ICA Model of Densities for Images (ICA_{DGGDD})

Dheeraj Tandon, Harsh Dev, Deepak Kumar Singh

***Abstract :** In data analysis we use ICA is a basic tool, the aim is that find out a co-ordinate system where are the components of the data are independent.Mostly ICA method such as fastICA and Jointapproximation and diogonalization of eigen matrix (JADE), uses kurtosis as a metric of non gaussianity. But the assumption of kurtosis (fourth order cumulant) may not always satisfies practically. So there are one possible solution is to use skewness (third order cumulant) instead of kurtosis. In this paper we are going to introduce ICA based method, that approach is good for heavy-tailed (fourth order kurtosis) as well as asymmetric data (third order skewness).*

***Keywords:** ICA, Skewness, Kurtosis.*

I. INTRODUCTION

If we want to unsupervised method for data analysis then we use independent component analysis (ICA) method. There are many applications of ICA like finance, fault deduction EEG analysis, MRI etc.

Maximization of non-gaussianity(by using kurtosis) was classical approach used in mostly ICA method like fast ICA, but practically kurtosis sources may not always satisfied, because we assume that underlie density is symmetric but it is not(or really) and data set is bounded. So for weak -kurtosis but skewed sources these methods could fail. Another metric using in ICA is skewness.

There is an approach present by the authors based on the maximum likelihood estimation. In which they found co-ordinate system that is optimally fitted to data and marginal densities also. Here authors model skewness that is based on division gaussian distribution that is well adopted for asymmetric data.

In [5,6] authors used combination of kurtosis and skewness that is projection index. This method find out good results for weak kurtosis and skewed data but modelling is the main difficulty in this approach.

In our work we use approximation the data density by product of division generalized gaussian distribution that is well suited for both model skewness and heavy-tailed (kurtosis) data.

II. RELATED WORK

Herault and Juttr in 1983 where present the firstICA method. They introduce algorithm based on neuro-memetic architecture anditerative real time [7]. By using third order cumulant (skewness) Gian-Nakis et al. find out issue of identifiability of ICAI in 1987, but it requirement and exhaustive search.Mathematical approach to the problem by using higher order statistics, that is measure of fitting independent component was introduced by Lacoume and Ruiz [9]. The algebraic property of fourth order comulant was find out by Cardoso [10,11]. Now a days it is popular approach.

Negentropy measures the fitting of independent component. FastICA uses negentropy. Entropy (differential) is concept of information theory used in negentropy. It measures independence of random variables that is mutual information. When we minimize MI and maximize the negentropy both have the same interpretation, so we can easily calculate without using additional parameters.

Another approach is maximum likelihood estimation to estimate ICA. It is related to infomax. Likelihood is directly proportional to the negative of MI. Now maximum-likelihood is most popular approach in ICA. In our paper we use maximum likelihood principle

III. MAXIMUM LIKELIHOOD APPROACH TO ICA

let a random vector y in R^d, that is generated with density F. We know that component of y is independent iff

$\exists\, 1 - D\ densities\ f_{1,......,}f_d\ \in$
$D_R\ where\ D_R\ is\ set\ of\ densities\ on\ R,$ s.t.
$F(y) = f_1\ (y_1) * \ldots * f_d\ (y_d), for\ y = (y_1, \ldots y_d) \in\ R^d$

let the components of y are not independent but why becomes independent by using basis A ($W=A^{-1}$)

$$F(y) = det\ (W) * f_1\ \left(\omega_1^T(y-m)\right) * \ldots * f_d\ \left(\omega_d^T(y - m, for\ y{=}y1{,}{\ldots}yd \in Rd\right. \quad (1)$$

Where $W = A^{-1}$
ω_i = i^{th} column of W
ω_i^T = convert i^{th} column into vector
m = center of basis
A = basis matrix
$\omega_1^T(y - m)$ = i^{th} coefficient of (y-m)
F $\subsetneq$ D_R (the set of all densities given by (1)

If we find out a basis, so that component becomes independent, we have to search matrix W and one dimensional densities such that the approximation

$$F(y) \approx det\ (W) * f_1\ \left(\omega_1^T(y-m)\right) * \ldots * f_d\ \left(\omega_d^T(y-m)\right), for\ y = (y_1, \ldots y_d) \in\ R^d$$

is optimal.

