An Efficient Pairwise Kurtosis Optimization Algorithm for Independent Component Analysis

Fei Ge and Jinwen Ma*

Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing, 100871, China
felixge@math.pku.edu.cn, jwma@math.pku.edu.cn

Abstract. In the framework of Independent Component Analysis (ICA), kurtosis has been used widely in designing source separation algorithms. In fact, the sum of absolute kurtosis values of all the output components is an effective objective function for separating arbitrary sources. In this paper, we propose an efficient ICA algorithm via a modified Jacobi optimization procedure on the kurtosis-sum objective he6re 7584(m)-7o.4(i)(e)dteer6ottCo1.997sel.2(G(C)2.f(g)-3

n be e l o ed n e n l. ne o ble w o e o o h d- nd fo h- o de m l n nd o e he ob b l de n f n on (df) b n e d ewo h ex n on 2 o m- h e ex n on 4 , m le m l n n bee l om ed. he e e m en le d o he h h-o de I m e h od .

De e he n of he df' on, he h h-o de m e h od e l ble n e ll o m l n n h he o e om onen e m ll nde e nden . om on 2 o ed h he m of e of fo h- o de m n l nd d ed m l n (o e) on f n on. e- m bl , b m x m n he e of fo h o de m n l m l n , Delfo e nd o b on o o ed defl on e on he e o e en ll ex- he o e n l , w h he e ll n eed on e en e o o e ol on . Unfo n el , he o e on e en e n he defl on m ode doe no n ee he m e l n m m e l (m l neo e on) fo m .

ll , m o e n fi n o on de he m m e l e on of he o e . In o e o wo , we n l ed he m of b ol e l e of o e of he o om on en on o ob e e f n on fo wh ened ob e on . e efe ed o - m ob e e f n on nd o ed h h no o m x m fo wo- o e m x n oble m . ll , m o e nd m h l o ded h ob e e f n on nd o o ed elf- d e m x m on l o h m , wh h m e ed he de m x n m x . l n o on b w oo om l ed fo m le m en on.

In h e , we o o e n effi en w e J ob o m on l o h m fo I m od fied J ob o ed e on he o - m ob e e f n on. m x m n he m of b ol e l e of o e of n of o om on en , he oble m ed ed o fi nd n n o m l o on n le, wh h n be ol ed d e l f om d . ollow n he J ob o ed e , he o - m ob e e f n on n be m x m ed b e e of l n o on . o e he om on o , we m e om e m od fi on on . he m l on ex e m en h ow h o o o ed I l o h m m o e om on ll effi en h n he I l o h m nde he m e e on e fo m n e.

In he e el, m x m n he o - m ob e e f n on n he w e o e n m ode n od ed n e on 2. In e on 3, we e en he w e o o m on l o h m fo I n he fo m of he m od fied J ob o - m on o ed e. hen n e on 4 he om on om lex n l ed nd om ed w h he I l o h m w h h d owe n on n e m l on ex e m en . e fin ll e b ef on l on n e on .

2 Pair ise Op imi a ion of he K r os S m

o e fo m o e e on f om wh ened ob e on x, we e o l o- o ed o w h n l o h m o m x m e he o - m ob e e f n on

$$J(\mathbf{W}) \quad \sum_{i=1}^{n} | \quad \{y_i\}| \quad \sum_{i=1}^{n} |E\{(\mathbf{w}_i^T\mathbf{x})^4 - 3\}| \ , \tag{1}$$

under the orthonormal constraint $\mathbf{W}^T\mathbf{W} = \mathbf{I}$, where $cum\{\cdot\}$ and $E\{\cdot\}$ denote the fourth-order cumulant and expectation of random variable, respectively. In order to apply the Jacobi-like algorithm, we consider the learning rule, and expect it to be applied to the orthonormal rows of $\mathbf{W}$. Since $\mathbf{W}$ is orthonormally constrained, in the learning procedure one rotation, i.e., only two rows of $\mathbf{W}$ are updated at each step by a Givens rotation. This procedure is also known as Jacobi algorithm, and all pairs of rows must be processed repeatedly until no further changes are required.

