Portfolio Diversification using Subspace Factorizations

Ruairí de Fréin & Konstantinos Drakakis & Scott Rickard
Complex & Adaptive Systems Laboratory
University College Dublin
Ireland
Email: rdefrein@ee.ucd.ie
Konstantinos.drakakis@ucd.ie scott.rickard@ucd.ie

***Abstract*— Successful investment management relies on allocating assets so as to *beat the stock market.* Asset classes are affected by different market dynamics or latent trends. These interactions are crucial to the successful allocation of monies. The seminal work on portfolio management by Markowitz prompts the adroit investment manager to consider the correlation between the assets in his portfolio and to vary his selection so as to optimize his risk-return profile. The factor model, a popular model for the return generating process has been used for portfolio construction and assumes that there is a low rank representation of the stocks. In this work we contribute a new approach to portfolio diversification by comparing a recently developed clustering technique, SemiNMF, with a new sparse low-rank approximate factorization technique, Sparse-semiNMF, for clustering stocks into latent trend based groupings as opposed to the traditional sector based groupings. We evaluate these techniques using a diffusion model based on the Black-Scholes options pricing model. We conclude that Sparse-semiNMF outperforms semiNMF when applied to synthetic stocks as the contribution of each trend to each stock is more disjoint for Sparse-semiNMF than for semiNMF, in an inter-class sense, meaning that the underlying trends for each stock are more readily apparent, whilst preserving the accuracy of the factorization. We conclude that the trend-based asset classes generated by Sparse-semiNMF should be considered in the investment management process to reduce the risk in portfolio selection.**

Keywords: Finance, Clustering, Low Rank Approximations, Portfolio Diversification.

I. Introduction

Let us set the scene for our work by sketching a brief outline of Modern Portfolio Theory (MPT). The foundation for MPT was laid down by Markowitz in [11] and [12]. His work on the effects of risk on efficient portfolio selection are known more formally as the *Markowitz Efficient Portfolio* and the *Markowitz Efficient Frontier*. In layman's terms, a portfolio on the *Markowitz Efficient Frontier* gives the optimum expected return for a given risk and the *Markowitz Efficient Portfolio* is the portfolio that has been diversified, so that there is no scope for further reduction of risk. These theories laid the ground for the *Capital Asset Pricing Model* (CAPM) proposed by Sharpe in [18]. Some of the basic tools for achieving the risk-reward balance are outlined in [11], [12] and [18], where volatility represents risk and is a function of the correlation of the assets in the portfolio, and the return is a function of the asset returns in the portfolio. Risk can be minimized by selecting a portfolio that contains assets that are anti-correlated, but reward is heavily dependent on risk. Diversification is the process of reducing the risk for a given portfolio return by spreading your bets. For example, if a portfolio contains a few assets and they are strongly dependent on the same underlying trends, the portfolio has a high volatility, and the return is uncertain. If the portfolio asset returns are anti-correlated, certain assets may lose value and others may rise in value but this will average out resulting in possibly a smaller reward but a reduced risk. The Black-Scholes options pricing model proposed in [2] combines a time dimension with volatility to calculate the fair market value for an option.

The goal of this work is to learn an alternative clustering of assets in the stock market by clustering their returns with a sparsity constraint on the assignment matrix yielding more descriptive latent trends or centroids, and consequently, families of stocks with approximate disjoint support. We argue that diversification of investment based on subspace factorizations with sparsity constraints could lead to improved reduction of volatility of a given portfolio.

Let us start by introducing recent work. The seminal application of Independent Component Analysis (ICA) to financial time series was by [1]. The goal was to find the latent factors of instantaneous stock returns, specifically for the daily closing prices of the Tokyo stock exchange. The Independent Components (I.C.) were weighted with respect to the first stock return and were sorted using the L_∞ norm. The central assumption made in this work was that the returns reflect the reaction of the stock market to "a few statistically independent time series", e.g., we have a low rank representation of the stocks. [3] discuss the factor model and its importance in many financial theories, such as, Modern Portfolio Theory and Arbitrage Pricing Theory. These theories assume that securities are represented as linear combinations of some factors. The authors apply ICA to discover the hidden factors and their corresponding sensitivities. Their work is a continuation of that in [1]. Prior to the application of ICA to financial data, Principle Component Analysis [15] was widely used to reveal the driving mechanism in returns. [1] reported that ICA revealed more readily interpretable underlying structure to the data than PCA. This movement from an orthogonality a constraint (PCA) to an independence constraint (ICA) yielded independent trends, and sensitivities which were posited to help minimize the risk for an investment model by helping to increase the diversity, by identifying the underlying independent factors in the market.

