PARAFAC BASED BLIND ESTIMATION OF MIMO SYSTEMS WITH POSSIBLY MORE INPUTS THAN OUTPUTS

Yuanning Yu, Athina P. Petropulu

Electrical & Computer Eng. Dept., Drexel University, Philadelphia, PA 19104
{yuanning, athina}@cbis.ece.drexel.edu

ABSTRACT

We consider the problem of frequency domain identification of a convolutive multiple-input multiple-output (MIMO) system driven by white, mutually independent unobservable inputs. In particular, we improve upon a method recently proposed in [1] that uses PARAFAC decomposition of a tensor that is formed based on third-order statistics of the system output. The approach of [1] utilizes only one slice of the output tensor to recover one row of the system response matrix. We here propose an approach that fully exploits the information in the output tensor. As a result, the proposed method not only achieves lower error values but also becomes applicable to MIMO systems with more inputs than outputs. We also extend the method to employ fourth-order statistics of the system output, thus making the proposed method applicable to a communications scenario.

2. PROBLEM FORMULATION

Let us consider a \( N_f \)-input \( N_o \)-output LTI system.

\[
x(k) = \sum_{l=0}^{L-1} h(l)s(k-l) + n(k)
\]

where \( s(n) \) is a \( N_f \) by 1 source vector; \( x(n) \) is a \( N_o \) by 1 observation vector; \( n(n) \) is the observation noise; \( h(l) \) is the FIR MIMO system impulse response matrix whose \((i,j)\) element is denoted by \( h_{ij}(n) \) with \( 1 \leq i \leq N_f, 1 \leq j \leq N_o \); \( L \) is the length of the longest \( h_{ij}(n) \).

Let \( \mathbf{H}(k) \) be an \( N_o \times N_f \) matrix defined as the \( N \)-point Discrete Fourier Transform of \( h(n) \), i.e.,

\[
\mathbf{H}(k) = \sum_{n=0}^{N-1} h(n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0, \ldots, N - 1 \text{ where } N > L.
\]

Our goal is to obtain an estimate of \( \mathbf{H}(k) \) within a column permutation ambiguity \( \mathbf{P} \), a constant diagonal scalar ambiguity \( \mathbf{A} \), and a linear phase term \( e^{j\frac{2\pi}{N}kM} \) (\( \mathbf{M} \) is a diagonal matrix with integer elements), i.e.,

\[
\hat{\mathbf{H}}(k) = \mathbf{H}(k)\mathbf{P}\mathbf{A}e^{j\frac{2\pi}{N}kM}
\]

Notation/Definitions: Let \( \mathbb{C}^n \) denote the \( n \)-way tensor; \( \mathbb{C} \) denotes a matrix; \( k_{\mathbf{A}} \) denotes the \( k \)-rank of matrix \( \mathbf{A} \). The matrix \( \mathbf{A} \) of size \( I \times J \) has \( k \)-rank \( k_{\mathbf{A}} = I \) if every \( I \) columns of \( \mathbf{A} \) are linearly independent, but either \( I = J \), or there exist a collection of \( I \) linearly dependent columns in \( \mathbf{A} \).

3. PROPOSED METHODS

3.1. Third-order statistics based approach

Assumptions: (A1) Each \( s_i(\cdot) \) is a zero mean, non-symmetrical distributed, independent identically distributed (i.i.d.), stationary process with nonzero skewness. The \( s_i's \) are mutually independent. (A2) \( n_i(\cdot), i = 1, \ldots, N_o \) are zero mean Gaussian stationary random processes with variance \( \sigma_n^2 \), mutually independent and independent of the inputs. (A3) The \( k \)-rank of \( \mathbf{H}(k) \) satisfies: \( 3k_H \geq 2N_i + 2 \) for every \( k \).

Assumption (A1) and (A2) are common in HOS based methods that involve third-order statistics. To get a sense of (A3), if the elements of the channel matrix are drawn independently from an absolutely continuous distribution, the \( \mathbf{H}(k) \) has both rank and \( k \)-rank equal to \( \min(N_f, N_o) \) with probability one [9]. For such channel, assumption (A3) is satisfied if \( 3\min(N_f, N_o) \geq 2N_i + 2 \). Under the
assumption $N_o \geq N_i$ that was made in [1], (A3) is easily satisfied as long as $N_o \geq 2$.

The $N \times N$ discrete-frequency cross-bispectrum of the outputs $x_i(k), x_i^*(k), x_j(k)$ is the two-dimensional Discrete Fourier transform of the third order cross-cumulant[10], and equals:

$$C_{ij}^3(k_1, k_2) = \sum_{p=1}^{N} \gamma_{ip}^3 H_{ip}(-k_1 - k_2) H_{ip}^*(-k_1) H_{jp}(k_2)$$  (3)

where $k_1, k_2 = 0, ..., N - 1$. For $k_1 = -m + r \delta, k_2 = \delta$, where $m, r, \delta$ integers, $C_{ij}^3((-m + r \delta, \delta)$ can be viewed as the $(l, i, j)-$th element of tensor $C^3((-m + r \delta, \delta) (N_o \times N_o \times N_o)$. Let us define:

$$A_r \triangleq \hat{H}(m - r \delta - \delta) \quad B_r \triangleq \hat{H}^*(m - r \delta) \Gamma^4 \quad C_r \triangleq \hat{H}(\delta)$$  (4)

By fixing the index $r$ we can rewrite (3) in matrix form as:

$$C_r^3((-m + r \delta, \delta) = B_r \Lambda r (A_r) C_r^T$$  (5)

In [1] it was shown that the $l-th$ row of $A_r$ can be recovered from $C^3((-m + r \delta, \delta)$ in an iterative way. Here we propose the following approach for recovering the entire matrix $A_r$.