The point of the Jacobi-like algorithm is as follows. In Eq. (1), when only two rows of $\mathbf{W}$ are updated under the orthonormal constraint, that is, when the l-th and k-th rows are rotated in a 2-D plane:

$$\begin{cases} \mathbf{w}'_l = \mathbf{w}_l \cos\theta + \mathbf{w}_k \sin\theta , \\ \mathbf{w}'_k = -\mathbf{w}_l \sin\theta + \mathbf{w}_k \cos\theta , \end{cases} \tag{2}$$

the $n-2$ elements $|cum\{y_i\}|$ $(i \neq l, i \neq k)$ in Eq. (1) will remain unchanged. Therefore, to process only the (l,k) pair, maximizing Eq. (1) is reduced to maximize

$$\begin{aligned} J_{l,k}(\theta) &= |E\{(\mathbf{w}_l'^T\mathbf{x})^4 - 3\}| + |E\{(\mathbf{w}_k'^T\mathbf{x})^4 - 3\}| \\ &= |E\{(y_l \cos\theta + y_k \sin\theta)^4\} - 3| + |E\{(-y_l \sin\theta + y_k \cos\theta)^4\} - 3|. \end{aligned} \tag{3}$$

This function is absolute, so we need some techniques to find the optimal solution of maximization; we define

$$J_1(\theta) = E\{(y_l \cos\theta + y_k \sin\theta)^4 + (y_k \cos\theta - y_l \sin\theta)^4\} - 6 , \tag{4}$$
$$J_2(\theta) = E\{(y_l \cos\theta + y_k \sin\theta)^4 - (y_k \cos\theta - y_l \sin\theta)^4\} , \tag{5}$$

and consider the learning rule form:

$$\theta^* = \arg\max_\theta \max\{|J_1(\theta)|, |J_2(\theta)|\} \tag{6}$$
$$= \begin{cases} \arg\max_\theta |J_1(\theta)|, & \text{if } \max_\theta |J_1(\theta)| > \max_\theta |J_2(\theta)|, \\ \arg\max_\theta |J_2(\theta)|, & \text{otherwise}. \end{cases} \tag{7}$$

It can be proved that $J_1(\theta)$ and $J_2(\theta)$ are sinusoidal functions as follows:

$$J_1(\theta) = A \sin(4\theta + \alpha) + c , \tag{8}$$
$$J_2(\theta) = B \sin(2\theta + \beta) , \tag{9}$$

with $A \geq 0, B \geq 0, c, \alpha, \beta$ being the parameters related to y_l and y_k. Note that when the expectations in Eq. (3) are replaced by the sample means, these expressions are still valid. The convenient way of determining these parameters is to connect Eq. (4)(5) with Eq. (8)(9) and let $\theta = 0$ to obtain

$$J_1(\theta)|_{\theta=0} = E\{y_l^4 + y_k^4\} - 6 = A \sin(\alpha) + c , \tag{10}$$
$$J_2(\theta)|_{\theta=0} = E\{y_l^4 - y_k^4\} = B \sin(\beta) , \tag{11}$$
$$J_1'(\theta)|_{\theta=0} = E\{4y_l^3 y_k - 4y_k^3 y_l\} = 4A \cos(\alpha) , \tag{12}$$
$$J_2'(\theta)|_{\theta=0} = E\{4y_l^3 y_k + 4y_k^3 y_l\} = 2B \cos(\beta) , \tag{13}$$
$$J_1''(\theta)|_{\theta=0} = E\{24 y_l^2 y_k^2 - 4y_k^4 - 4y_l^4\} = -16A \sin(\alpha) . \tag{14}$$

he bo e fi e e on e o o de e e $J(\theta)$ d $J(\theta)$. he e e e be ol ed follow

$$c \quad \frac{3}{4}E\{y_l^4+y_k^4\}+\frac{3}{2}E\{y_l^2y_k^2\} - \;, \qquad (1\ )$$

