Jurnal Teknologi

Full Paper

A THEORETICAL MODEL FOR PORTFOLIO OPTIMIZATION DURING CRISIS

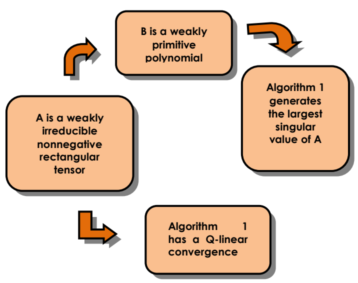
Colin Cheong Kiat Gan[a*], Sudarso Kaderi Wiryono[a], Deddy Priatmodjo Koesrindartoto[a], Budhi Arta Surya[b]

[a]School of Business and Management, Bandung Institute of Technology, Jalan Ganesha 10, Bandung 40132, Indonesia
[b]School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

Article history
Received
12 August 2015
Received in revised form
14 December 2015
Accepted
11 January 2016

*Corresponding author
fadhilah@umt.edu.my

Graphical abstract



Abstract

The premise of this paper is providing a theoretical model for a novel way to portfolio optimization using generalized hyperbolic distribution during crisis with risk measures, expected shortfall and standard deviation. Getting good expected returns from investing in portfolio assets like stocks, bonds and currencies during crisis period chosen is harder where the risks cannot be diverted because of disruptive financial jolts i.e. sudden and unprecedented events like subprime mortgage crises in 2008. Multivariate generalized hyperbolic distribution on joint distribution of risk factors from stocks, bonds and currencies is used because it can simplify the risk factors calculation by allowing them to be linearized. The results show the premise is true. The contributions are discovering both the appropriate probability distribution and risk measure will determine whether the portfolio is optimal or not. The practical application will be taking care of the risk to take care of the profit.

Keywords: Generalized hyperbolic distribution, portfolio optimization

1.0 INTRODUCTION

During crisis, financial data do not follow the Gaussian distribution because of the large and frequent jumps. Mandelbrot [1] shares that logarithmic relative price changes on financial and commodity markets display heavy-tailed distributions. Barndorff-Nielsen [2] introduced the original generalized hyperbolic (GH) distribution where variance gamma is a special case. Madan and Seneta [3] provide a lévy process with variance gamma distribution of increments to model logarithmic price processes. Barndorff-Nielsen [4] shows subclasses of GH distributions are excellent fits to empirically observe financial log price processes increments especially logarithmic return distributions.

Eberlein and Keller [5] discuss about hyperbolic distribution and Barndorff-Nielsen [4] determines the canonical Levy kind decomposition of the process. Some examples of statistical properties of asset returns are distributional properties, tail properties and extreme fluctuations, and pathwise regularity. Cont [6] talks about empirical return distributions often display excess kurtosis and heavy tail. Generalized hyperbolic skew student's t distribution has the important property of two tails exhibiting different behavior. One is polynomial, while the other is exponential. Aas and Haff [7] discuss this is a perfect fit to skew financial data exhibiting such tail behaviors. Student's t and Gaussian distributions are limit distributions of GH. Surya and Kurniawan [8] mention that GH family distributions are popularly used because they fit well to financial return data and extend to common student's t and normal distributions. Robust and fast estimation procedures are rare in a limited data environment. Its alternative class is with random vectors that are stochastically independent and generalized hyperbolic marginals affine-linearly changed. They have good approximation attributes and have appealing reliance framework. Tail dependence of extreme events can be modelled with them. Schmidt *et al.* [9] talks about the essential approximation and arbitrary number creation methods. Surya and Kurniawan [8] talks about the reasons of using GH distribution as the appropriate distribution for the portfolio loss distribution. First, it is the linearity property of GH distribution. Second, it encompasses the Generalized Inverse Gaussian and Multivariate Normal Mean Variance Mixture distributions with flexibility to model a wide range of portfolio loss distribution. Jarque and Bera [10] obtain the normality of observations and regression disturbances tests through Lagrange multiplier procedure or score test on the Pearson family of distributions. They have optimum asymptotic power properties and good finite sample performance. Gnanadesikan and Kettenring [11] address the commutativity of robust estimators of multivariate location by applying estimators after a preliminary transformation to principal components coordinates or to sphericized coordinates robustified with a data-dependent transformation when the sample covariance or correlation matrix is used for obtaining the transformation. The objectives of techniques for detecting multivariate outliers are intertwined with those methods of assessing the joint normality of multiresponse data. Gopikrishnan *et al.* [12] discovered an asymptotic power-law behavior for the cumulative distribution with an exponent $a \approx 3$ outside the Levy regime ($0 < a < 2$). A non-linear fractional covariance matrix is generalized by a non-linear transformation with returns as Gaussian variables using covariance matrix to measure dependence between the non-Gaussian returns. It becomes the definite fat tail framework of the fundamental marginal distributions for firmness and good control. The portfolio distribution is a mapping to particle physics ϕ^q field theory using Feynman diagrammatic approach and large divergence theory for multivariate Weinbull distributions. Substantial empirical tests on the foreign exchange market prove the theory. Sornette *et al.* [13] provides an ample prediction of risks of a portfolio hinges much more on the appropriate description of the tail structure not on their interdependence for fat tail distributions.