ICA is applied to real returns in [3] by transforming the securities to returns and then learning the I.C.'s and fitting a number of the I.C.'s (Low rank) and weights or sensitivities to the "Independent Factor Model". Further work in the same spirit is contributed in [4], where the Minimum Description Length Principle is used to determine the number of factors k to use in the factor model. They explore a number of measures to determine the properties of the learnt factors so that they

can select k factors. These measures are: L_2, L_∞, kurtosis and a measure based on the Wald-Wolfowitz Test (a test that scores the randomness of a sequence at $100(1-\alpha)\%$ confidence level). These measures are used to sort the I.C.'s in terms of energetic significance, maximum value of the factors, non gaussianity, and randomness of the sequence. They conclude that it is more appropriate to assume that the underlying trends are independent rather than uncorrelated. Related work on financial time series was presented in [20] where a model for chain stores was developed. Cashflow is the observable mixture of products (sources) in a store and they have the cashflow for the same number of stores as they have products. They use association rules mining to find related variables that can be considered to be from the same class and reduce the dimensionality of their ICA model. They try to estimate the distributions of each product at a given time period.

In this paper we combine techniques and ideas from disparate communities e.g., source separation and finance to tackle what is essentially a Blind Source Separation (BSS) task. We note that previous work has applied algorithms blindly to real financial data, making the assumption that the generative factor model was composed of uncorrelated or independent latent trends. We propose using a diffusion model based on the Black-Scholes PDE as a test-bed for algorithmic development. We illustrate the performance of Sparse-semiNMF by comparing it with semiNMF using our Black-Scholes synthetic data.

In Section II we discuss how we generate the market data using the closed form solution of the Black-Scholes PDE. We give an illustration of portfolio diversification in Section III. We discuss semiNMF in Section IV and its suitability for the market generative model we consider. We introduce Sparse-semiNMF in section V. Finally, we illustrate the different clustering methods in Section VI and make our conclusions in Section VII.

We shall use the convention that $|.|$ denotes the absolute value function, $\sum_{i=1}^{N}|x_i|$ the L_1 norm and $||.||_2$ the L_2 norm in the following sections.

II. Terse description of the Black-Scholes PDE and diffusion model

Let us generate families of stocks governed by latent trends. We consider a family of n stocks whose prices are governed by the Black-Scholes PDE [19]. We assume that this family collectively depends on m independent realizations of a normal random walk (the interesting case being when $m < n$). Denoting the price of the ith stock by $S_i(t)$, $i = 1, \ldots, n$, we have, according to the Black-Scholes PDE:

$$dS_i(t) = S_i(t)\left(r_i dt + \sum_{j=1}^{m}\sigma_{ij}dW_j(t)\right). \tag{1}$$

$S_i(t)$ follows a linear combination of independent Brownian motions, $W_j(t)$, with constant drift r_i and volatility σ_{ij}. This can be solved in closed form to yield:

$$S_i(t) = S_i(0)\exp\left(\left(r_i - \frac{1}{2}\sum_{j=1}^{m}\sigma_{ij}^2\right)t + \sum_{j=1}^{m}\sigma_{ij}W_j(t)\right), \tag{2}$$

where $i = 1, \ldots, n$. It follows that

$$\ln(S_i(t)) = \ln(S_i(0)) + \left(r_i - \frac{1}{2}\sum_{j=1}^{m}\sigma_{ij}^2\right)t + \sum_{j=1}^{m}\sigma_{ij}W_j(t). \tag{3}$$

We propose setting $D[\cdot]$ to be the detrend operator, namely the operator that removes any linear trend, leaving the factor model:

$$D[\ln(S_i)](t) = \sum_{j=1}^{m}\sigma_{ij}W_j(t). \tag{4}$$

Assuming further that we observe prices at discrete times $t = k\Delta t$, $k = 1, \ldots, K = \frac{T}{\Delta t}$, where T is the total observation interval and Δt the discrete time step, we can represent the observations as a large $n \times K$ array:

$$\mathbf{S}_\pm = \mathbf{\Sigma}_+\mathbf{W}_\pm, \tag{5}$$

where we use the notation $\boldsymbol{S}_{ik},\ \Sigma_{ij},\ W_{ik}$, to denote the elements of each array and $\boldsymbol{S}_\pm$ to denote that the elements of $\boldsymbol{S}$, for example, can lie anywhere in $\Re^{n\times K}$ where as $\mathbf{\Sigma}$ is restricted to the positive orthant of $\Re^{n\times m}$.

$$\boldsymbol{S}_{ik} = D[\ln(S_i)](k\Delta t),\ \mathbf{\Sigma}_{ij} = \sigma_{ij},\ \boldsymbol{W}_{ik} = W_i(k\Delta t) \tag{6}$$

$\mathbf{S} \in \Re^{(n\times K)}$, $\mathbf{\Sigma} \in \Re^{n\times m}$, and $\mathbf{W} \in \Re^{m\times K}$. This is strongly reminiscent of the semi non-negative matrix factorization formulation, see Section IV. We want to cluster the positive dependencies of the stocks on the underlying trends, e.g., the random walks.

III. Motivating Portfolio Diversification

We motivate our paper by considering the following thought experiment which we analyze numerically in our experiments in Section VI.