$$A \quad \sqrt{(E\{y_l^4+y_k^4\} - - c)^2 + (E\{y_l^3y_k - y_k^3y_l\})^2}\;, \qquad (1\ )$$

$$B \quad \sqrt{(E\{y_l^4-y_k^4\})^2 + (2E\{y_l^3y_k + y_k^3y_l\})^2}\;, \qquad (1\ )$$

$$\alpha \quad \begin{cases} {}^{-1}((E\{y_l^4+y_k^4\} - - c)/A), & \text{f o } \alpha > 0, \\ \pi - {}^{-1}((E\{y_l^4+y_k^4\} - - c)/A), & \text{o he w e,} \end{cases} \qquad (1\ )$$

$$\beta \quad \begin{cases} {}^{-1}(E\{y_l^4-y_k^4\}/B), & \text{f o } \beta > 0, \\ \pi - {}^{-1}(E\{y_l^4-y_k^4\}/B), & \text{o he w e.} \end{cases} \qquad (1\ )$$

ll we eed o o e $J(\theta)$ d $J(\theta)$, d he lo e he x of E . (3). o d E . ( ), f $|c|+A>B$, we ho ld hoo e

$$\theta \quad \max_\theta |J(\theta)| \quad \begin{cases} (\frac{\pi}{2}-\alpha)/4, & \text{f } c \geq 0, \\ (-\frac{\pi}{2}-\alpha)/4, & \text{f } c < 0. \end{cases} \qquad (20)$$

o he w e we ho ld hoo e

$$\theta \quad \max_\theta |J(\theta)| \quad (\frac{\pi}{2}-\beta)/2, \qquad (21)$$

he θ be he o l o le fo he b- oble . I , e e he fi e fo h-o de o e

$$E\{y_l^4\},\ E\{y_k^4\},\ E\{y_l^3y_k\},\ E\{y_ly_k^3\} \text{ d } E\{y_l^2y_k^2\} \qquad (22)$$

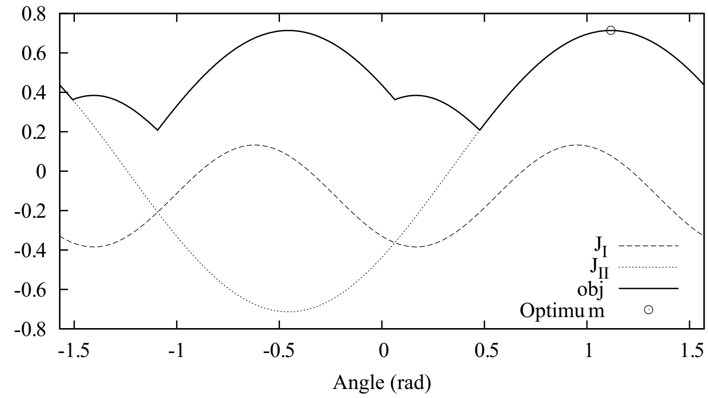
o l l e lo ed-fo ol o of θ. ll he le o e w ll be ed e d. I he follow we de o e he o e o d le o e of E . (22) b $\mu_{4,0}$, $\mu_{0,4}$, $\mu_{3,1}$, $\mu_{1,3}$ d $\mu_{2,2}$, e e el .

he le α d β l o be l l ed b he e e e f o . If θ de e ed E . (20), he ol o e le o he E e o 10 . If θ de e ed E . (21), e le o he E e o 11 . Howe e E d E e o we e de ed de wo- o e x e . ol o h e o be h b d of he e wo e o , E . ( ) e e o e w h e h . U e o e o l o h , he e o eed o e l e he w h oeffi e " e el .

o de o e, we e h . 1 he ob e e f o (3) well $J(\theta)$, $J(\theta)$ d he l l ed o l le, f o w e e of fo o e x oble . he le o e we e $\mu_{4,0}$ 2. 1 , $\mu_{0,4}$ 3.0 3 , $\mu_{3,1}$ 0.0 1 0 , $\mu_{1,3}$ 0.220 4 d $\mu_{2,2}$ 1.0 0 , l l ed ll 4000 le .