Revised Manuscript Received on 14 September, 2019.
Dheeraj Tandon, SHIMT (Sitapur), Uttar Pradesh, India.
(Email: dheerajtandon9@gmail.com)
Dr. Harsh Dev, PSIT (Kanpur), Uttar Pradesh, India.
(Email: doctorharshdev@rediffmail.com)
Dr. Deepak Kumar Singh, Integral University (Lucknow), Uttar Pradesh, India.
(Email: d.k.singh@ieee.org)

Maximum likelihood estimation is used to solve above approximation of F(y) that will give formulation of ICA problem.

3.1 Difficulties :

Super Gaussian logistic distribution is very sensitive to outliers and for asymmetric data the approximation cannot give expected result. Divide normal distribution for heavy tails, this is (skew) model is not fit, that are very common in many data sets.

So that we introduce new idea, model of densities that would not have two above disadvantages. We give the concept of estimation of both (Kurtosis and Skewness) measures simultaneously Divide Generalized Gaussian Density Distribution (DGGDD).

The following formula is used for maximizing the MLE for the approximation of the above density equation.

$$\frac{1}{|y|}\sum_{i=1}^{d}\sum_{y\,\in Y}\ln\left(f_i\left(\omega_i^T(y-m)\right)\right)+\ln(\det(W))$$

Where

$y \subsetneq R^d$ (is data set)

$F \subsetneq D_R$ (is set of all densities)

W = Unmixing matrix

m = center of data

$f_{1,\ldots\ldots,}f_d \in F$ (Densities)

should be maximized

In classical ICA we were used super Gaussianlogistic distribution, basic difference among Gaussian and Super Gaussianis presence of heavy tails. In this method limitation is asymmetry. Second methodis based on division of gaussian density, but in this method limitation is heavy tails.Another method is generalized Gaussian density distribution. It is good in comparison tosuper gaussian Logistic model, for symmetric PDF. It is widely used model, but the limitation in this model is asymmetry. Next model we propose is division of generalized gaussiandensity distribution. In this model there is no any limitation for asymmetry. Vigyan abbreviated by DGGDD (divide generalized gaussian density distribution).

A algorithm developed by amari et al. [13],but this algorithm developed independently by cardoso the name given by him relative gradient algorithm [3, 14].

$$W^{K+1} = W^K + \eta\,[I - \varphi(y)y^T]W^K \qquad (2)$$

Where η> 0 called learning rate. How we can choose it give in reference (4) and φ (.) is vector of score function. The optimal components are given by

$$\varphi(y_i) = \frac{d}{dy_i}\log P_i(y_i) = -\frac{P_i'(y_i)}{P_i(y_i)} \qquad (3)$$

IV. SPLIT GAUSSIAN DISTRIBUTION

The most famous distribution approach for skewness (asymmetric data) is SN distributionand for heavy tails is General Gaussian (GG) distribution. our approach is based on Divide generalized Gaussian Density Distribution (DGGDD),that will estimateasymmetric data.

Normal distribution faces the limitation of symmetry, so we control the shape of PDF by calculatingthe deviation of sampleto the mean without modelling tails.

The Gaussian (normal) PDF can be modify by describe the deviation from symmetry. Two different left and right variance in Gaussian PDFreplaces the variance. It will yield asymmetric split Gaussian model.

$$P_{SN}(y; m, \sigma^2, \tau^2) = \begin{cases} d\,.\exp\left[-\frac{1}{2\sigma^2}(y-m)^2\right], where\ y \le m \\ d\,.\exp\left[-\frac{1}{2\sigma^2\tau^2}(y-m)^2\right], where\ y > m \end{cases}$$

Whered $= \sqrt{\frac{2}{\pi}}\sigma^{-1}(1+\tau)^{-1}$

Let γ1 and γ2are two different parameters that calulate the skewness and sharpness of a random variable and distribution.

