Lecture 3: Fourth Order BSS Method

Jack Xin *

1 Introduction

The approach combines second and fourth order statistics to perform BSS of instantaneous mixtures. It applies for any number of receivers if they are as many as sources. It is a batch algorithm that uses non-Gaussianity and stationarity of source signals. It is linear algebra based direct method, reliable and robust, though large dimensions of sources may slow down the computation significantly. It is however limited to instantaneous mixtures.

1.1 Moments and Cumulants

For a complex random vector $X = (X^1, ..., X^p) \in \mathbb{C}^p$, its moments are:

$$m^i = E(X^i), \quad m^{ij} = E(X^i X^j), \quad m^{ijk} = E(X^i X^j X^k), \quad \cdots . \tag{1.1}$$

Consider the following moment generating function defined in $\xi = (\xi_1, ..., \xi_p) \in \mathbb{C}^p$:

$$M(\xi) = E(\exp(\xi \cdot X)). \tag{1.2}$$

Taylor expansion gives with the convention of summation over repeated indices:

$$M(\xi) = 1 + \xi_i m^i + \frac{1}{2!}\xi_i \xi_j m^{ij} + \frac{1}{3!}\xi_i \xi_j \xi_k m^{ijk} + \cdots \tag{1.3}$$

The cumulants c^i, c^{ij}, $\cdots$ are defined as coefficients in the expansion of the generating function:

$$K(\xi) = \log M(\xi) = \xi_i c^i + \frac{1}{2!}\xi_i \xi_j c^{ij} + \frac{1}{3!}\xi_i \xi_j \xi_k c^{ijk} + \cdots . \tag{1.4}$$

*Department of Mathematics, UC Irvine, Irvine, CA 92697, USA.

To find the connection between cumulants and moments, we write

$$M(\xi) = \exp(K(\xi)) = \exp\left(\xi_i c^i + \frac{1}{2!}\xi_i\xi_j c^{ij} + \frac{1}{3!}\xi_i\xi_j\xi_k c^{ijk} + \cdots\right)$$

and expand to find

$$M(\xi) = 1 + \xi_i c^i + \xi_i\xi_j(c^{ij} + c^i c^j)/2! + \cdots . \tag{1.5}$$

Combining terms and using symmetry, we express moments in terms of cumulants:

$$m^i = c^i \tag{1.6}$$
$$m^{ij} = c^{ij} + c^i c^j \tag{1.7}$$
$$m^{ijk} = c^{ijk} + (c^i c^{jk} + c^j c^{ik} + c^k c^{ij}) + c^i c^j c^k = c^{ijk} + c^i c^{jk}[3] + c^i c^j c^k \tag{1.8}$$
$$m^{ijkl} = c^{ijkl} + c^i c^{jkl}[4] + c^{ij} c^{kl}[3] + c^i c^j c^{kl}[6] + c^i c^j c^k c^l. \tag{1.9}$$

The number in the bracket represents the number of terms in the combination as the indices rotate, e.g. $i \to j$, $j \to k$ etc.

Taylor expanding $\log M(\xi)$ with $\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots$ gives the inverse formula:

$$c^i = m^i \tag{1.10}$$
$$c^{ij} = m^{ij} - m^i m^j \tag{1.11}$$
$$c^{ijk} = m^{ijk} - m^i m^{jk}[3] + 2m^i m^j m^k \tag{1.12}$$
$$c^{ijkl} = m^{ijkl} - m^i m^{jkl}[4] - m^{ij} m^{kl}[3] + 2m^i m^j m^{kl}[6] - 6m^i m^j m^k m^l. \tag{1.13}$$

Example 1: In the single variable ($p = 1$) case, $m_1 = E(X)$ (mean), $m_{11} = E(X^2)$ (variance), $m_{111} = E(X^3)$ (skewness), $m^{1111} = E(X^4)$ (kurtosis). The corresponding cumulants are:

$$\begin{aligned}
c^1 &= m^1 = E(X), \\
c^{11} &= m^{11} - (m^1)^2 = E(X^2) - (E(X))^2, \\
c^{111} &= m^{111} - 3m^1 m^{11} + 2(m^1)^3 = E(X^3) - 3E(X)E(X^2) + 2(E(X))^3, \\
c^{1111} &= m^{1111} - 4m^1 m^{111} - 3(m^{11})^2 + 12(m^1)^2 m^{11} - 6(m^1)^4 \\
&= E(X^4) - 4E(X)E(X^3) - 3(E(X^2)^2 + 12(E(X))^2 E(X^2) - 6(E(X))^4.
\end{aligned}$$