3 Pair ise K r o Op imi a ion Algori hm

o n o o o e , f ll J ob wee o $n(n-1)/2$ o e fo ll of o e . If θ o ffi e l lo e o 0 o $\pm\pi/2$,



F . 1. E ample of pair ise sub-problem solution

h r he n ro ed. nd rd J ob l o r h m e r f o r m J ob wee one f e r no h e r, n l no r h bee n ro ed n one wee .

o r m x m n h e r o - m ob e e f n on rw e ro e n n b h m ode, h e o e r ll o m on o l m o h e o of l l n h e m le m o m e n n d m n h n e (ro on) o h e d . h e n d r d J ob l o r h m h n ob o d r w b wh e n o m n r l r r (l, k), f h e n e h e r h e d $y_l(t)$ n o r h e d $y_k(t)$ h bee n h n ed n e h e l e ro on of h r, h r l re d o m l n d l l n θ m e l d 0, $\pi/2$ o r $-\pi/2$, wh h ele .

o o d h n n e e r l l on , we n ee fl $F(l, k)$ f o r e h r (l, k) n d h e bef o r e ro on e m . If h e fl l re d m r ed do n e", do n o l l e θ o h e r w e m r h fl do n e" n d on n e l. h e n e e r ro on l ed o r (p, q), l e r ll fl $F(p, i)$ n d $F(j, q)$ (f o r ll i / p, j / p) be e h ro on ffe o m e o h e r r . h e e r m n n r e r o n f o r h e J ob l o r h m be o m e wh e n h e fl of ll r re m r ed do n e". h m od fi on doe n o h n e h e beh o r of h e J ob l o r h m.

In m l on ex e r m e n we l o ob e r ed h l h ro on n o l el o ffe o h e r r h re o m l e, h we dded n h e r r l e o f r h e r e o m e o m on o , w h o n fi n de r d on of e - r on l . h r l e h f θ n o f r f r o m 0, $\pi/2$ o r $-\pi/2$ (b ro on ll n eeded), do n o l e r n fl . o m m r e, o r rw e r o o - m on l o r h m . h e m od fied J ob ro ed re n be l ed f o ll ow .

INPUT wh e n ed ob e r on d $\{\mathbf{y}(t)\}$, t $1, \ldots, N$.
( ) *Initialize* e h e fl $F(i, j)$ 0, e $\mathbf{h}$ $(h_1, \ldots, h_n)^T$ $\mathbf{0}$.
( ) e r f o r m J ob wee h o h e h r n h e o r de r $(1, 2), (1, 3), \ldots,$ $(1, n), (2, 3), \ldots, (2, n), \ldots, (n-1, n)$. o r e h r (p, q), do
( ) If $F(p, q)$ 1, re r n o e ( ) n d ele h e n ex r

(b) l l e $a(t)$ $y_p(t)^2$, $b(t)$ $y_q(t)^2$, $c(t)$ $y_p(t)y_q(t)$ $(t$ $1,\ldots,N)$.
( ) If h_p 0, l l e $\mu_{4,0}$ $\sum_{t=1}^{N} a(t)^2/N$, o e w e e $\mu_{4,0}$ h_p .
If h_q 0, l l e $\mu_{0,4}$ $\sum_{t=1}^{N} b(t)^2/N$, o e w e e $\mu_{0,4}$ h_q .
l l e $\mu_{2,2}$ $\sum_{t=1}^{N} c(t)^2/N$, $\mu_{3,1}$ $\sum_{t=1}^{N} a(t)c(t)/N$,
$\mu_{1,3}$ $\sum_{t=1}^{N} b(t)c(t)/N$.
(d) b θ o d . (1)-(21). d θ $|\theta| \leq \pi/4$.
(e) If $|\theta| < \theta$, o o ( ).
(f) o e e d $\{y_p(t), y_q(t)\}'$ $\{y_p(t)$ o $\theta + y_q(t)$ θ,
$-y_p(t)$ $\theta + y_q(t)$ o $\theta\}$ $(t$ $1,\ldots,N)$.
l l e h_p $\mu_{4,0}$ o $^4\theta + 4\mu_{3,1}$ o $^3\theta$ $\theta +$ $\mu_{2,2}$ o $^2\theta$ $^2\theta$
$+4\mu_{1,3}$ o θ $^3\theta + \mu_{0,4}$ $^4\theta - 3$,
d h_q $\mu_{4,0}$ $^4\theta - 4\mu_{3,1}$ $^3\theta$ o $\theta +$ $\mu_{2,2}$ o $^2\theta$ $^2\theta$
$-4\mu_{1,3}$ θ o $^3\theta + \mu_{0,4}$ o $^4\theta - 3$.
( ) If $|\theta| > \theta$, le ll fl $F(p,i)$ $(i \; / \; q)$ d $F(j,q)$ $(j \; / \; p)$.
( ) e $F(p,q)$ 1.
( ) If $F(i,j)$ 1 fo ll $j > i$, o . e w e o o ( ).

