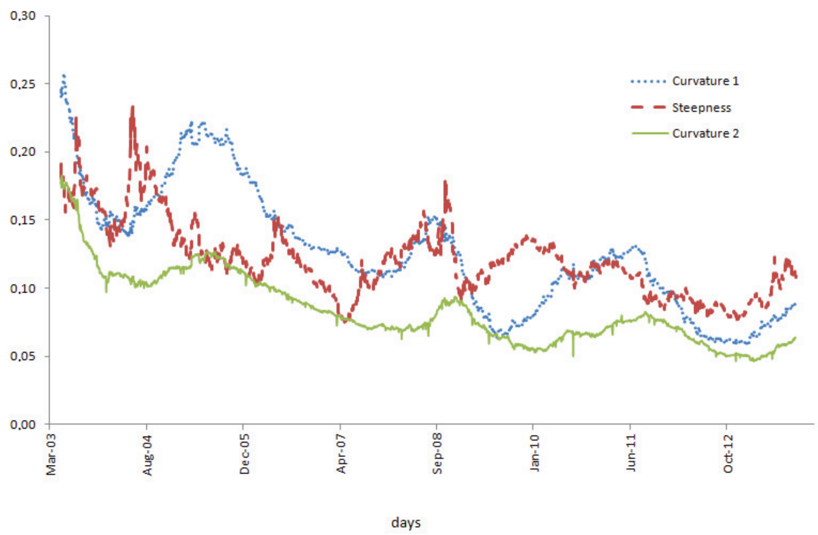

FIGURE 5. Factors for yield curve using NNMF.

TABLE 1. Approximation errors for PCA, NS and NNMF.

Factor Models	Approximation Error
PCA	5.97301
NS	16.46829
NNMF	0.00181

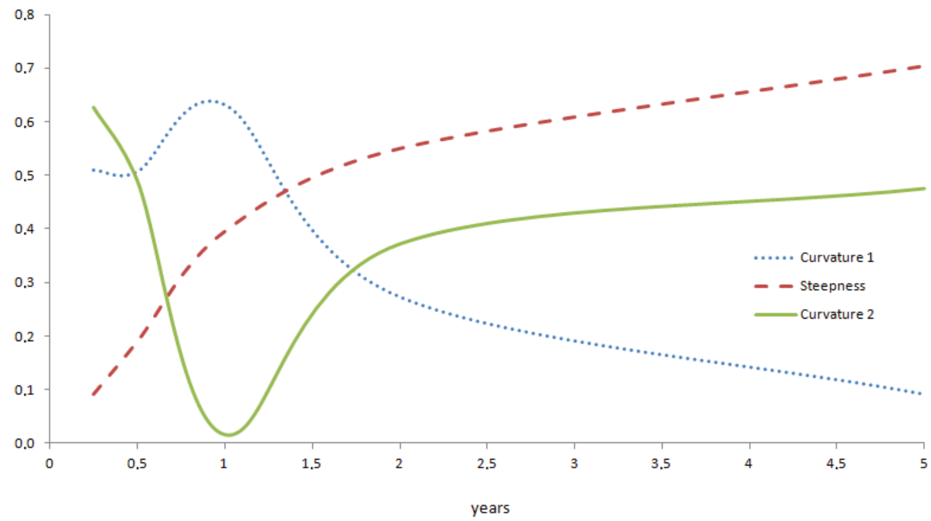

FIGURE 6. Factor loadings for yield curve using NNMF.

CONCLUSIONS

In this paper, NNMF technique was applied for the first time to find factors for yield curves. The data adjustment is far better than that obtained from usual techniques such as PCA or NS. The obtained factors and factor loadings due to their non-negativeness easy the interpretation for economic and financial analysis. Additionally, the NNMF factors are relatively smooth when compared to that from NS and they are quite different from the classical ones from Litterman and Sheinkman (1991). Actually, for the Brazilian data the three NNMF factors are identified as the steepness and two different curvatures. One of the curvatures is that one identified by Litterman and Sheinkman (1991) and the other is the called Svensson's factor. In this work, NNMF was found to be a suitable factorization model for yield curves.

REFERENCES

1. P. Paatero and U. Tapper, "Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics* **5** (2), 111–126 (1994).
2. D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature* **401**, 788–791 (1999).
3. M. W. Berry (editor), *Computational information retrieval*, Philadelphia: Society for Industrial and Applied Mathematics, 2001.
4. D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," *Advances in Neural Information Processing Systems* **13**, 556–562 (2001).
5. P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 177–180 (2003).
6. K. Drakakis, S. Rickard, R. de Fréin and A. Cichocki, "Analysis of financial data using non-negative matrix factorization," *International Mathematical Forum* **3** (38), 1853–1870 (2008).
7. R. Litterman and J. Sheinkman, "Common factors affecting bond returns," *J. Fixed Income* **1**, 62–74 (1991).

8. C. Nelson and A. F. Siegel, "Parsimonious modeling of yield curves," *J. Business* **60**, 473–489 (1987).
9. F. X. Diebold and C. Li, "Forecasting the Term Structure of Government Bond Yields," *J. Econometrics* **130**, 337–364 (2006).
10. L. E. O. Svensson, "Estimating and interpreting forward interest rates: Sweden 1992-1994," *IMF Working Paper* **94** (114), 1–49 (1994).
11. I. S. Dhillon and S. Sra, "Generalized nonnegative matrix approximations with Bregman divergences," *Advances in Neural Information Processing Systems* **18**, 283–290 (2005).
12. K. Devarajan, "Molecular pattern discovery using non-negative matrix factorization based on Renyi's information measure," *Proceedings of 20th Annual International Conference on Statistics, Combinatorics, Mathematics and Applications*, December, 2005.
13. A. Cichocki, R. Zdunek and S.-I. Amari, "Csiszar's divergences for non-negative matrix factorization: family of new algorithms," *Lecture Notes in Computer Science* **3889**, 32–39 (2006).
14. R. Kompass, "A generalized divergence measure for nonnegative matrix factorization," *Neural Computation* **19** (3), 780–791 (2007).
15. A. Cichocki, H. Y. Lee, Y. D. Kim and S. Choi, "Non-negative matrix factorization with alpha-divergence," *Pattern Recognition Letters* **29** (9), 1433–1440 (2008).
16. C. Févotte, N. Bertin and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis," *Neural Computation* **21** (3), 793–830 (2009).
17. K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine* **2** (11), 559–572 (1901).
18. G. H. Golub and C. F. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, 1996.
19. L. Li, G. Lebanon and H. Park, "Fast Bregman divergence NMF using Taylor expansion and coordinate descent," *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, August, 2012.
20. D. Bolder and D. Streliski, "Yield curve modelling at the Bank of Canada," *Technical Reports* **84**, Bank of Canada, 1999.
21. M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," Computational Statistics and Data Analysis **52**, 155–173 (2007).