$\gamma1 = \text{skew}(y) = \frac{m3\,(y)}{\sqrt{m_2^3(y)}}$;

$\gamma2 = \text{kurt}(y) = \frac{m4\,(y)}{\sqrt{m_2^2(y)}} - 3$,

where

$m_2(y)$ = Second order moment,

$m_3(y)$ = Third order moment,

$m_4(y)$ = Fourth order moment

the values

γ1 (skewness) = 0 (for symmetric distribution)

γ1 (skewness) = Non zero (for asymmetric distribution)

and

γ2 (kurtosis) = +ve (super - gaussian)

γ2 (kurtosis) = 0 (gaussian)

γ2 (kurtosis) = -ve (sub - gaussian)

The relationship between γ1 and γ2 and freedom degree parameters in GG and DGGDD we can determined in next section

Kurtosis another measure of non-gaussianity in terms of shape. The value of kurtosis is 3 for Gaussian, if the kurtosis is larger (smaller) than 3 then the sharpness of PDF shape is higher (lower) than the corresponding gaussian function. Generalized symmetric PDF is a good model for variable sharpness. The most famous model used generalized symmetric PDF model with variable sharpness is the generalized Gaussian [17, 18]

$$P_{GG}(y; m, \alpha, d) = \frac{d}{2\alpha\,\Gamma\left(\frac{1}{d}\right)}\exp\left[-\frac{|y-m|^d}{a^d}\right]$$

For

$\alpha, d \in R_+ \ and\ m \in R$ where

$\Gamma(.)$ = Gamma function, $\quad F(y) = \int_0^\infty t^{y-1}e^{-t}\,dt$

d > 0 (influence the model sharpness)

when we change the value of d (d>0), we get different sharpness in the family of distributions.

when d = 2 (distribution PDF reduced to gaussian)

d > 0 (distribution PDF becomes sub gaussian)

that is γ2< 0 (not heavy tail)

otherwise d < 2 (distribution PDF become super gaussian)

that is γ2 > 0 (heavy tail).

d can not be determined directly from data samples but we can obtain kurtosis (γ2). For this reason we (y)must have the relationship between c and γ2 (kurtosis). From definition the second and fourth order moment can obtain

$$m2 = \int_{-\infty}^{\infty} y^2 P_{GG} dy = \frac{\Gamma\left(\frac{3}{d}\right)}{\gamma^2 \Gamma\left(\frac{1}{d}\right)}, \quad m4 = \int_{-\infty}^{\infty} y^4 P_{GG}(y) dy = \frac{\Gamma\left(\frac{5}{d}\right)}{\gamma^4 \Gamma\left(\frac{1}{d}\right)}$$

So, $\gamma = \sqrt{\frac{\Gamma\left(\frac{3}{d}\right)}{m\Gamma\left(\frac{1}{d}\right)}}$,

by using the definition of $\gamma 2$, $\gamma 2 = \frac{m4}{m_2^2} - 3 = \frac{\Gamma\left(\frac{5}{d}\right)\Gamma\left(\frac{1}{d}\right)}{\Gamma^2\left(\frac{3}{d}\right)} - 3$ (4)

By the value of $\gamma 2$ we can get value of d by above equation but due to definition of gamma function ($\Gamma(.)$) it is not possible to represent d in terms of γ2 (kurtosis), analytically exact expression. So that we take approximation. Our best approximation was found by (least square method) LSM [19]. The LSM estimate the following result

$$d = \sqrt{\frac{5}{\gamma 2 - 1.865}} - 0.12 \quad (5)$$

by above analysis the score function based on generalized gaussian density PDF in the algorithm equation (2) can be obtained from equation (3)32

$$\varphi_{i,GG}(y_i) = d\,\gamma^d |y_i - \mu y_1|^{d-1} sgn\,(y_i - \mu y_1) (6)$$

V. DGGDD MODEL

Symmetry is the main limitation of general gaussian model so left and right variance are replaced by the variance in the gaussian PDF to obtain the asymmetric gaussian model and we obtain asymmetric division generalized gaussian model

Choi et. al. [16] given the idea of ICA algorithm when the score function is obtained from DGGD model. This function is controlled by d. The limitation of GGDD model is it symmetry but DGGDD model is based on asymmetry DGGDD. This model is based on two 2[nd]order parameters σ_l^2, σ_r^2 are called left and right variance they are represented and given by

$$\widehat{\sigma_l^2} = \frac{1}{N_l - 1} \sum_{i=1,\, y_i < m_y}^{N_l} \left(y_i - \widehat{m_y}\right)^2 \quad and$$

$$\widehat{\sigma_r^2} = \frac{1}{N_r - 1} \sum_{i=1,\, y_i > m_y}^{N_r} \left(y_i - \widehat{m_y}\right)^2$$