Alternative risk measures have to be considered because financial return data are non-Gaussian with heavy tails and volatility cannot capture extreme large losses. Volatility only measures financial return deviations from its mean. Value-at-Risk (VaR) determines the point of relative loss level exceeded at a specified degree. It can measure the behavior of negative return distributions at a point far from the expected return when adjusted suitably. However, it has a serious disadvantage. Artzner *et al.* [14] says that it can lead to a centralized portfolio when applied to non-elliptical distributions violating the diversity principle. Portfolio optimization becomes an expensive computational problem. Artzner *et al*. [14] highlights expected shortfall (ES) as a risk measure responds to VaR's disadvantage. It is a coherent risk measure which always results in a diversified portfolio. It shows how the distributions' tails behave like VaR with a magnified scope. These attributes make it more favorable than VaR.

Financial return commensurate with risk taken. During crisis, the financial risk escalates with diminishing returns and ballooning losses. Using the wrong distribution and risk measure can exacerbate the loss during crisis where certain huge losses in the financial markets happen at far higher frequencies. It is even more important to determine the right distribution and risk measure during crisis.

The key difference in the approach to portfolio optimization during crisis is the portfolio loss is minimized given the targeted expected return because risk can be present in the entire loss distribution. On the contrary, in a non-crisis situation, maximizing the targeted expected return given the standard deviation is the key because risk may not be present in the entire loss distribution.

Understanding the appropriate approach to portfolio optimization during crisis and non-crisis situations will allow the targeted expected return to be obtained given the risk measure taken. This has practical applications for both retail and institutional investors.

The paper discusses the general properties of GH distribution, the expected shortfall as coherent risk measure and elaborates the asset structures of the portfolio, discusses the profit and loss (P & L) distribution on multivariate GH distribution, constructs the portfolio optimization problems and provides the conclusion of the paper.

2.0 GENERALIZED HYPERBOLIC DISTRIBUTION

McNeil *et al*. [15] talk about the Generalized Hyperbolic Distribution is built upon the Generalized Inverse Gaussian and Multivariate Normal Mean Variance Mixture Distributions.

2.1 Definition Generalized Inverse Gaussian Distribution (GIG).

The random variable, Z, is a GIG represented by $Z \sim N^{-}(\lambda, \chi, \psi)$ if its probability density function is:

$$f(z) = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}^{\lambda}\right)}{2K_\lambda\left(\sqrt{\chi\psi}\right)} z^{\lambda-1} exp\left(-\frac{1}{2}\left(\chi z^{-1} + \psi z\right)\right), z, \chi, \psi > 0, \quad (1)$$

K_λ is a modified Bessel function of the third kind with index λ fulfilling the parameters:

$\chi > 0, \psi \geq 0 \; if \; \lambda < 0$

$\chi > 0, \psi > 0 \; if \; \lambda = 0$

$\chi \geq 0, \psi > 0 \; if \; \lambda > 0$

2.2 Definition Multivariate Normal Mean Variance Mixture Distribution (MNMVM)

A random variable $X \epsilon R^d$ is MNMVM if it is represented by the following:

$$X = \mu + W\gamma + \sqrt{W}AZ \qquad (2)$$

with $\mu, \gamma \epsilon R^d$ and $A \epsilon R^{d \times k}$, a matrix, as distribution parameters, $Z \sim N_k(0, I_k)$ is a standard multivariate normal random variable. W is non-negative, scalar mixing random variable independent of Z. $\Sigma := AA'$ must be positive definite. In a univariate model, Σ is replaced by σ^2.

Barndorff-Nielsen [2] mentions this as a new class of distribution first proposed for multivariate GH distribution. Barndorff-Nielsen *et al.* [16] mentions this was further developed. Its parameters are: μ, location parameter; γ, skewness parameter; Σ, scale parameter; W, shock factor for skewness and scale.

The GH distribution is defined from Equation (2) linked to the lévy process where the levy-ito decomposition allows the linear transformation of risk factors.

2.3 Definition Generalized Hyperbolic Distribution (GH)

A random variable $X \epsilon R^d$ is GH-distributed if it is represented by

$$X \sim GH(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \qquad (3)$$

if and only if it has Equation (2) with $W \sim N^-(\lambda, \chi, \psi)$ is a scalar GIG distributed random variable. X is symmetric if and only if $\gamma = 0$.

$$f(x) = c\frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{\left(\chi+\frac{(x-\mu)'(x-\mu)}{\Sigma}\right)\left(\psi+\frac{\gamma'\gamma}{\Sigma}\right)}\right)e^{\frac{(x-\mu)'\gamma}{\Sigma}}}{\left(\sqrt{\left(\chi+\frac{(x-\mu)'(x-\mu)}{\Sigma}\right)\left(\psi+\frac{\gamma'\gamma}{\Sigma}\right)}\right)^{\frac{d}{2}-\lambda}} \qquad (4)$$

Barndorff-Nielsen [2] mentions the above pdf is consistent with the definition of GH first proposed. It has the following normalizing constant:

$$c = \frac{\sqrt{\chi\psi}^{-\lambda}\psi^{\lambda}\left(\psi+\frac{\gamma'\gamma}{\Sigma}\right)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})} \qquad (5)$$

Equation (2) contributes significantly to the linearity property of GH distribution. The following theorem is central to solving optimal portfolio selection problems discussed more in Section 6.