If $E(X) = 0$, the first three cumulants are same as moments: $c^1 = 0$, $c^{11} = E(X^2)$, $c^{111} = E(X^3)$; difference starts at 4th order:

$$c^{1111} = E(X^4) - 3(E(X^2))^2.$$

Example 2: In several variable ($p \geq 2$) case,

$$\begin{aligned} c^{12} &= E(X^1X^2) - E(X^1)E(X^2) \\ c^{123} &= E(X^1X^2X^3) - E(X^1)E(X^2X^3) - E(X^2)E(X^1X^3) \\ &\quad -E(X^3)E(X^1X^2) + 2E(X^1)E(X^2)E(X^3). \end{aligned}$$

If all random variables have mean zero, the 2nd and 3rd cross cumulants agree with cross correlations: $c^{12} = E(X^1X^2)$, $c^{123} = E(X^1X^2X^3)$.

Exercise 1: Write down the full expression of c^{1234} in terms of moments, and show that if all random variables have mean zero, then:

$$\begin{aligned} c^{1234} &= E(X^1X^2X^3X^4) - E(X^1X^2)E(X^3X^4) \\ &\quad -E(X^1X^3)E(X^2X^4) - E(X^2X^3)E(X^1X^4). \end{aligned}$$

Example 3: [**Multivariable normal distribution**] The PDF of multivariate Gaussian is

$$\frac{1}{(2\pi)^{p/2}|\sigma^{ij}|^{1/2}} \exp\left(-\frac{1}{2}(x^i - m^i)\sigma_{ij}(x^j - m^j)\right)$$

where $\sigma^{ij} = E((X^i - m^i)(X^j - m^j))$, $(\sigma_{ij}) = (\sigma^{ij})^{-1}$. By integration:

$$M(\xi) = \exp(\xi_i m^i + \frac{1}{2}\xi_i\xi_j\sigma^{ij}).$$

Hence $c^i = m^i$, $c^{ij} = \sigma^{ij} = E((X^i - m^i)(X^j - m^j))$. **All cumulants above order two vanish!** $\square$

Exercise 2: Show directly from the cumulant formulas that for a mean zero random variable X: (1) $c^{111} = 0$ if its PDF is symmetric around the origin; (2) $c^{1111} = 0$ if X is Gaussian.

If $c^{1111} > 0$, X is called super-Gaussian (leptokurtic or long-tailed); if $c^{1111} < 0$, X is called sub-Gaussian (platykurtic or short-tailed). The size of $|c^{1111}|$ is a measure of deviation from Gaussian, either peakedness (+) or flatness (-).

Exercise 3: Show that a generalized Gaussian random variable discussed in Lecture 1 is super-Gaussian if $r \in [1, 2)$, sub-Gaussian if $r \in (2, \infty)$.

1.2 Cumulants under Affine Transforms

Suppose we have an affine transformation

$$Y^r = a^r + a_i^r X^i, \quad (Y = a + AX) \tag{1.14}$$

Then the cumulants of Y are

$$c_Y^r = a^r + a_i^r c^i, \quad c_Y^{rs} = a_i^r a_j^s c^{ij}, \quad c_Y^{rst} = a_i^r a_j^s a_k^t c^{ijk}, \quad c_Y^{rstu} = a_i^r a_j^s a_k^t a_l^u c^{ijkl} \tag{1.15}$$

and so on. On the other hand, the moments of Y are much more complicated, unless $a^r = 0$.

To prove (1.15), we use the generating function and note that

$$\begin{aligned} K_Y(\xi) &= \log E\left(\exp\left(\xi \cdot (a + AX)\right)\right) = \log\big(\exp(\xi \cdot a) E[\exp(\xi_r a_i^r X^i)]\big) \\ &= \xi \cdot a + K_X(a_i^r \xi_r) \end{aligned}$$