Where m_y = estimated mode (that must not conside with mean in asymmetric distribution

$N_l(N_r)$

$= Number\ of\ y_i\ less\ than\ \widehat{m_y}\ \ (or\ greater\ than\ \widehat{m_y})$

So now replace the variance of General Gaussian Distribution PDF with $\sigma_l$and σ_r ,we can get DGGDD model.

$$P_{DGGDD}(y; m, \sigma_l, \sigma_r, d) = \begin{cases} \frac{c\gamma}{\Gamma\left(\frac{1}{d}\right)} .\exp\left[-\gamma_l^d\left[-\left(y - m_y\right)\right]^d\right], & where\ y < m_y \\ \frac{c\gamma}{\Gamma\left(\frac{1}{d}\right)} .\exp\left[-\gamma_r^d\left[-\left(y - m_y\right)\right]^d\right], & where\ y \geq m_y \end{cases} \quad 7)$$

Where $\gamma = \frac{1}{\sigma_l + \sigma_r}\sqrt{\frac{\Gamma\left(\frac{3}{d}\right)}{\Gamma\left(\frac{1}{d}\right)}}$, $\gamma_l = \frac{1}{\sigma_l}\sqrt{\frac{\Gamma\left(\frac{3}{d}\right)}{\Gamma\left(\frac{1}{d}\right)}}$, $\gamma_r = \frac{1}{\sigma_r}\sqrt{\frac{\Gamma\left(\frac{3}{d}\right)}{\Gamma\left(\frac{1}{d}\right)}}$ $and\ m_y = mode$

from definition DGGDD model satisfies following properties of density function.

$$P_{DGGDD}(y; m, \sigma_l, \sigma_r, d) > 0, \int_{-\infty}^{\infty} P_{DGGDD}(y)\ dy = 1$$

same time PDGGDD is continuous when y = m_y from above

CASE 1: when $\sigma_l^2 = \sigma_r^2$the PDF is same as GGD model (that is symmetric distribution). $\sigma_l^2 = \sigma_r^2$ and γ_2 (Kurtosis) = 0 then DGGDD model consider with Gaussian model and if γ_2 (kurtosis) > 0 it consider with super gaussian model otherwise γ_2 (kurtosis) < 0 it consider with sub gaussian PDF.

CASE 2: when $\sigma_l^2 \neq \sigma_r^2$ and γ_2 = 0 then we get general asymmetric gaussian. From the definition second and fourth order moment of DGGDD can be obtained.

$$m2 = (\sigma_r - \sigma_l)^2 \left(1 - \frac{\left(\Gamma\left(\frac{2}{d}\right)\right)^2}{\Gamma\left(\frac{1}{d}\right)\Gamma\left(\frac{3}{d}\right)}\right) + \sigma_l\sigma_r \quad (8)$$

$$m4 = \frac{\Gamma\left(\frac{5}{d}\right)\Gamma(d)}{\left(\Gamma\left(\frac{3}{d}\right)\right)^2}(\sigma_l^4 - \sigma_r^3\sigma_l + \sigma_r^2\sigma_l^2 - \sigma_r\sigma_l^3 + \sigma_l^4) - 4\frac{\Gamma\left(\frac{2}{d}\right)\Gamma(4)}{\left(\Gamma\left(\frac{3}{d}\right)\right)^2}(\sigma_r^2 + \sigma_l^2)(\sigma_r - \sigma_l)^2 + 6\frac{\left(\Gamma\left(\frac{2}{d}\right)\right)^2}{\Gamma\left(\frac{3}{d}\right)\Gamma\left(\frac{1}{d}\right)}(\sigma_r^2 - \sigma_r\sigma_l +$$

σl2σr− σl2−3 Γ2d2Γ3dΓ1d2σr− σl4 (9)

Relationship between γ_2 (kurtosis) and d can be derived from equation (4), equation (8) and equation (9) from equation (3) we obtain score function for DGGDD.

$$\varphi_{i,DGGDD}(y_i) = \begin{cases} -d\,\gamma_l^d\left(-\left(y_i - m_{yi}\right)\right)^{d-1} & y_i < m_{yi} \\ d\,\gamma_r^d\left(y_i - m_{yi}\right)^{d-1} & y_i \geq m_{yi} \end{cases}$$