2.4 Theorem

If $\boldsymbol{X} \sim GH_d(\lambda, \chi, \psi, \mu, \boldsymbol{\Sigma}, \gamma)$ and $\boldsymbol{Y} = \boldsymbol{BX} + \boldsymbol{b}$ given $\boldsymbol{B} \epsilon R^{k \times d}$ and $\boldsymbol{b} \epsilon R^k$, then

$$\boldsymbol{Y} \sim GH_k(\lambda, \chi, \psi, \boldsymbol{B}\mu + \boldsymbol{b}, \boldsymbol{B\Sigma B}', \boldsymbol{B}\gamma) \qquad (6)$$

McNeil *et al.* [15] provide the theoretical proof for Theorem 2.4 is found in Proposition 3.13 showing the parameters coming from the generalized inverse Gaussian distribution stays the same under linear operations.

3.0 WEAKER CONVERGENT CONDITION

An algorithm for finding the largest singular value of an irreducible nonnegative rectangular tensor was proposed by Chang *et al*. [7]. Later, it was updated by Zhou *et al*. [11]. In this section, we will prove that the algorithm is convergent for weakly irreducible nonnegative rectangular tensors.

Algorithm 1 [11]

Step 0: Choose $\rho > 0, x^{(1)} > 0$ and $y^{(1)} > 0$. Set $k = 1$.

Step 1: Calculate $\xi^{(k)} = B_x\left(x^{(k)}, y^{(k)}\right)$ and $\eta^{(k)} = B_y\left(x^{(k)}, y^{(k)}\right)$. Let

$$\underline{\mu}_k = \min_{x_i^{(k)}>0, y_j^{(k)}>0}\left(\frac{\xi_i^{(k)}}{\left(x_i^{(k)}\right)^{M-1}}, \frac{\eta_j^{(k)}}{\left(y_j^{(k)}\right)^{M-1}}\right),$$

$$\bar{\mu}_k = \max_{x_i^{(k)}>0, y_j^{(k)}>0}\left(\frac{\xi_i^{(k)}}{\left(x_i^{(k)}\right)^{M-1}}, \frac{\eta_j^{(k)}}{\left(y_j^{(k)}\right)^{M-1}}\right).$$

Step 2: If $\underline{\mu}_k = \bar{\mu}_k$, then stop. Otherwise, compute

$$x^{(k+1)} = \frac{\left(\xi^{(k)}\right)^{\left[\frac{1}{M-1}\right]}}{\left\|\left(\xi^{(k)}, \eta^{(k)}\right)^{\left[\frac{1}{M-1}\right]}\right\|} \quad \text{and} \quad y^{(k+1)} = \frac{\left(\eta^{(k)}\right)^{\left[\frac{1}{M-1}\right]}}{\left\|\left(\xi^{(k)}, \eta^{(k)}\right)^{\left[\frac{1}{M-1}\right]}\right\|},$$

replace k with $k+1$ and go to Step 1.

Let $\mu_0 = \underline{\mu}_k = \bar{\mu}_k$. The largest singular value of A is $\mu_0 - \rho$. Zhou *et al*. [11] have shown that this algorithm is convergent, if A is an irreducible nonnegative rectangular tensor. We will now show that Algorithm 1 is convergent if A is a weakly irreducible nonnegative rectangular tensor.

We define the polynomial map $P = (P_1, \ldots, P_N)^T : R_+^N \to R_+^N$ through:

$$P(z) = \begin{pmatrix} Ax^{p-1}y^q \\ Ax^p y^{q-1} \end{pmatrix},$$

where $N = m + n$, $z = \begin{pmatrix} x \\ y \end{pmatrix}$. Let P_i be a polynomial with degree, $d_i > 1$. Suppose that the coefficient of each monomial in P_i is nonnegative. The associated graph of P is the directed graph $G(P) = (V, E(P))$, where the vertices $V = \{1, 2, \ldots, N\}$ and the edge $(i, j) \in E(P)$ if the coefficient of variable z_j appears in the expression of P_i.

Definition 2: Let $P = (P_1, \ldots, P_N)^T : R_+^N \to R_+^N$ be a polynomial map, where each P_i is a homogeneous polynomial of the degree $d \geq 1$ with nonnegative coefficients. We call P weakly irreducible if $G(P)$ is strongly connected. If the directed graph $G(P)$ is strongly connected, and the great common divisor (gcd) of the lengths of its circuits is equal to one, then we say P is weakly primitive.

Another way to check the gcd of a graph's lengths of is to observe the diagonal of its associated matrix. An irreducible matrix has a nonzero main diagonal entry if and only if the associated directed graph has a loop, a closed path with length equals to one.

We can show that P is weakly primitive by proving that the associated matrix of its graph is primitive. Let $M(G(P))$ be the associated matrix of graph $G(P)$. We can say that $M(G(P))$ is primitive if the graph is strongly connected, and if the gcd of its lengths is equal to one.

Definition 3: A rectangular tensor A is weakly irreducible if P is weakly irreducible.

Let $B(z)=\begin{pmatrix} Ax^{p-1}y^q+\rho x^{[M-1]} \\ Ax^p y^{q-1}+\rho y^{[M-1]} \end{pmatrix}$ and let $I(z)=\begin{pmatrix} \rho x^{[M-1]} \\ \rho y^{[M-1]} \end{pmatrix}$. Hence we have $B(z)=P(z)+I(z)$. Now we prove that Algorithm 1 is convergent, if tensor A is weakly irreducible.

We can now present our results for this section.

Lemma 1: If A is a weakly irreducible nonnegative rectangular tensor with the order p,q and the $m\times n$ dimension, then $B(z)$ is a weakly primitive polynomial.