A. A synthetic diversification problem

Consider a simplified stock market comprising of the stocks in equation (7). There are $F = 2$ families of stocks $\boldsymbol{S}_{ik}^f$ with $n = 3$ stocks per family. Each family is indexed by $f = 1, \ldots, F$ and each stock is indexed by $i = 1, \ldots, n$ and we have $K = 200$ returns. Each family's behavior is governed by $m = 2$ random walks, $\boldsymbol{W}_{j,k}^f$, where $j = 1, \ldots, m$. Examining $\mathbf{\Sigma}$, the volatility matrix, we see that stocks from families $f = 1$ and $f = 2$ are disjoint in an inter-familial sense, that is, stocks from family 1 are not constructed using the random walks from family 2. In this work we cluster stocks into underlying trend based groupings as opposed to the traditional sector based groupings, e.g. health care, technology. The minimum description length principle states that any regularity in a given set of data can be used to compress data, and in the case of the investor the quality of the compression, e.g. understanding of the underlying factors in the market, is related to the financial gain [10]. Having identified the underlying trends in a given set of financial series, we can cluster the data based on the weights of those trends in a given mixture. We ask the question: *Is there an underlying trend that says a company that would be traditionally classed in the technology sector, e.g. IBM, actually behaves like a company that would be traditionally classed as health care sector company, e.g. Pfizer?* For example, consider the traditional groupings versus

$$\begin{pmatrix} \boldsymbol{S}^1_{1,1} & \boldsymbol{S}^1_{1,2} & \dots & \boldsymbol{S}^1_{1,200} \\ \boldsymbol{S}^1_{2,1} & \boldsymbol{S}^1_{2,2} & \dots & \boldsymbol{S}^1_{2,200} \\ \boldsymbol{S}^1_{3,1} & \boldsymbol{S}^1_{3,2} & \dots & \boldsymbol{S}^1_{3,200} \\ \boldsymbol{S}^2_{1,1} & \boldsymbol{S}^2_{1,2} & \dots & \boldsymbol{S}^2_{1,200} \\ \boldsymbol{S}^2_{2,1} & \boldsymbol{S}^2_{2,2} & \dots & \boldsymbol{S}^2_{2,200} \\ \boldsymbol{S}^2_{3,1} & \boldsymbol{S}^2_{3,2} & \dots & \boldsymbol{S}^2_{3,200} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}^1_{1,1} & \boldsymbol{\Sigma}^1_{1,2} & 0 & 0 \\ \boldsymbol{\Sigma}^1_{2,1} & \boldsymbol{\Sigma}^1_{2,2} & 0 & 0 \\ \boldsymbol{\Sigma}^1_{3,1} & \boldsymbol{\Sigma}^1_{3,2} & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}^2_{1,3} & \boldsymbol{\Sigma}^2_{1,4} \\ 0 & 0 & \boldsymbol{\Sigma}^2_{2,3} & \boldsymbol{\Sigma}^2_{2,4} \\ 0 & 0 & \boldsymbol{\Sigma}^2_{3,3} & \boldsymbol{\Sigma}^2_{3,4} \end{pmatrix} \begin{pmatrix} \boldsymbol{W}^1_{1,1} & \boldsymbol{W}^1_{1,2} & \dots & \boldsymbol{W}^1_{1,200} \\ \boldsymbol{W}^1_{2,1} & \boldsymbol{W}^1_{2,2} & \dots & \boldsymbol{w}^1_{2,200} \\ \boldsymbol{W}^2_{1,1} & \boldsymbol{W}^2_{1,2} & \dots & \boldsymbol{W}^2_{1,200} \\ \boldsymbol{W}^2_{2,1} & \boldsymbol{W}^2_{2,2} & \dots & \boldsymbol{W}^2_{2,200} \end{pmatrix} \quad (7)$$

TABLE I
TRADITIONAL VERSUS TREND BASED CLUSTERS

Financial series	Traditional sector	Latent trend based
$S^1_{1,1:200}$	health care	Family 1
$S^1_{2,1:200}$	health care	Family 1
$S^1_{3,1:200}$	technology	Family 1
$S^2_{1,1:200}$	health care	Family 2
$S^2_{2,1:200}$	technology	Family 2
$S^2_{3,1:200}$	technology	Family 2

trend based clustering of the stocks in equation (7) and summarized in Table I.

The reason for asking this question is that an investor might think they have diversified their portfolio by investing in different traditional sectors, e.g. health care and technology by purchasing stocks $\boldsymbol{S}^1_1$ and $\boldsymbol{S}^1_3$, but that supposed diversification might actually be ill-founded as $\boldsymbol{S}^1_3$, traditionally classed as a technology stock, might actually behave like a health care stock, meaning that their portfolio might actually be homologous in nature. Our contention is that investing in stocks belonging to different clusters (identified by a sparse low rank decomposition), but not necessarily the different traditional sectors might offer genuine portfolio diversification opportunities.