(10)

VI. MULTIDIMENSIONAL DIVIDE GENERALIZED GAUSSIAN DISTRIBUTION

The density of multidimensional divide generalized gaussian distribution is given by

$P_{DGGDD}(y; m, \sigma_l, \sigma_r, d) = |\det(W)| \prod_{i=1}^{d} DGG\left(\omega_j^T(y -\right.$ *m ;0, σlj, σrj, d,* (11) Where

$$m = (m_1, \ldots, m_d)^T, \sigma_l = (\sigma_{l1}, \ldots, \sigma_{ld}), \sigma_r = (\sigma_{r1}, \ldots, \sigma_{rd}) \text{ and}$$

$$\omega_j \text{ is } i - th \text{ column of non singular matrix } W \text{ and } d \text{ is a constant.}$$

VII. PROPOSED ALGORITHM

1. Calculate output F(y) by observation f(y) and matrix W
2. Approximate the skewness and kurtosis of mixed image F(y)
3. When value of skewness = 0 or very close to zero
then if $\sigma_l^2 = \sigma_r^2$
find c from equation (5) using equation (4) according the value $\widehat{\gamma_2}$
else if $\gamma_1\ (skewness) > 0.1$
find c from the equation (5) using equation (4), equation (8) and equation (9) according the value $\widehat{\gamma_2}$.
4. Calculate the score function by equation (6) or equation (10).
5. updating W using equation (2)
If $\|W^{K+1} - W^K\| \geq \epsilon$, K = K + 1
then go to step 1 otherwise
Exit

VIII. EXPERIMENTS AND RESULTS

For comparison our method with classical once we use Tuskers Congress Coefficient [61] defined by

$$C_r(S_i, \bar{S_i}) = \frac{\sum_{i=1}^{d} S_i^j S_i^j}{\sqrt{\sum_{j=1}^{d} S_i^j \sum_{j=1}^{d} S_i^{-j}}}$$

Where $-1 \leq C_r \leq +1$. It is used to compare the similarity of extracted factor between different sample. Mostly congruence coefficient = 0.9 (high degree of factor similarity) while congruence coefficient = 0.95 or higher (factor are virtually identical)

8.1Computation Time :

ICADGGDD gives comparable result with NGPP if we increase the number of component to 2 to 40 fastICA, infomaxand JADE are effective in respect to computation time but do not solve image separation

8.2 Images Separation :

ICADGGDD method essentially gives good result in comparison to ICAGG, fastICA, infomax, JADE.

Table I : Mixing and Separation of two images

Img No.	Original Images	Sum & Substraction of Images	Separated by ICA_{SGG}	Separated by ICA_{DGGDD}
1				
2				

IX. CONCLUSION

Our approach based on data density by product of division generalized gaussian density distribution which is suitable for both skewness and heavy tail model. It give the better result in image separation but worst computational time in comparison with classical algorithms.

we verify our approach on images only. This algorithm better recover original images. We used the skewed and heavy tail data only.