I e bo e l o , θ e e old e e fo e o o , d θ $> \theta$ e ble o e le. e e o h o e e l l ed l fo o de le o e , w be o ed w o olde l e of e o e , b ex d

$$E\{y_p'^4\} \quad E\{(y_p \;\; o \; \theta + y_q \;\; \theta)^4\}, \quad E\{y_q'^4\} \quad E\{(-y_p \;\; \theta + y_q \;\; o \; \theta)^4\} \; . \quad (23)$$

4 Comp a ion Comple i and Sim a ion

ow we l e e o o o lex of o o o ed l o . e o o o e l ed flo (flo o o e o , defi ed l l o followed b dd o 2). o de $N \gg n$, e o e f ll J ob wee ox el eed $Nn(n-1)$ flo $3N$ flo e (.b), o e $3N$ flo e (.) o l ex e fo e e l l e e N o $4N$, d $4N$ flo e (.f). o w e o o o l o w ll o e o o e , o e o l be of flo e e d be e ed b l o .

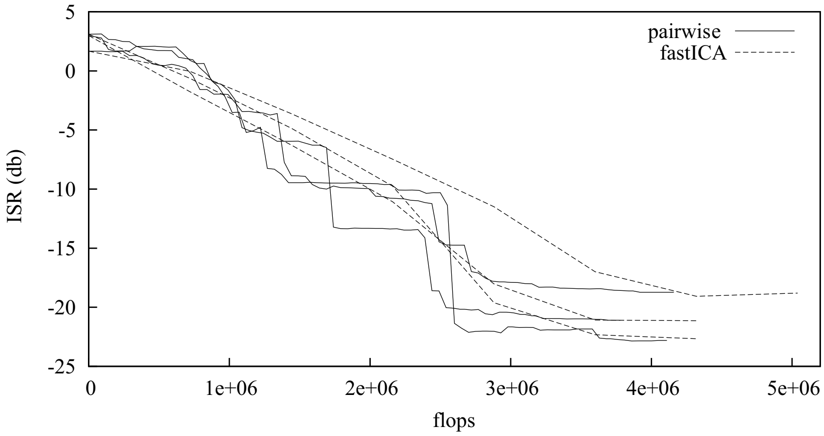
o o o , we l o o de e f o I l o o e ode w e d owe o l e . e o o o of I l o e o e fixed- o e o e

$$\mathbf{w}_i^{+} \quad \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t) y_i(t)^3 \;\; - 3\mathbf{w}_i \;\; , \qquad (24)$$

w eed $2nN + 2N$ flo . o d e w ole x $\mathbf{W}$ e $n(2n+2)N$ flo . e I l o ll o e e le we e o , w e effi e .