Proof. Since A is weakly irreducible then $P(z)$ is a weakly irreducible polynomial. By Definition 2, the graph of $P(z)$, $G(P(z))$ is strongly connected. By Theorem 1, the matrix of $G(P(z))$ is irreducible. We know that $G(I(z))$, the graph of $I(z)$, has a self-loop at each vertices. Therefore the matrix of $G(I(z))$ is a diagonal matrix. Hence, by Corollary 1, the matrix of $G(B(z))$ is primitive. By Theorem 2, $G(B(z))$ is strongly connected, and has a gcd that is equal to one. This implies by Definition 2 that $B(z)$ is a weakly primitive polynomial.

The following theorem is the main result of this paper.

Theorem 6. Let A be a weakly irreducible rectangular tensor of the p,q-th order and the $m\times n$ dimension. Suppose that (μ_0,x_0,y_0) is the solution of equation (5). Then, Algorithm 1 yields the value of μ_0 through a finite number of steps, or generate two convergent sequences $\{\underline{\mu}_k\}$ and $\{\bar{\mu}_k\}$, both of which converge to μ_0. The largest singular value of A is $\mu_0-\rho$.

Proof. By Lemma 1 and Corollary 5.1 [21], Algorithm 1 converges when the rectangular tensor A is weakly irreducible.

4.0 RATE OF CONVERGENCE

In this section, we will show that Algorithm 1 has Q-linear convergence, when A is a nonnegative weakly irreducible rectangular tensor of p,q-th order and $m\times n$ dimensional. We use the same argument as Zhou, Qi and Wu's study [22].

Define:

$$F(z)=B(z)=\begin{pmatrix} Ax^{p-1}y^q+\rho x^{[M-1]} \\ Ax^p y^{q-1}+\rho y^{[M-1]} \end{pmatrix},$$

$$D(z)=F(z)^{\left[\frac{1}{M-1}\right]}, \qquad H(z)=\frac{D(z)}{\phi(D(z))},$$

where $\phi: R_+^N \to R_+$ is defined as:

$$\phi(z)=z_1=\sum_{i=1}^N z_i,$$

for any nonnegative $z\in R_+^N$. We can see that the sequence $\{z^{(k)}\}$ in Algorithm 1 is generated by

$$z^{(k+1)}=H\left(z^{(k)}\right), \quad k=1,2,\ldots, \tag{1}$$

and $\phi\left(z^{(k)}\right)=1$ for all $k=1,2,\ldots$.

Lemma 2. Let A,μ_0,x_0 and y_0 be as in Theorem 6 and let $H'(z_0)$ be the Jacobian of the function H at z_0. Then, $\rho(H'(z_0))<1.$

Proof. Let μ_0 be the largest singular value of B and z_0 be the corresponding eigenvector. We have $H(z_0)=D(z_0)/\phi(D(z_0))$. We want to show that:

$$\rho(H'(z_0))=\rho\left(\frac{D'(z_0)\phi(D(z_0))-D(z_0)\phi'(D(z_0))}{\phi^2(D(z_0))}\right)<1.$$

We already have $F(z_0)=B(z_0)=\mu_0 z_0^{[M-1]}$ and $\phi(z_0)=1.$

Hence, $D(z_0)=(F(z_0))^{\left[\frac{1}{M-1}\right]}=\mu_0^{\left[\frac{1}{M-1}\right]}z_0.$ Let $\mu_1=\mu_0^{\left[\frac{1}{M-1}\right]}$, so we have $D(z_0)=\mu_1 z_0.$

Now we compute $D'(z_0)$, i.e. the Jacobian of D at z_0. Let

$$D(z_0)=(F(z_0))^{\left[\frac{1}{M-1}\right]}=\begin{bmatrix}(F_1(z_0))^{\left[\frac{1}{M-1}\right]}\\(F_2(z_0))^{\left[\frac{1}{M-1}\right]}\\\vdots\\(F_N(z_0))^{\left[\frac{1}{M-1}\right]}\end{bmatrix},$$

$$\nabla\left((F_1(z_0))^{\left[\frac{1}{M-1}\right]}\right)=\frac{1}{M-1}(F_1(z_0))^{\left[\frac{2-M}{M-1}\right]}\nabla F_1(z_0).$$

By the same method, we can get:

$$\nabla\left((F_i(z_0))^{\left[\frac{1}{M-1}\right]}\right)=\frac{1}{M-1}(F_i(z_0))^{\left[\frac{2-M}{M-1}\right]}\nabla F_i(z_0)\ ,\ i=1,\dots,N.$$

Thus the Jacobian of D at z_0 is given below:

$$D'(z_0)=\nabla\left((F(z_0))^{\left[\frac{1}{M-1}\right]}\right)$$

$$=\begin{bmatrix}\nabla\left((F_1(z_0))^{\left[\frac{1}{M-1}\right]}\right)\\\nabla\left((F_2(z_0))^{\left[\frac{1}{M-1}\right]}\right)\\\vdots\\\nabla\left((F_N(z_0))^{\left[\frac{1}{M-1}\right]}\right)\end{bmatrix}=\begin{bmatrix}\frac{1}{M-1}(F_1(z_0))^{\left[\frac{2-M}{M-1}\right]}\nabla F_1(z_0)\\\frac{1}{M-1}(F_2(z_0))^{\left[\frac{2-M}{M-1}\right]}\nabla F_2(z_0)\\\vdots\\\frac{1}{M-1}(F_N(z_0))^{\left[\frac{2-M}{M-1}\right]}\nabla F_N(z_0)\end{bmatrix}$$