We explore a clustering type approach, for example, Non negative Matrix Factorization (NMF), as anti-correlation methods such as PCA [15] fail to acknowledge the multi-trend nature of stock prices. PCA algorithms only use second order statistics and give projections of the data in the direction of maximum variance in the remaining orthogonal subspaces. Principle components are less meaningful than ICA components which enforce a stronger condition, that is, statistical independence. We investigate NMF and its variant semiNMF as the factors are more intuitive than standard ICA. It is tempting to apply vanilla NMF [6] blindly to stock data as it is non-negative, but knowledge of the Black-Scholes generative model, in equation (4) causes us to consider semiNMF as (4) directs us to cluster the stocks, which are *positively* correlated to the underlying random walks. In this work we contribute a new approach, Sparse-semiNMF, which exploits the sparsity of the underlying assignment matrix $\boldsymbol{\Sigma}$ outlined in the synthetic market in the previous section, and learns intuitive factors. The Sparse-semiNMF type approach, leads to more disjoint inter-familial clustering where the factors mirror the block diagonal structure of $\boldsymbol{\Sigma}$ in equation (7).

IV. CLUSTERING APPROACHES

The advent of non-negative matrix factorization as a clustering technique [6] and especially the introduction of the concept of a *Separable Factorial Articulation Family* for a unique decomposition in [7] and its resonance in light of our formulation for the stock generation process using the Black-Scholes model lead us to consider a NMF variant for our performance based clustering.

A. Variants on the NMF theme: Semi-NMF

Traditionally, NMF [6] considers the following problem: Given $\boldsymbol{Y} = [\boldsymbol{y}(1), \boldsymbol{y}(2), \dots, \boldsymbol{y}(T)] \in \Re^{m \times T}$, the data matrix, NMF decomposes the $\boldsymbol{Y}$ into the product of two matrices: a basis, signatures or mixing matrix $\boldsymbol{D} \in \Re^{m \times r}$ and the source component or activation matrix $\boldsymbol{C} \in \Re^{T \times r}$, where all matrices have non-negative elements.

These decompositions are approximative in nature, i.e.,

$$\boldsymbol{Y}_+ \approx \boldsymbol{D}_+ \boldsymbol{C}_+^T. \quad (8)$$

NMF has been used in a wide range of clustering applications, such as, document clustering [17] and gene clustering [14]. Indeed, in [8] we presented our initial work on portfolio selection. We decomposed the daily closing prices of the 30 stocks which make up the Dow Jones Industrial Average, into underlying trends and governing weights, and showed that NMF revealed consistent groupings, which differed from the traditional latent trend based groupings.

One of several extensions of the NMF technique by [5] allows NMF to be applied in a k-means type framework with one of the factors constrained to be non-negative and the data and the other factor unconstrained, e.g. can have mixed signs. They illustrate the connection between semiNMF and k-means. Given the generative model

$$\boldsymbol{Y}_\pm \approx \boldsymbol{D}_\pm \boldsymbol{C}_+^T, \quad (9)$$

they minimize the objective $||\boldsymbol{Y}_\pm - \boldsymbol{D}_\pm \boldsymbol{C}_+^T||_2^2$ where the columns of $\boldsymbol{D}_\pm$ contain the cluster centroids and the elements $\boldsymbol{C}_{pq_\pm}$ are soft assignments compared to the hard assignment and centroids given by k-means. This model is the transpose of equation (5). We will refer to $\boldsymbol{D}$ and $\boldsymbol{C}$ as opposed to $\boldsymbol{D}_\pm$ and $\boldsymbol{C}_+$ as the signs are clear from the context in the remainder of this work. [5] conclude that NMF variants give a better clustering than k-means when clustering accuracy as well as matrix approximation is considered. They also state that their factors are more interpretable than k-means. Given a mixed k-means-canonical pseudo inverse initialization, they learn a matrix factorization using an element-wise multiplicative rule for $\boldsymbol{C}$ and a closed-form rule for $\boldsymbol{D}$:

$$\boldsymbol{D} = \boldsymbol{Y}\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{C})^{-1}, \quad (10)$$

$$\boldsymbol{C}_{pq} \longleftarrow \boldsymbol{C}_{pq}\sqrt{\frac{[(\boldsymbol{Y}^T\boldsymbol{D})^+]_{pq} + [\boldsymbol{C}(\boldsymbol{D}^T\boldsymbol{D})^-]_{pq}}{[(\boldsymbol{Y}^T\boldsymbol{D})^-]_{pq} + [\boldsymbol{C}(\boldsymbol{D}^T\boldsymbol{D})^+]_{pq}}}, \quad (11)$$

where

$$[(\boldsymbol{B})^{+}]_{pq} = (|\boldsymbol{B}_{pq}| + \boldsymbol{B}_{pq})/2, \quad (12)$$

$$[(\boldsymbol{B})^{-}]_{pq} = (|\boldsymbol{B}_{pq}| - \boldsymbol{B}_{pq})/2, \quad (13)$$