For example, collecting 4th order cumulants gives:

$$c_Y^{rstl} \xi_r \xi_s \xi_t \xi_l = c^{ijkl}(a_i^{r_1} \xi_{r_1})(a_j^{r_2} \xi_{r_2})(a_k^{r_3} \xi_{r_3})(a_l^{r_4} \xi_{r_4})$$

therefore

$$c^{ijkl} a_i^{r_1} a_j^{r_2} a_k^{r_3} a_l^{r_4} = c_Y^{r_1 r_2 r_3 r_4}.$$

Suppose we know ε is a random variable with multivariable normal distribution (or constant plus white Gaussian noise, to be used later), and is independent of X. Define

$$Y = \varepsilon + AX. \tag{1.16}$$

Then, by the same argument, we find that

$$\begin{aligned} K_Y(\xi) &= \log E\left(\exp\left(\xi \cdot (\varepsilon + AX)\right)\right) = \log\left(E[\exp(\xi \cdot \varepsilon)] E[\exp(\xi \cdot AX)]\right) \\ &= \text{second order poly. of } \xi + K_X(A^\top \xi). \end{aligned}$$

So, the cumulants of 3rd or higher order are *not* affected by ε !

Proposition 1 *If $A \in \mathbb{C}^{n\times n}$ satisfies $A^H A = I$, namely $a_i^{j,*} a_k^j = \delta_{ik}$, (implies $AA^H = I$ and A is unitary) then*

$$\sum_{i,j,k,l} |c_Y^{ijkl}|^2 = \sum_{i,j,k,l} |c^{ijkl}|^2$$

Proof: By (1.15),

$$\begin{aligned}
&\sum_{r,s,t,u} |c_Y^{rstu}|^2 = \sum_{r,s,t,u} \left| \sum_{i,j,k,l} a_i^r a_j^s a_k^t a_l^u c^{ijkl} \right|^2 \\
&= \sum_{r,s,t,u} \sum_{i,j,k,l} \sum_{i_1,j_1,k_1,l_1} a_i^r a_j^s a_k^t a_l^u a_{i_1}^{r,*} a_{j_1}^{s,*} a_{k_1}^{t,*} a_{l_1}^{u,*} c^{ijkl} c^{i_1 j_1 k_1 l_1,*} \\
&= \sum_{i,j,k,l} \sum_{i_1,j_1,k_1,l_1} \left(\sum_{r,s,t,u} a_i^r a_j^s a_k^t a_l^u a_{i_1}^{r,*} a_{j_1}^{s,*} a_{k_1}^{t,*} a_{l_1}^{u,*} c^{ijkl} c^{i_1 j_1 k_1 l_1,*} \right) \\
&= \sum_{i,j,k,l} |c^{ijkl}|^2. \qquad \square
\end{aligned}$$

From the definition (1.4), it is clear that the value of ξ_k does not affect the value of c^{ij} if $k \neq i, j$. In other words, c^{ij} only depends on X^i and X^j. Moreover by (1.15), the dependence of c^{ij} on X^i and X^j is bilinear. Similar properties hold for other high order cumulants. Let us introduce the notation:

$$c^{ijkl} = Cum(X^i, X^j, X^k, X^l).$$

In the next section, we will consider in particular $Cum(X^i, X^{j,*}, X^k, X^{l,*})$ for complex random variables. A few simple and useful properties are:

(1) Symmetry: $Cum(X^i, X^j, X^k, X^l)$ is the same under permutation of X^i, X^j, X^k, X^l.

(2) Additivity:

$$Cum(X^1+Y^1, X^2, X^3, X^4) = Cum(X^1, X^2, X^3, X^4) + Cum(Y^1, X^2, X^3, X^4).$$

2 Cumulants, Linear Algebra and JADE

2.1 Preliminary

The general mixing model of n sources and m receivers is

$$y(t) = As(t) + \varepsilon(t), \tag{2.17}$$

where $y(t) = (y_1(t), ..., y_m(t))^\top \in \mathbb{C}^m$, $s(t) = (s_1(t), ..., s_n(t))^\top \in \mathbb{C}^n$, $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ and is independent of t, $\varepsilon(t)$ is the noise independent of signal. We know $\{y_i(t)\}$ and that s_i and s_j are independent for $i \neq j$. Our goal is to recover $s(t)$ under the assumption that $\varepsilon(t)$ is Gaussian white noise independent of s.

For any diagonal matrix Λ and any permutation matrix P, $\tilde{s}(t) = P\Lambda s(t)$ is also independent. We can write

$$y(t) = (A\Lambda^{-1}P^{-1})(P\Lambda s(t)) + \varepsilon(t) = \tilde{A}\tilde{s} + \varepsilon(t),$$

where $\tilde{A} = A\Lambda^{-1}P^{-1}$.