T b 1. Source separation results from 100 Monte Carlo simulation runs

	Algorithm	ISR			Flops
		0.25 quantile	median	0.75 quantile	(average)
$n = 4$	FastICA	-27.60 dB	-24.77 dB	-21.76 dB	8.92×10^5
	pair ise	-27.66 dB	-24.79 dB	-21.84 dB	6.635×10^5
$n = 8$	FastICA	-21.38 dB	-19.74 dB	-18.51 dB	4.356×10^6
	pair ise	-21.48 dB	-19.88 dB	-18.51 dB	3.927×10^6
$n = 16$	FastICA	-16.86 dB	-15.76 dB	-14.43 dB	2.614×10^7
	pair ise	-16.95 dB	-15.94 dB	-14.88 dB	2.207×10^7



F . 2. T pical performance curves ith 8 arbitar sources

e o d ed l o ex e e d l e e ed d o e . o e o e ed o e l o , N 000 l e w e e ed f o (b l o e) o e of e follow d b o (1) f o d b o , (2) b $\{-1, 1\}$ w e l ob b l , (3) e d b o e(2, 2), (4) l e x e w $p(x)$ $\frac{1}{2\sqrt{\pi}}e^{-(x-\sqrt{2}/2)^2} + \frac{1}{2\sqrt{\pi}}e^{-(x+\sqrt{2}/2)^2}$, ( ) l d b o , ( ) e bo e d b o , ( ) t-d b o w f eedo , ( ) t-d b o w f eedo 13. e o e we e o l ed d e x x w do l e e ed o o l x, o de o b e fl e e of ew e e .

e e ed ee d ffe e d e o we n 4, n d n 1 , e e el . I e l o , e w e l o d e I l o we e e ed w e e x e d . 100 e d o we e e fo ed fo e d e o , d ble 1 e e e ed e fo e . e e o l e ed b e e I e fe e e o l owe o (I), w f o of **R WA** e l o

$$I \quad (\mathbf{R}) \quad 10 \, \mathrm{lo}_{10} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{n} r_{ij}^2}{\mathrm{x}_{1 \le j \le n} \, r_{ij}^2} - 1 \right) \right\} . \qquad (2\ )$$

[illegible] . 2, [illegible] n .

[illegible] θ [illegible] 0.002 [illegible] θ 0.02 . [illegible]

5 Conclusions

[illegible]

Acknowledgements

[illegible] .D. [illegible] 200 0001042.

References

1. Cardoso, J.F.: Blind Signal Separation: Statistical Principles. Proceedings of the IEEE 86, 2009–2025 (1998)
2. Comon, P.: Independent Component Analysis – a New Concept? Signal Processing 36, 287–314 (1994)
3. Bell, A., Sejnowski, T.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation 7, 1129–1159 (1995)
4. Amari, S.I., Cichocki, A., Yang, H.: A New Learning Algorithm for Blind Separation of Sources. Advances in Neural Information Processing 8, 757–763 (1996)
5. Cardoso, J.F.: High-order contrasts for Independent Component Analysis. Neural Computation 11, 157–192 (1999)
6. Delfosse, N., Loubaton, P.: Adaptive Blind Separation of Independent Sources: a Deflation Approach. Signal Processing 45, 59–83 (1995)
7. Ge, F., Ma, J.: Analysis of the Kurtosis-Sum Objective Function for ICA. In: Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W. (eds.) ISNN 2008, Part I. LNCS, vol. 5263, pp. 579–588. Springer, Heidelberg (2008)
8. Moreau, E., Macchi, O.: High-order Contrast for Self-adaptive Source Separation. International Journal of Adaptive Control and Signal Processing 10, 19–46 (1996)
9. Hyvarinen, A.: Fast and Robust Fixed-point Algorithms for Independent Component Analysis. IEEE Trans. Neural Networks 10, 626–634 (1999)
10. Zarzoso, V., Nandi, A.K.: Blind Separation of Independent Sources for Virtually any Source Probability Density Function. IEEE Trans. Signal Processing 47, 2419–2432 (1999)
11. Zarzoso, V., Nandi, A.K., Herrmann, F., Millet-Roig, J.: Combined Estimation Scheme for Blind Source Separation with Arbitrary Source PDFs. Electronic Letters 37, 132–133 (2001)