$$=\begin{bmatrix}\frac{1}{M-1}(F_1(z_0))^{\left[\frac{2-M}{M-1}\right]} & & 0\\ & \ddots & \\ 0 & & \frac{1}{M-1}(F_N(z_0))^{\left[\frac{2-M}{M-1}\right]}\end{bmatrix}\begin{bmatrix}\nabla F_1(z_0)\\\nabla F_2(z_0)\\\vdots\\\nabla F_N(z_0)\end{bmatrix}$$

$$=diag\left(\frac{1}{M-1}(F(z_0))^{\left[\frac{2-M}{M-1}\right]}\right)F'(z_0)$$

$$=\frac{1}{M-1}diag\left((\mu_1 z_0)^{[2-M]}\right)F'(z_0),$$

where $\frac{1}{M-1}diag\left((\mu_1 z_0)^{[2-M]}\right)$ is a constant with $\mu_1>0$, and z_0 is a positive vector. Therefore $G(D'(z_0))=G(F'(z_0))$. For the graph of B, by definition, there exists an edge between i and j, if variable z_j appears in the expression of B_i. Notice that the graph of B is similar to the graph of D', $G(D'(z_0))=G(F'(z_0))=G(B(z_0))$. Lemma 1 states that B is weakly primitive, therefore, the graph of B is strongly connected. Hence the graph of D' is also strongly connected, and D' is therefore irreducible. The term $I(z)$ in B ensures that the diagonal is nonzero, and that implies D' be primitive matrix. Since $D'(z_0)$ is a primitive matrix, by Theorem 3, the eigenvalues $v_1,v_2,\dots,v_N$ of $D'(z_0)$ can be ordered as follows:

$$v_1=\rho(D'(z_0))>|v_2|\ge|v_3|\ge\dots\ge|v_N|.$$

For all $t>1$, we expand $D(tz_0)$ about z_0 by using Taylor's Series, and obtains:

$$\begin{aligned}t\mu_1 z_0&=D(tz_0)\\&=D(z_0)+D'(z_0)(tz_0-z_0)+o(\|tz_0-z_0\|)\\&=\mu_1 z_0+(t-1)D'(z_0)z_0+o(t-1)\\(t-1)\mu_1 z_0&=(t-1)D'(z_0)z_0+o(t-1),\end{aligned}$$

which implies that $D'(z_0)z_0=\mu_1 z_0$. Since $D'(z_0)$ is a primitive matrix, and $z_0>0$, by referring to the Theorem 3, z_0 is an eigenvector of $D'(z_0)$ associated with the largest eigenvalue $\mu_1=v_1$. Therefore, $\phi(D(z_0))=\phi(\mu_1 z_0)=\mu_1$.

We also have $\phi(D(z_0))=D_1(z_0)+D_2(z_0)+\dots+D_N(z_0)$, and $\phi'(D(z_0))=D_1'(z_0)+D_2'(z_0)+\dots+D_N'(z_0)=eD'(z_0)$, where e is the row vector of ones with N dimension. From $H(z_0)=D(z_0)/\phi(D(z_0))$, and after some manipulations we attain:

$$\begin{aligned}H'(z_0)&=\frac{D'(z_0)\phi(D(z_0))-D(z_0)\phi'(D(z_0))}{\phi^2(D(z_0))}\\&=\frac{D'(z_0)-z_0eD'(z_0)}{\mu_1}.\end{aligned}$$

Let $S=D'(z_0)$ and $Q=S-z_0eS$. Therefore the above equation can be written as $H'(z_0)=Q/\mu_1$. Here let it be reminded that we want to prove that $\rho(H'(z_0)=\rho(Q/\mu_1)<1.$ We can achieve this by showing that the spectral radius of Q is equal to $|v_2|$. We can also show that the spectrum of Q is $\{0,v_2,v_3,\dots,v_N\}$.

We have $1=\phi(z_0)=(z_0)_1+(z_0)_2+\dots+(z_0)_N=ez_0$, so $ez_0=1$ and $Q=S-z_0eS,\ Q^Te^T=(S-z_0eS)^Te^T=0.$ We can conclude that e^T is an eigenvector of Q^T, associated with the eigenvalue 0.

There are two possible cases of S^T.

Case 1: The matrix $S^T=D'(z_0)^T$ is diagonizable, that is, S^T is semisimple. For $i=2,3,\dots,N$, we assume $S^Tw^i=v_iw^i$, where w^i is an eigenvector of S^T that is associated with the eigenvalue v_i. Suppose that the set of eigenvector $\{w^1,w^2,\dots,w^N\}$ is linearly independent.