So without loss of generality, we may assume

$$R_s = E(ss^H) = (E(s_i(t)s_j^*(t)))_{i,j=1,...,n} = I_n. \tag{2.18}$$

Then $R_y = E(yy^H) = AA^H + E(\varepsilon(t)\varepsilon(t)^H) = AA^H + \sigma I_m \in \mathbb{R}^{m\times m}$. If $m \geq n$, σ can be estimated as the average of the $m-n$ smallest eigenvalues of R_y. Denoting $\mu_1, ..., \mu_n$ the n largest eigenvalues and $\boldsymbol{h}_1, ..., \boldsymbol{h}_n$ the corresponding eigenvectors of R_y. Define $W \in \mathbb{C}^{n\times m}$ by

$$W = [(\mu_1 - \sigma)^{-1/2}\boldsymbol{h}_1, ..., (\mu_n - \sigma)^{-1/2}\boldsymbol{h}_n]^H.$$

Then

$$I_n = W(R_y - \sigma I)W^H = W(AA^H)W^H. \tag{2.19}$$

Let $z = Wy$, where W is called a whitening matrix, then

$$R_z = WR_yW^H = I_n + \sigma WW^H = I_n + \sigma \begin{pmatrix} \frac{1}{\mu_1-\sigma} & & \\ & \ddots & \\ & & \frac{1}{\mu_n-\sigma} \end{pmatrix}. \tag{2.20}$$

Now, $z = WAs + W\varepsilon = Us + W\varepsilon$ with $U = WA \in \mathbb{C}^{n\times n}$. Because of (2.19),

$$UU^H = I_n,$$

namely, U is a unitary matrix. Now, the problem is reduced to finding U from z. Once we find U, since we know z and $z = Us + W\varepsilon$, s can be approximated by $U^H z$ which is the (noise corrupted) source signals and A can be approximated by $W^\# U$ where $W^\# \in \mathbb{C}^{m\times n}$ is the pseudoinverse of W, namely $WW^\# W = W$, $W^\# WW^\# = W^\#$, $(WW^\#)^* = WW^\#$ and $(W^\# W)^* = W^\# W$.

2.2 Identification of U

Suppose $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_n)$. For any $n \times n$ matrix $M = (m_{ij})$, define a cumulant matrix denoted by

$$N = Q_z(M) \quad \text{defined by} \quad n_{ij} = \sum_{k,l=1}^{n} Cum(z_i, z_j^*, z_k, z_l^*) m_{lk}. \tag{2.21}$$

We will make use of the property of the affine transformation (1.16). Plugging $z_i = u_{ij}s_j + \tilde{\varepsilon}_i$ into (2.21) with $(\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_n)$ being Gaussian random variable (noise), we get

$$\begin{aligned} n_{ij} &= \sum_{k,l} Cum(u_{i,k_1}s_{k_1}, u_{j,k_2}^* s_{k_2}^*, u_{k,k_3}s_{k_3}, u_{l,k_4}^* s_{k_4}) m_{lk} \\ &= \sum_{k,l}\sum_{p} Cum(s_p, s_p^*, s_p, s_p^*) u_{i,p}u_{j,p}^* u_{k,p}u_{l,p}^* m_{lk} \end{aligned} \tag{2.22}$$

which can be rewritten as

$$N = \sum_{p=1}^{n} (c_p\, \boldsymbol{u}_p^H M \boldsymbol{u}_p)\, \boldsymbol{u}_p \boldsymbol{u}_p^H, \quad \text{or} \quad N = U\Lambda_M U^H, \quad NU = U\Lambda_M, \tag{2.23}$$

where $c_p = Cum(s_p, s_p^*, s_p, s_p^*)$ and $\Lambda_M = \mathrm{diag}(c_1 \boldsymbol{u}_1^H M \boldsymbol{u}_1, ..., c_n \boldsymbol{u}_n^H M \boldsymbol{u}_n)$. In deriving the above equality, we have used the fact that the independence s_i's implies

$$Cum(s_i, s_j^*, s_k, s_l^*) = 0 \quad \text{when } i, j, k, l \text{ are not the same.} \tag{2.24}$$

Note that unlike moments, (2.24) holds without s_i's being mean zero. It follows from (1.13) as long as s_i's are independent. This is an advantage of cumulants.

Identity (2.23) says that any cumulant matrix is diagonalized by U. Or $NU = U\Lambda_M$, the column vectors of U are eigenvectors of any cumulant matrix. We shall seek a joint diagonalizer of cumulant matrices to identify U next.