V. Motivation for a sparse extension

Considering the disjointness of the matrix formed by placing each $\boldsymbol{\Sigma}^f$ block along the diagonal used to generate the data in equation 7, and where by disjointness we mean sparsity and independence of occurrence, we propose to exploit this sparsity and learn a more interpretable factorization for the data. We note that in this representative problem the sparsity is not readily apparent due to the dimensions of the exemplar. Considering the Dow Jones Index, it would be more reasonable to search for up to 10 families of stocks, where a reasonable number of stocks per family could be as large as 10, which would yield tall $\boldsymbol{\Sigma}^f$ matrices and consequently a very sparse assignment matrix $\mathbf{C}_{DJ}$. For example, according to the low-rank assumption made in [12] and [9], a possible assignment matrix for the Dow Jones Index, $\mathbf{C}_{DJ}$, could have the following form:

$$\begin{pmatrix} [\boldsymbol{\Sigma}^1]_{5\times 3} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\boldsymbol{\Sigma}^2]_{6\times 3} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & [\boldsymbol{\Sigma}^3]_{4\times 2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Sigma}^4]_{5\times 3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Sigma}^5]_{5\times 2} \end{pmatrix} \quad (14)$$

where there are $F = 5$ families of stocks with 2 or 3 underlying trends per family and 4, 5 or 6 stocks per family. This underlines the need for an additive sparsity constraint on the latent assignment matrix in the objective as in (15).

A vector $\boldsymbol{x}$ is considered to be sparse if most of its elements are relatively small [13]. We consider the L_1 norm of a vector $\boldsymbol{x}$, $\sum_{n=1}^{N} |\boldsymbol{x}_n|$, as our sparsity measure in this paper as the assignment matrix $\boldsymbol{C}$ we learn is guaranteed to remain non-negative due to the multiplicative form of its update if it is initialized in the positive orthant. This makes taking the partial derivatives of the additive sparsity term in (15) trivial.

SemiNMF gives a soft assignment and this results in a more interpretable factorization than the hard assignment learnt by k-means. We propose to learn a sparse assignment matrix where variation of the sparsity will tune the "softness" of the assignment to a cluster centroid while maintaining the interpretability of an NMF factorization and increasing the disjointness between the columns of $\boldsymbol{C}^T$ as the sparsity increases.

A. Sparse semiNMF

We present an extension to semiNMF in this section where non-negative quadratic programming (NQP) is used to minimize the regularized objective in (15) with respect to $\boldsymbol{C}$. Our technique alternates between a closed form solution for $\boldsymbol{C}$ and an iterative update for $\boldsymbol{D}$, where $\boldsymbol{D}$ is normalized by the L_2 norm after the iterative $\boldsymbol{D}$ update has converged.

Consider the objective:

$$\boldsymbol{Q} = ||\boldsymbol{Y}_{\pm} - \boldsymbol{D}_{\pm}\boldsymbol{C}_{+}^{T}||_2^2 + \lambda \sum_{n=1}^{rT} |\text{vec}(\boldsymbol{C}^T)| \quad (15)$$

Taking the partial derivatives of (15) with respect to $\boldsymbol{D}$ we calculate the closed form expression for $\boldsymbol{D}$, which is equivalent to equation (10). We now hold $\boldsymbol{D}$ fixed and take the partial derivatives with respect to $\boldsymbol{C}$. We use the Non-negative quadratic programming rule presented in [16] to iteratively learn $\boldsymbol{C}$.

NQP is a general framework where, if the objective can manipulated into the standard NQP form, e.g. (16), the multiplicative rule, (18), is guaranteed to minimize the objective at each iteration. Given that the assignment matrix $\boldsymbol{C}$ is non-negative and the feasible region for the solution of NQP is the positive orthant we manipulate the cost (15), into the standard quadratic form (16),

$$\frac{1}{2}\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \boldsymbol{b}^T\boldsymbol{v}, \quad (16)$$

where vec(.) is a function that vectorizes a matrix column-wise, $\bigotimes$ denotes the Kronecker product, $\mathbf{I}_m$ and $\mathbf{1}_{rT}$ are the identity matrix of dimension $m \times m$ and a vector of ones of dimension $rT \times 1$,

$$\begin{aligned} \boldsymbol{Q} = \frac{1}{2}&\text{vec}(\boldsymbol{C}^T)^T(\mathbf{I}_m \bigotimes 2\boldsymbol{D}^T\boldsymbol{D})\text{vec}(\boldsymbol{C}^T) \\ &+ \text{vec}(-2\boldsymbol{D}^T\boldsymbol{Y})^T\text{vec}(\boldsymbol{C}^T) \\ &+ \lambda\mathbf{1}_{rT}^T\text{vec}(\boldsymbol{C}^T), \end{aligned} \quad (17)$$