We can write $v_iz_0^Tw^i=z_0^Tv_iw^i=z_0^TS^Tw^i$, for $i=2,3,\dots,N.$ We already have $D'(z_0)z_0=Sz_0=\mu_1 z_0.$ So, $(Sz_0)^T=(\mu_1 z_0)^T$, and

$$z_0^TS^T=\mu_1 z_0^T. \qquad (2)$$

Hence, $v_iz_0^Tw^i=z_0^TS^Tw^i=\mu_1 z_0^Tw^i,\ (v_i-\mu_1)z_0^Tw^i=0.$ So, for $i=2,3,\dots,N,$ it is either $v_i=\mu_1$ or $z_0^Tw^i=0$. However $v_i\ne\mu_1$ for $i=2,3,\dots,N$. Therefore $z_0^Tw^i=0.$

Now we have $Q^Tw^i=(S-z_0eS)^Tw^i=S^Tw^i-0.$ Since we assume that $S^Tw^i=v_iw^i$, so then $Q^Tw^i=v_iw^i$. The vector w^i is an eigenvector of Q^T associated with the eigenvalue v_i for $i = 2,3,\ldots,N.$

Now we prove that the set of eigenvectors of Q, $\{e^T, w^2, w^3, \ldots, w^N\}$ is linearly independent. Suppose that:

$$\alpha_1 e^T + \alpha_2 w^2 + \ldots + \alpha_N w^N = 0, \tag{3}$$

and $v_i \neq 0$ for $i = 2,3,\ldots,p$ and $v_j = 0$ for $j = p+1,\ldots,N.$ We know that $Q^T e^T = 0e^T$ and $Q^T w^j = v_i w^j$ for $i = 2,3,\ldots,N.$ Adding these two equations on LHS and RHS respectively yield:

$$Q^T e^T + Q^T w^2 + \ldots + Q^T w^N = 0e^T + v_2 w^2 + \ldots + v_p w^p.$$

Now, substitute $e^T, w^2, w^3, \ldots, w^N$ with $\alpha_1 e^T, \alpha_2 w^2, \alpha_3 w^3, \ldots, \alpha_N w^N$ of equation (8) and obtain

$$\begin{aligned} &\alpha_1 Q^T e^T + \alpha_2 Q^T w^2 + \ldots + \alpha_N Q^T w^N \\ &= \alpha_2 v_2 w^2 + \ldots + \alpha_p v_p w^p \end{aligned} \tag{4}$$

$$Q^T\left(\alpha_1 e^T + \alpha_2 w^2 + \ldots + \alpha_N w^N\right) = \alpha_2 v_2 w^2 + \ldots + \alpha_p v_p w^p = 0,$$

Since we consider the set $\{w^2, w^3, \ldots, w^N\}$ to be linearly independent, we then get $\alpha_2 = \alpha_3 = \ldots = \alpha_p = 0,$ and we can now write equation (8) as:

$$\alpha_1 e^T + \alpha_{p+1} w^{p+1} + \ldots + \alpha_N w^N = 0, \tag{5}$$

$$S^T\left(\alpha_1 e^T + \alpha_{p+1} w^{p+1} + \ldots + \alpha_N w^N\right) = 0,$$

$$\alpha_1 S^T e^T + \alpha_{p+1} S^T w^{p+1} + \ldots + \alpha_N S^T w^N = 0.$$

Since $S^T w^j = v_i w^j$ for $j = p+1,\ldots,N,$ we then get

$$\alpha_1 S^T e^T + \alpha_{p+1} v_{p+1} w^{p+1} + \ldots + \alpha_N v_N w^N = 0.$$

Since $v_j = 0$ for $j = p+1,\ldots,N,$ it yields

$$\alpha_1 S^T e^T = 0.$$

We then get $\alpha_1 = 0$ since $S^T e^T > 0$ and S is diagonalizable. From equation (10), we have

$$\alpha_{p+1} w^{p+1} + \ldots + \alpha_N w^N = 0. \tag{6}$$

We know that the set $\{w^{p+1}, w^{p+2}, \ldots, w^N\}$ is linearly independent, so $\alpha_{p+1} = \ldots = \alpha_N = 0.$ So we get $\alpha_1 = \ldots = \alpha_N = 0.$ This means that the set $\{e^T, w^2, w^3, \ldots, w^N\}$ is linearly independent and the spectrum of Q is $\{0, v_2, v_3, \ldots, v_N\}.$

Case 2: Consider that S^T is not diagonalizable or defective. We know that a defective matrix has less than N different eigenvalues. Assume that S^T has $q < N$ different eigenvalues where $v_1 = \mu_1, v_2, \ldots, v_q,$ and these eigenvalues can be written as follows:

$$v_1 = \mu_1 > |v_2| \geq |v_3| \geq \ldots \geq |v_q|. \tag{7}$$

So, S^T has the form $S^T = XJX^{-1}$, where the $J = diag\{J_1, J_2, \ldots, J_q\}$ is in a canonical form. Suppose that the square matrices $J_i,\ i = 1,2,\ldots,q$ be the Jordan blocks with various sizes, in the form of:

$$J_i = \begin{bmatrix} v_i & 1 & 0 & \cdots & 0 \\ 0 & v_i & 1 & \ddots & \vdots \\ \vdots & \ddots & v_i & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & v_i \end{bmatrix},$$

where v_i is an eigenvalue of S^T. Let $J_1 = [\mu_1],$ and X_i is the i th column vector of $X,\ i = 1,2,\ldots,N.$ Let l_i be the size of J_i of each Jordan block, where $i = 1,2,\ldots,q.$ We now have $S^T = XJX^{-1}$, and therefore

$$S^T X = XJ,$$

$$S^T\begin{bmatrix} X_1 & X_2 & X_3 & X_4 & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & \cdots \end{bmatrix} \begin{bmatrix} \mu_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 1 & \ddots & \vdots \\ \vdots & \ddots & v_2 & \ddots & 0 \\ \vdots & 0 & \ddots & v_2 & 1 \\ 0 & \cdots & \cdots & 0 & \ddots \end{bmatrix}.$$