REFERENCES

1. P. Spurek, J. Tabor et al., ICA based on data asymmetry, arXiv:1802.05550v1 [stat.ML] 14 Feb 2018
2. Fasong Wang, Hongwei Li, Adaptive ICA Algorithm Based on Asymmetric GeneralizedGaussian Density Model , National Natural Science Foundation of China(GrantNo.60472062) and Natural Science Foundation of Hubei Province, China(GrantNo.2004ABA038).
3. Relative Gradient Learning for Independent Subspace Analysis, Heeyoul "Henry Choi etal., Conference Paper • January 2006
4. P. Spurek, J. Tabor, P. Rola, M. Ociepka, Ica based on asymmetry, Pattern Recognition67 (2017) 230–244.
5. T. Blaschke, L.Wiskott, Cubica: Independent component analysis by simultaneous third-and fourth-order cumulant diagonalization, IEEE Transactions on Signal Processing 52(5) (2004) 1250–1256.
6. J. Virta, K. Nordhausen, H. Oja, Projection pursuit for non-gaussian independentcomponents, arXiv preprint arXiv:1612.05445.
7. C. Jutten, J. Herault, Blind separation of sources, part i: An adaptive algorithm based onneuromimetic architecture, Signal processing 24 (1) (1991) 1–10.
8. G. B. Giannakis, Y. Inouye, J. M. Mendel, Cumulant based identification of multichannelmoving-average models, Automatic Control, IEEE Transactions on 34 (7) (1989) 783–787.
9. J.-L. Lacoume, P. Ruiz, Separation of independent sources from correlated inputs, SignalProcessing, IEEE Transactions on 40 (12) (1992) 3074–3078.
10. J.-F. Cardoso, Super-symmetric decomposition of the fourth-order cumulant tensor.blind identification of more sources than sensors, in: Acoustics, Speech, and SignalProcessing, 1991. ICASSP-91., 1991 International Conference on, IEEE, 1991, pp.3109–3112.
11. J.-F. Cardoso, High-order contrasts for independent component analysis, Neuralcomputation 11 (1) (1999) 157–192.
12. M. Gaeta, J.-L. Lacoume, et al., Source separation without a priori knowledge: themaximum likelihood solution, in: Proc. EUSIPCO, Vol. 90, Barcelona, Spain, 1990, pp.621–624.
13. Amari, S.(1998)Natural gradient works efficiently in learning. Neural Computation 10:251-276
14. Cardoso, J.F., Laheld, B.(1996)Equivariant adaptive source separation. IEEE Trans. onSignal Processing 44: 3017-3030
15. Chamber, J.A., Jafari, M.G., McLaughlin, S.(2004)Variable step-size EASI algorithmfor sequential blind source separation. Electronics Letters 40: 393-394

16. Cao, J., Murata, N., Amari, S., Cichocki, A., Takeda, T. (2003)A Robust Approach toIndependent Component Analysis of Signals With High-Level Noise Measurements.IEEE Tans. on Neural Network 14: 631-645
17. A. K. Nandi, D. M¨ampel, An extension of the generalized gaussian distribution toinclude asymmetry, Journal of the Franklin Institute 332 (1) (1995) 67–75.
18. J. Miller, J. Thomas, Detectors for discrete-time signals in non-gaussian noise, IEEETransactions on Information Theory 18 (2) (1972) 241–250.
19. Tesei, A., Regazzoni, C.S.(1996)Use of four-order statistics for non-Gaussian noisemodeling: the generalized Gaussian Pdf in terms of kurtosis. In: Ramponi, G.,Sicuranza, G.L., Carrato, S., Marsi, S.(eds.) Proc. EUSIPCO'96. EdizioniLINT, pp. 671- 674.

AUTHOR PROFILE

Dheeraj Tandon is a Associate Professor in Dept. of Master of Computer Application, he is also holding a position of HOD-MCA at Sacred Heart Institute of Management & Technology, Sitapur, India. He completed B.E. in Computer Science & Engineering from G. B. Pant Engineering College, Pauri-Garhwal, India and completed M.Tech in Computer Science & Engineering from HBTI, Kanpur, India. He has published more than five papers in journals and conferences and various seminars organized by professional bodies. Currently he is pursuing Ph.D. From Dr. APJAKTU, Lucknow, India.

Dr. Harsh Dev is a Professor in Dept. Of Computer Science and Engineering, he is also holding a position of Dean Research at Pranveer Singh Institute of technology, Kanpur, India. He got his M.Sc. Degree in 1995 from Lucknow University, India & Ph.D. degree in Computer Science in 2005 from Babasaheb Bhimrao Ambedkar University, Lucknow, India. He has 23 years of teaching experience and 16 years of research experience in the field of Computer Graphics, Cryptography, Software Engineering and Data Mining. He has publishedmore than 40 International and National publications. Seven students have been awarded Ph.D. degree in Computer Science under his guidance. He is a member of editorial board of reputed Journal. He is a member of Cofmputer Society of India, Indian Science Congress and International Association of Engineers (IAENG).

Dr. Deepak Kumar Singh completed his doctorate from Indian Institute of Technology, Kanpur and has over eight years experience in the field of academics and IT industry. His core areas of expertise include IoT & real-life applications of artificial intelligence. Apart from teaching in a prominent institution in Lucknow, he is also consultant of several IT firms. He has published more than fifteen papers in international journals and conferences and various seminars organized by professional bodies and industry associations worldwide. Dr. Deepak, holds professional associations with IEEE Technical Committee, IEEE Young Professionals and Association of Knowledge Workers of Lucknow