2.3 Joint Diagonalization

Let $\mathcal{N} = \{N_r; r = 1, ..., s\}$ be a set of s matrices of size $n \times n$. A joint diagonalizer of the set $\mathcal{N}$ is defined as a unitary *minimizer* of

$$\sum_{r=1}^{s} |\mathrm{off}(V^H N_r V)|^2$$

where $|\mathrm{off}(A)|^2$ is the sum of squares of all off-diagonal elements of matrix A. Note that the Frobenious norm is preserved under Hermitian transformation ((2.28)), when V is unitary, $\|V^H N_r V\|_F^2 = \|N_r\|_F^2 = \sum_{i,j} |N_r(i,j)|^2$. Hence a joint diagonalizer V can also be characterized as the unitary *maximizer* of

$$\sum_{r=1}^{s} |\mathrm{diag}(V^H N_r V)|_{l_2}^2.$$

Proposition 2 *For any d-dimensional complex random vector v, there exist d^2 real number $\lambda_1, ..., \lambda_{d^2}$ and d^2 matrices $M_1, ..., M_{d^2}$, called eigenmatrices so that*

$$Q_v(M_r) = \lambda_r M_r, \qquad tr(M_r M_s^H) = \delta_{r,s}.$$

Proof: The relation $N = Q_v(M)$ is put into matrix vector form $\tilde{N} = \tilde{Q}\tilde{M}$. $\tilde{Q}_{ab} = Cum(v_i, v_j^*, v_k, v_l^*)$ with $a = i + (j-1)d$, $b = l + (k-1)d$. $\tilde{M} = (m_{11}, m_{21}, ..., m_{12}, m_{22}, m_{32}, ...)^\top$. Similarly for $\tilde{N}$.

$$\tilde{Q}_{ab}^* = Cum(v_i^*, v_j, v_k^*, v_l) = Cum(v_l, v_k^*, v_j, v_i^*) = \tilde{Q}_{ba}.$$

So $\tilde{Q}$ is Hermitian. Hence it has d^2 real eigenvalues and corresponding eigen vectors which are orthogonal to each other. □

2.4 Algorithm

JADE (Joint Approximate Diagonalization of Eigen-matrices)

- Form the sample covariance R_y and compute a whitening matrix W.
- Form the sample 4th order cumulants Q_z of the whitened process $z = Wy$. Compute the n most significant eigenpairs $\{\lambda_r, M_r; r = 1, ..., n\}$.
- Jointly diagonalize the set $\mathcal{N} = \{\lambda_r M_r; r = 1, ..., n\}$ by a unitary matrix U.
- An estimate of A is $A = W^{\#}U$ where $W^{\#} \in \mathbb{C}^{m\times n}$ is the pseudoinverse of W.

2.5 Jacobi Method for Symmetric Eigenvalue Problem

Suppose A is Hermitian, and we want to find a unitary J so that $B = J^H AJ$ is more "diagonal" than A. The idea of Jacobi method is to systematically reduce the quantity

$$\mathrm{off}(A) = \sqrt{\sum_{i=1}^{n}\sum_{j\neq i} |a_{ij}|^2}. \tag{2.25}$$

Let $c \in \mathbb{R}^+$, $s \in \mathbb{C}$ and $|s|^2 + |c|^2 = 1$,

$$J(p,q,c,s) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s^* & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \tag{2.26}$$

where the c's and s's are at p, q rows and columns. We want to choose c and s so that

$$\begin{pmatrix} b_{pp} & 0 \\ 0 & b_{qq} \end{pmatrix} = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}^H \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \tag{2.27}$$

is diagonal. At each step, a symmetric pair of zeros is introduced into the matrix, though previous zeros may be destroyed. However, the "energy" of the off diagonal elements is decreasing with each step. First the Frobenius norm is conserved:

$$\|B\|_F^2 = \sum_i \sum_j |b_{ij}|^2 = \text{Tr}(B^H B) = \text{Tr}(J^H A^H J J^H A J) = \text{Tr}(A^H A) = \|A\|_F^2. \tag{2.28}$$

Because (2.27) is a Hermitian transformation, we have

$$|a_{pp}|^2 + |a_{qq}|^2 + 2|a_{pq}|^2 = |b_{pp}|^2 + |b_{qq}|^2.$$

Note that diagonal entries of A remain unchanged except (p,p) and (q,q) entries. So,

$$\begin{aligned}
\text{off}(B)^2 &= \|B\|_F^2 - \sum_i |b_{ii}|^2 \\
&= \|A\|_F^2 - \sum_i |a_{ii}|^2 + (|a_{pp}|^2 + |a_{qq}|^2 - |b_{pp}|^2 - |b_{qq}|^2) \\
&= \text{off}(A)^2 - 2|a_{pq}|^2.
\end{aligned}$$

It is in this sense that A moves closer to a diagonal form with each Jacobi step.