and we use the element-wise multiplicative update rule from [16] :

$$\boldsymbol{v}_i \longleftarrow \frac{-\boldsymbol{b}_i + \sqrt{\boldsymbol{b}_i^2 + 4\boldsymbol{a}_i\boldsymbol{c}_i}}{2\boldsymbol{a}_i}\boldsymbol{v}_i \quad (18)$$

where

$$\boldsymbol{a}_i = [(\mathbf{I}_m \bigotimes 2\boldsymbol{D}^T\boldsymbol{D})^{+_{nqp}}\text{vec}(\boldsymbol{C}^T)]_i, \quad (19)$$

$$\boldsymbol{c}_i = [(\mathbf{I}_m \bigotimes 2\boldsymbol{D}^T\boldsymbol{D})^{-_{nqp}}\text{vec}(\boldsymbol{C}^T)]_i, \quad (20)$$

$$\boldsymbol{b}_i = [(\text{vec}(-2\boldsymbol{D}^T\boldsymbol{Y}) + \lambda\mathbf{I})^T]_i, \quad (21)$$

where,

$$\boldsymbol{A}_{ij}^{+_{nqp}} = \begin{bmatrix} \boldsymbol{A}_{ij} & \text{if } \boldsymbol{A}_{ij} > 0 \\ 0 & \text{otherwise} \end{bmatrix}, \quad (22)$$

$$\boldsymbol{A}_{ij}^{-_{nqp}} = \begin{bmatrix} |\boldsymbol{A}_{ij}| & \text{if } A_{ij} < 0 \\ 0 & \text{otherwise} \end{bmatrix}, \quad (23)$$

and we assume the matrix $\boldsymbol{A}$ is symmetric and semi-positive definite so that the objective (15) is bounded below and its optimization is convex.

VI. Experiments

We illustrate the differences between semiNMF and Sparse-semiNMF using synthetic data generated using the Black-Scholes diffusion model. Consider the synthetic detrended stock market, consisting of 365 returns illustrated in row 1 of Figure 1. This market is constructed of 3 families of stocks with 4 stocks per family, giving a total of 12 stocks, with 2 underlying trends per family, e.g. 6 latent trends in total. To illustrate the performance of Sparse-semiNMF we ask the question: *Given that the number of underlying trends is known, can we cluster the stocks, dependent on the same underlying trends, in the same groupings?* . We initialize the assignment matrix, $\boldsymbol{C}$, of both semiNMF and Sparse-semiNMF with the same initial factors, using k-means clustering, as both techniques are dependent on their initializations. $\boldsymbol{D}$, the trends are initialized using the update rule (10) and we normalize

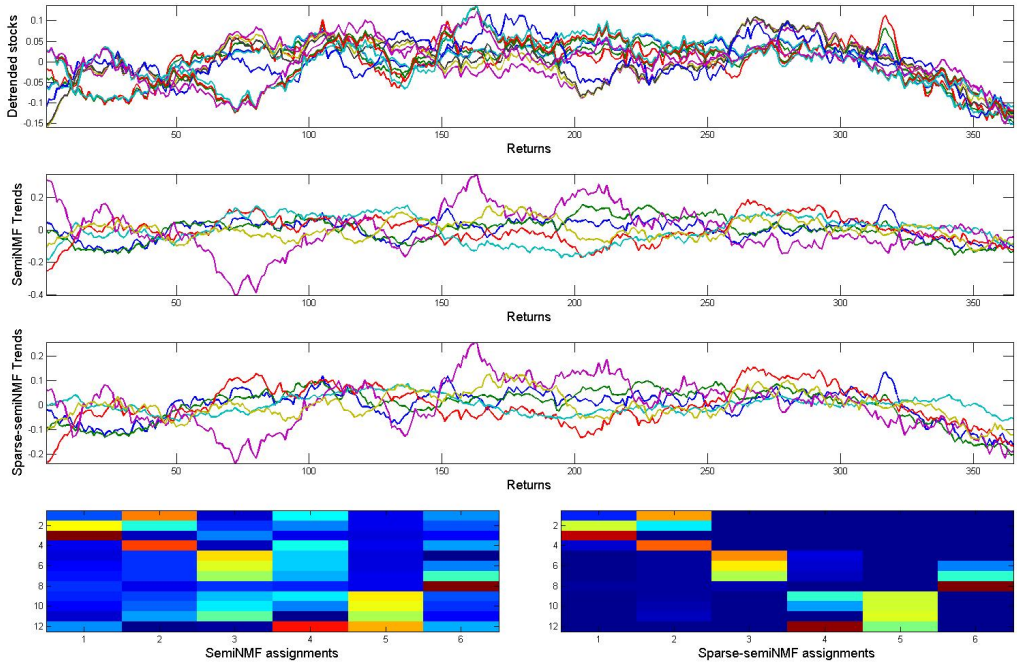

Fig. 1. The upper plot shows 12 de-trended stocks generated by the Black-Scholes diffusion model using uniformly distributed random numbers for the volatility matrix and drift. There are 365 returns. 6 underlying trends were used to generate these stocks. There are 3 families of stocks each relying on 2 trends each. Rows 1 and 2 show the latent trends found by semiNMF (row 1) and Sparse-semiNMF (row 3). The assignment matrices on row 4 for semiNMF (column 1) and Sparse-semiNMF (column 2) show the advantage of a sparsity constraint. Sparse-semiNMF yields more disjoint columns giving a better representation of the inter-familial disjointness. SemiNMF produces a noisier assignment of stocks to families. SemiNMF yields an SNR of 216.6032 dB where as sparse-semiNMF yields an SNR of 86.5903 dB.