From the above equation, we get:

$$\begin{aligned} S^T X_2 &= v_2 X_2, \\ S^T X_3 &= X_2 + v_2 X_3, \\ S^T X_4 &= X_3 + v_2 X_4, \\ &\vdots \\ S^T X_{l_2+1} &= X_{l_2} + v_2 X_{l_2+1}, \\ S^T X_{l_2+2} &= v_3 X_{l_2+2}, \\ S^T X_{l_2+3} &= X_{l_2+1} + v_3 X_{l_2+2}, \\ &\vdots \end{aligned}$$

Just like in Case 1, $S^T X_2 = v_2 X_2$ and based on the equation (7), $v_2 z_0^T X_2 = z_0^T v_2 X_2 = z_0^T S^T X_2 = \mu_1 z_0^T X_2,$ $(v_2 - \mu_1) z_0^T X_2 = 0.$ From equation (12), $v_2 \neq \mu_1.$ So $z_0^T X_2 = 0.$ Hence,

$$Q^T = \left(S - z_0 eS\right)^T,$$

$$Q^T X_2 = \left(S - z_0 eS\right)^T X_2 = S^T X_2 - S^T e^T z_0^T X_2 = S^T X_2 - 0,$$

which means that $Q^T X_2 = v_2 X_2.$ This implies that X_2 is an eigenvector of Q^T associated with the eigenvalue $v_2.$

From the equation $S^T X_3 = X_2 + v_2 X_3,$ we get

$$v_2 z_0^T X_3 = z_0^T v_2 X_3 = z_0^T (S^T X_3 - X_2) = z_0^T S^T X_3 - 0.$$

By equation (7), we get $v_2 z_0^T X_3 = \left(\mu_1 z_0^T\right) X_3.$ Consequently, $(v_2 - \mu_1) z_0^T X_3 = 0.$ By equation (12), and since $v_2 \neq \mu_1,$ we obtain $z_0^T X_3 = 0.$ Therefore,

$$Q^T = \left(S - z_0 eS\right)^T,$$

$$\begin{aligned} Q^T X_3 &= \left(S - z_0 eS\right)^T X_3 = S^T X_3 - S^T e^T z_0^T X_3 \\ &= S^T X_3 = X_2 + v_2 X_3. \end{aligned}$$

Likewise, we obtain:

$$
\begin{aligned}
Q^T X_2 &= v_2 X_2 \\
Q^T X_3 &= X_2 + v_2 X_3 \\
Q^T X_4 &= X_3 + v_2 X_4 \\
&\vdots \\
Q^T X_{l_2+1} &= X_{l_2} + v_2 X_{l_2+1} \\
Q^T X_{l_2+2} &= v_3 X_{l_2+2} \\
Q^T X_{l_2+3} &= X_{l_2+1} + v_3 X_{l_2+2} \\
&\vdots
\end{aligned}
$$

Like in Case 1, we want to show that the set $\{e^T, X_i, i = 2,3,\ldots,N\}$ is linearly independent. Let $Y = [e^T, X_i, i = 2,3,\ldots,N]$. Therefore, $Q^T Y = Ydiag\{[0], J_2, \quad \ldots, J_q\}$. We now have the spectrum of Q, $\{0, v_2, v_3, \ldots, v_q\}$ which is similar to the spectrum of Q^T. The spectral radius of Q is $|v_2|$. Therefore we get the following result:

$$\rho(H'(z_0)) = \rho\left(\frac{Q}{\mu_1}\right) = \frac{|v_2|}{\mu_1} < 1,$$

since $\mu_1 > |v_2|$.

Now we can determine the convergence rate of Algorithm 1.

Theorem 7. Let A and $\{z_0^{(k)}\}$ be as in Theorem 6. Then the convergence rate of the sequence $\{z_0^{(k)}\}$ is Q-linear, which means, there exists a vector norm $\|\bullet\|$ such that

$$\limsup_{k\to\infty} \frac{\|z^{(k+1)} - z_0\|}{\|z^{(k)} - z_0\|} < 1.$$

Proof. By Proposition 1, there exist an $\epsilon > 0$ and a spectral norm $\|\bullet\|$ such that $\|H'(z_0)\|_\epsilon \le \rho(H'(z_0)) + \epsilon$. By Lemma 2:

$$\|H'(z_0)\|_\epsilon \le \rho(H'(z_0)) + \epsilon < 1. \tag{8}$$

Hence, by equation (6), we have $z^{(k+1)} = H(z^{(k)})$, $k = 1,2,\ldots,$ and $z_0 = H(z_0)$. Therefore, $z^{(k+1)} - z_0 = H(z^{(k)}) - H(z_0)$. Expand $z^{(k)}$ at z_0 by using the Taylor expansion, we get:

$$H(z^{(k)}) = H(z_0) + H'(z_0)(z^{(k)} - z_0) + o\left(\|z^{(k)} - z_0\|_\epsilon\right),$$

$$z^{(k+1)} - z_0 = H'(z_0)(z^{(k)} - z_0) + o\left(\|z^{(k)} - z_0\|_\epsilon\right),$$

$$\frac{\|z^{(k+1)} - z_0\|_\epsilon}{\|(z^{(k)} - z_0)\|_\epsilon} = \|H'(z_0)\|_\epsilon,$$

From equation (13), we can get

$$\limsup_{k\to\infty} \frac{\|z^{(k+1)} - z_0\|}{\|z^{(k)} - z_0\|} < 1.$$

Therefore Algorithm 1 is Q-linear convergence.