2.6 Joint Diagonalizer

For each choice of $p \neq q$, find $c \in \mathbb{R}^+$ and $s \in \mathbb{C}$ to minimize the following objective function

$$\sum_{k=1}^{K} \left[\text{off}(J(p,q,c,s) N_k J^H(p,q,c,s))\right]^2 \tag{2.29}$$

with J defined in (2.26).

Theorem 2.1 *For any set $\mathcal{N} = \{N_k; k = 1, ..., K\}$ of $n \times n$ matrices and a given pair (p,q) of indices, define a 3×3 real symmetric matrix G by*

$$G = Real\left(\sum_{k=1}^{K} h^H(N_k) h(N_k)\right)$$

where $h(N) = (a_{pp} - a_{qq}, a_{pq} + a_{qp}, j(a_{qp} - a_{pq})) \in \mathbb{C}^{1\times 3}$*. Then the objective function*

$$f(c, s) = \sum_{k=1}^{K} \left[\text{off}(J(p, q, c, s)N_k J^H(p, q, c, s))\right]^2$$

is minimized at

$$c = \sqrt{\frac{x + r}{2r}}, \quad s = \frac{y - jz}{\sqrt{2r(x + r)}}, \quad r = \sqrt{x^2 + y^2 + z^2}$$

where $(x, y, z)^\top$ *is an eigenvector associated with the largest eigenvalue of* G*.*

Consider 2×2 matrices ($n = 2$):

$$N_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix}$$

A complex rotation matrix is:

$$J = \begin{bmatrix} \cos\theta & -e^{j\varphi}\sin\theta \\ e^{-j\varphi}\sin\theta & \cos\theta \end{bmatrix}$$

Let a'_k, b'_k, c'_k, d'_k be the four elements of $J^H N_k J$, the optimization is to find angles (θ, φ) to maximize $\sum_k |a'_k|^2 + |d'_k|^2$. Notice that:

$$2(|a'_k|^2 + |d'_k|^2) = |a'_k - d'_k|^2 + |a'_k + d'_k|^2,$$

and trace $a'_k + d'_k$ is invariant, maximization is on the objective:

$$Q = \sum_k |a'_k - d'_k|^2 \tag{2.30}$$

Define vectors:

$$\begin{aligned} u &= [a'_1 - d'_1, \cdots, a'_K - d'_K]^T, \\ v &= [\cos 2\theta, \sin 2\theta\cos\varphi, \sin 2\theta\sin\varphi]^T, \\ g_k &= [a_k - d_k, b_k + c_k, j(c_k - b_k)]^T \in \mathbb{C}^{3\times 1} \\ G &= [g_1, \cdots, g_K]^T \end{aligned}$$

Then the transform:

$$a'_k - d'_k = (a_k - d_k)\cos 2\theta + (b_k + c_k)\sin 2\theta\cos\varphi + j(c_k - b_k)\sin 2\theta\sin\varphi,$$

is put in matrix-vector form $u = Gv$ and:

$$Q = u^H u = v^T G^H G v = v^T \mathrm{real}\,(G^H G) v. \tag{2.31}$$

Direct calculation shows that $v^T v = 1$. Maximizing Q is same as finding an eigenvector corresponding to the largest eigenvalue of the real symmetric 3×3 matrix real$(G^H G)$. This is also known as the Rayleigh quotient problem for symmetric matrices [3].

References

[1] J-F. Cardoso and A. Souloumiac, *Blind beamforming for non Gaussian signals*, IEE-Proceedings-F (Radar and Signal Processing) , vol 140, no 6, (1993) 362–370.

[2] J-F. Cardoso and A. Souloumiac, *Jacobi angles for simultaneous diagonalization*, SIAM Journal on Matrix Analysis and Applications, Vol 17, (1996) 161–164.

[3] G. H. Golub and C. F. Van Loan, “Matrix Computations”, 3rd edition, Johns Hopkins University Press, (1996).

[4] P. McCullagh, “Tensor Methods in Statistics”, Chapman and Hall, (1987).