each column of $\boldsymbol{Y}$ using the L_2 norm. Each algorithm iterates for $10,000$ iterations to ensure convergence and we illustrate the resulting factorization into the 6 latent trends, in rows 2 and 3, and assignment matrices, in row 4, columns 1 and 2 of Figure 1. The desideratum, after the successful application of semiNMF or Sparse-semiNMF, is a permuted and scaled version of the initial assignment or volatility matrix, a block diagonal matrix, which was used to generate the data. Once we reverse the permutation introduced in both algorithms, and have identified the latent trends that should be grouped together, we can then evaluate the dependence of each stock on the latent trends by identifying which columns of $\boldsymbol{C}^T$ have the most energy and assign stocks to families in this fashion.

We undo the permutation ambiguity by forming a confusion matrix $\boldsymbol{W} = \boldsymbol{H}\boldsymbol{C}^T$, where $\boldsymbol{H}$ consists of ideal columns of the assignment matrix $\boldsymbol{C}^T$ along its rows. For example, one of the rows of $\boldsymbol{H}$ for Figure 1 would consist of $[0,0,0,0,1,1,1,1,0,0,0,0]$, which yields a large inner product with columns 4 and 5 of the assignment matrix, $\boldsymbol{C}^T$, for the Sparse-semiNMF assignment matrix in Figure 1. For example, we calculate the percentage of energy in row $p = 9$ of $\boldsymbol{C}^T$ in the correct positions, S_p, e.g. the percentage of the energy in the columns corresponding to the trends for family f defined by the indices $I_f = [4,5]$, and give a score for the accuracy of the clustering based on this percentage.

$$S_p = \frac{\sum_{q \in I_f} \boldsymbol{C}_{pq}^2}{\sum_{q=1}^{r} \boldsymbol{C}_{pq}^2}, \tag{24}$$

where I_f denotes the set of indices for the pair of columns, as there are two underlying trends per family, of $\boldsymbol{C}^T$ that correspond to the trends of family f. Given that we have identified the pairings of columns of the dominant weights for the latent trends for each family and calculated the percentage score for each row of the assignment matrix, we quantify the measure of accuracy of the clustering by averaging the percentages for each row over the whole assignment matrix.

A. The advantage of Sparsity in this setting

It is immediately evident from the assignment matrices on row 4 of Figure 1 that Sparse-semiNMF learns harder assignments due to the additive sparsity constraint in the objective (15). The increase in the sparsity comes at a cost of reducing the SNR, e.g. from an SNR of 216.6032dB for semiNMF to an SNR of 86.5903dB for Sparse-semiNMF due to the trade-off in the objective. The extracted trends in row 3 of Figure 1 are more expressive allowing the assignments to represent the data more efficiently. Consequently, the columns of the assignment matrix for Sparse-semiNMF are more disjoint in an inter-familial sense, meaning that the structure of the block-diagonal matrix used to generate the families is readily visible.

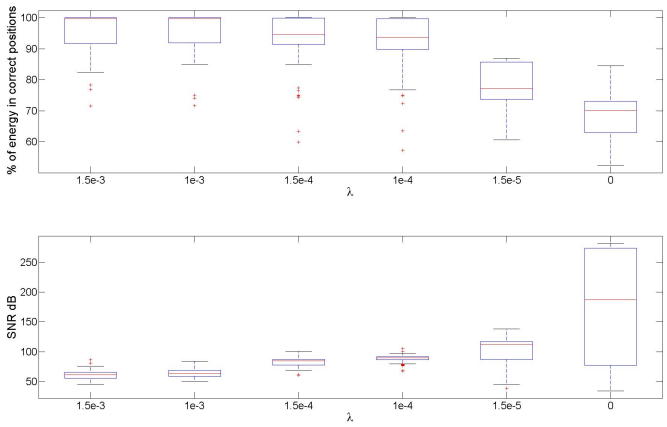

Fig. 2. Box plots are used to illustrate the increase in accuracy of the clustering as sparsity, λ, is increased. Each plot represents the median and interquartile range of the results. The trade off between sparsity and SNR in clearly evident, yet a decomposition with an SNR of 50dB is a worthy reconstruction.