5.0 CONCLUSION

Within this paper, we proved that the algorithm for finding the largest singular value of nonnegative rectangular tensors, as proposed by Zhou *et al*. [11], is convergent under weak irreducibility condition and has a Q-linear rate of convergence. This paper only presents the convergence properties of Algorithm 1. In regards to numerical tests, the reader can refer to the referenced studies [7, 11].

The study of rectangular tensors is relatively new. Another method for determining the largest singular value of rectangular tensors can be found in Zhang's study [23], and it has been proven to be convergent under some assumptions. Algorithm 1 has also been generalised to nonnegative polynomials, as presented in Ibrahim's study [24]. The method is also convergent.

Acknowledgement

The second author is thankful to the Universiti Pendidikan Sultan Idris (UPSI) and the Ministry of Higher Education (Malaysia) for financial support throughout this study, under the RAGS grant Vot 2014-0122-101-72.

References

[1] Knowles, J. K. and Sternberg, E. 1975. On The Ellipticity of The Equations of Nonlinear Elastostatics for a Specialmaterial. *J. Elasticity*. 5: 341-361.

[2] Knowles, J. K. and Sternberg, E. 1977. On The Failure of Ellipticity of The Equations for Finite Elastostatic Plane Strain. *Arch. Ration. Mech. Anal.* 63: 321-336.

[3] Rosakis, P. 1990. Ellipticity and Deformations with Discontinuous Deformation Gradients in Finite Elastostatics. *Arch. Ration. Mech. Anal.* 109: 1-37.

[4] Wang, Y. and Aron, M. 1996. A Reformulation of The Strong Ellipticity Conditions for Unconstrained Hyperelasticmedia. *J. Elasticity*. 44: 89-96.

[5] Dahl, D., Leinass, J. M. Myrheim, J. and Ovrum, E. 2007. A Tensor Product Matrix Approximation Problem in Quantum Physics. *Linear Algebra Appl.* 420: 711-725.

[6] Einstein, A., Podolsky, B. and Rosen, N. 1935. Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Phys. Rev.* 47: 777-780.

[7] Chang, K., Qi, L. and Zhou, G. 2010. Singular Values of a Real Rectangular Tensor. *Journal of Mathematical Analysis and Applications*. 370: 284-294.

[8] Wood, R. J. and O'Neill, M. J. 2007. Finding The Spectral Radius of A Large Sparse Non-Negative Matrix. *ANZIAM J.* 48: C330-C345.

[9] Wood, R. J. and O'Neill, M. J. 2007. An Always Convergent Method for Finding The Spectral Radius of An Irreducible Non-Negative Matrix. *ANZIAM J.* 45: C474-C485.

[10] Ng, M., Qi, L. and Zhou, G. 2009. Finding The Largest Eigenvalue of A Nonnegative Tensor. *SIAM J. Matrix Anal. Appl*. 31: 1090-1099.

[11] Zhou, G., Caccetta, L. and Qi, L. 2013. Convergence of An Algorithm for The Largest Singular Value of A Nonnegative Rectangular Tensor. *Linear Algebra and its Applications*. 438: 959-968.

[12] Jordehi, A. R. 2015. Enhanced Leader PSO (ELPSO): A New PSO Variant for Solving Global Optimisation Problems. *Applied Soft Computing.* 26: 401-417.

[13] Jordehi, A. R., Jasni, J., Wahab, N. A., Kadir M. Z., and Javadi, M. S. 2015. Enhanced leader PSO (ELPSO): A New Algorithm for Allocating Distributed TCSC's in Power Systems. *International Journal of Electrical Power & Energy Systems.* 64: 771-784.

[14] Jordehi, A. R. 2015. Brainstorm Optimisation Algorithm (BSOA): An efficient Algorithm for Finding Optimal Location and Setting of FACTS Devices in Electric Power Systems. *International Journal of Electrical Power & Energy Systems*. 69: 48-57.

[15] Jordehi, A. R. 2014. A Chaotic-Based Big Bang-Big Crunch Algorithm for Solving Global Optimisation Problems. *Neural Computing and Applications*. 25: 1329-1335.

[16] Jordehi, A. R. 2015. Chaotic Bat Swarm Optimisation (CBSO). *Applied Soft Computing.* 26: 523-530.

[17] Varga, R. 1965. *Matrix Iterative Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Englewood Cliffs.

[18] Minc, H. 1988. *Nonnegative Matrices*. New York: John Wiley and Sons.

[19] Golub, G. H. and Loan, C. F. V. 1996. *Matrix Computations*. Baltimore: Johns Hopkins University Press.

[20] Horn, R. A. and Johnson, C. R. 1985. *Matrix Analysis*. Cambridge University Press.

[21] Friedland, S., Gaubert, S. and Han, L. 2011. Perron-Frobenius Theorem for Nonnegative Multilinear forms and Extensions. *Linear Algebra Appl.* 438: 738-749.

[22] Zhou, G., Qi, L., and Wu, S. 2013. Efficient Algorithms for Computing The Largest Eigenvalue of A Nonnegative Tensor. *Front. Math. China*. 8: 155-168.

[23] Zhang, L. 2013. Linear Convergence of An Algorithm for Largest Singular Value of A Nonnegative Rectangular Tensor. *Front. Math. China*, 8: 141-153.

[24] Ibrahim, N. F. 2014. An Algorithm for The Largest Eigenvalue of Nonhomogeneous Nonnegative polynomials. *Numerical Algebra, Control and Optimization.* 4: 75-91.