VII. Results

We run 100 Monte Carlo experiments and illustrate the statistical break-down of the results in Figure 2, where we iterate each technique for $10,000$ iterations. For each experiment we initialize our factors as in the previous section. We use the same initial conditions for each Sparse-semiNMF tuned by the different sparsity parameters λ for a single Monte Carlo run and also for semiNMF. We have 3 families with 2 trends per family and 4 stocks per family. The trade-off between SNR and sparsity, tuned by the parameter λ, in the objective (15) is clearly identifiable. An SNR of 50dB is more than sufficient for an accurate representation of the data. Figure 2 tells us that, as we increase λ the degree of hardness of the clustering increases making the following assignment of stocks to families step easier as the dependence on latent trends, specifically the trends for a certain family, is more clearly defined. semiNMF learns a softer clustering which gives a more accurate representation of the data, but for clustering purposes, Sparse-semiNMF outperforms semiNMF as there is a clear delineation between the support of the assignment matrix for the underlying trends. The disjoint support of the weights in the assignment matrices resulting from the application of Sparse-semiNMF with a $\lambda > 1 \times 10^{-4}$ yields far a superior measure of the percentage of the energy in the correct columns compared to standard semiNMF.

VIII. Conclusions

In conclusion, we have presented a new technique, Sparse-semiNMF for clustering stocks based on latent trends, so that this information can be leveraged to minimize the risk when selecting a portfolio of holdings. We have used a detrended diffusion model derived from the Black-Scholes option pricing model as a benchmark for comparing our technique with the state-of-the-art. Sparse-semiNMF combines the advantages of an intuitive factorization, with sparse assignments, to decompose synthetic stock data into a meaningful assignment matrix and latent trends matrix, where the hardness of the clustering assignment can be tuned to reveal assignments with disjoint support. We note that harder assignments yield factors which were more in line with the factors used to generate the data as they mirror the underlying inter-familial disjointness of the stock market families. Sparse-semiNMF lends itself to the detrended Black-Scholes diffusion factor model as the model dictates that we should only consider positive correlations in the assignment matrix and mixed-signs trends. We conclude that sparse-semiNMF identifies valuable clusters, worthy of consideration in the diversification of a portfolio.

Acknowledgment

Supported by Science Foundation Ireland and the Irish Research Council for Science Engineering and Technology. We thank Paul O'Grady for advice.

References

[1] A.D. Back and A.S. Weigend. A first application of independent component analysis to extracting structure from stock returns. *Int. Journal of Neural Systems*, 8(4):473–484, August 1997.

[2] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, May - June 1973.

[3] Siu-Ming Cha and Lai-Wan Chan. Applying independent component analysis to factor model in finance. In *IDEAL '00: Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents*, pages 538–544, London, UK, 2000. Springer-Verlag.

[4] L. Chan and S. Cha. Selection of independent factor model in finance. *In The Proceedings of the Third International Conference on Independent Component Analysis and Signal Saparation*, 2001.

[5] C. Ding, T. Li, and M.I. Jordan. Convex and seminonnegative matrix factorizations. Technical Report 60428, Lawrence Berkeley National Laboratory, 2006.

[6] Daniel D.Lee and H.Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[7] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts, 2003.

[8] K. Drakakis, S. Rickard, R. de Frein, and A. Cichocki. Analysis of financial data using non-negative matrix factorization. *International Journal of Mathematical Sciences*, 6(2), June 2007.

[9] A. Hyvarinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, June 2000.

[10] Jr. Kelly, J. A new interpretation of information rate. *Information Theory, IEEE Transactions on*, 2(3):185–189, Sep 1956.

[11] Harry M. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[12] Harry M. Markowitz. *Portfolio Selection and Efficient Diversification of Investments*. John Wiley and Sons, Inc., New York, 1959.

[13] Paul D. O'Grady, Barak A. Pearlmutter, and Scott T. Rickard. Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*, 15(1):18–33, 2005. Blind Source Separation and De-convolution in Imaging and Image Processing.

[14] Alberto D. Pascual-Montano, Francisco Tirado, Pedro Carmona-Saez, Jos Mara Carazo, and Roberto D. Pascual-Marqui. Two-way clustering of gene expression profiles by sparse matrix factorization. In *CSB Workshops*, pages 103–104. IEEE Computer Society, 2005.

[15] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 1901.

[16] F. Sha, L. Saul, and D. Lee. Multiplicative updates for non-negative quadratic programming in support vector machines, 2002.

[17] Farial Shahnaz, Michael W. Berry, V. Paul Pauca, and Robert J. Plemmons. Document clustering using nonnegative matrix factorization. *Inf. Process. Manage.*, 42(2):373–386, 2006.

[18] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, September 1964.

[19] K. Sircar and G. Papanicolaou. General black-scholes models accounting for increased market volatility from hedging strategies, 1996.

[20] Shangming Yang and Yi Zhang. Fast ica for online cashflow analysis. In *ISNN (2)*, pages 891–896, 2005.