Let us slice the tensor $C_r^3((-m + r \delta, \delta)$ differently by fixing the index $i$ as:

$$C_r^3((-m + r \delta, \delta) = C_i \Lambda r (B_i) A_i^T$$  (6)

we can stack the matrices $C_r^3((-m + r \delta, \delta)$ for $i = 1, ..., N_o$ to form a tall matrix $U_A(r)$. It holds:

$$U_A(r) = (B_1 \otimes C_1) \Lambda 1 (A_1) \otimes C_1 \otimes \hat{H}(\delta)$$  (7)

where $\otimes$ is the Khatri-Rao (column-wise Kronecker) Product.

**Lemma 1:** The Khatri-Rao product $(B_r \otimes C_r)$ has a left inverse for all $r = 0, ..., N - 1$ under (A3).

**Proof:** It has been proved in [9] that $(B_r \otimes C_r)$ has full column rank $N_i$ once $k_B + k_C \geq N_i + 1$. By noting that $k_A = k_B \leq N_i$ and $k_B + k_C \geq 2N_i + 2$ (A3), we get $k_B + k_C \geq N_i + 1$ and thus $(B_r \otimes C_r)$ has left inverse.

We can then solve (7) for $A_r$ as:

$$A_r = ((B_r \otimes C_r)^{-1} U_A(r))^T$$  (8)

where $(B_r \otimes C_r)^{-1}$ is the left pseudo inverse of $(B_r \otimes C_r)$.

By using the above avoiding trick, we avoid the need for the existence of the inverse of $B_r$ that was required in [1]. As it will be seen later, this makes the method applicable to systems with $N_i > N_o$ once (A3) is satisfied.

To find the system response we need to compute $A_r$ for $r = 0, ..., N - 1$, we apply PARAFAC decomposition once to the tensor $C_r^3((-m, \delta)$, and then use an iterative approach to find all the $A_r$‘s. Under assumption (A3), PARAFAC decomposition of $C_r^3((-m, \delta)$, yields [2]:

$$\hat{A}_0 \triangleq \hat{H}(m - (r - 1) \delta) \Gamma^4 \quad \hat{B}_0 \triangleq \hat{H}^*(\delta) \Gamma^4 \quad \hat{C}_0 \triangleq \hat{H}(\delta) \Gamma^4$$  (9)

where $\hat{P}$ is a permutation matrix and $\hat{A}_0$ is a complex diagonal matrices that satisfy: $A_2 \hat{A}_1; A_3 = I$. Subsequently, by noting that $B_r = A_r \Gamma^4$, let us define an iteration for $r = 1, ..., N$ as:

$$\hat{A}(r) = ((\hat{A}^*(r - 1) \otimes \hat{C}_0)^{-1} U_A(r))^T$$  (10)

It can be shown that (see Appendix I):

$$\hat{A}(r) = \hat{H}(m - r \delta - \delta) \Lambda r (A_r) \otimes C_r \otimes H(\delta)$$  (11)

As it will be seen later, the phase ambiguity in eq. (11) can be solved in the same manner as in [1], which is included here for the readers’ convenience. Define:

$$\hat{A}(r)[\hat{A}^{-1}(N) \hat{A}(0)]^{r/N}$$

where $k$ is an integer. Applying an $N/2$ point IDFT on the even samples of $\hat{H}(k)$ of (11), we get an unsampled by $\delta$ version of $h$, circularly shifted by $k$ and modulated due to the $m - \delta$ term in (11). Once we extract the $L$-samples long segment (modulo $N$) with the maximum energy based on its absolute value, we can cancel the modulating factor, and compute the amount of circular shift.

The steps consisting of applying equations (9),(10) and (11) define the proposed method, which we will refer to as the Improved Single PARAFAC decomposition (ISPFD) method.

3.2. Extension to fourth-order statistics

Assumptions: In addition to (A2) we assume: (A4) Each $s_i(\cdot)$ is a zero mean, i.d., stationary process with nonzero kurtosis. The $s_i$’s are mutually independent. (A5) The $r$-rank of $\hat{H}(k)$ satisfies: $4k_H \geq 2N_i + 3$ for every $k$.

Assumption (A4) requires that the fourth-order cumulants of the inputs are not identically zero. Assumption (A4), unlike assumption (A1), is satisfied by most communication signals. For a channel matrix with independent taps, assumption (A5) is satisfied if $4\min(N_i, N_o) \geq 2N_i + 3$. We can see that, for $N_o \geq N_i, (A3)$ and (A6) are satisfied as long as $N_i \geq 2$.

Based on (A2) and (A4) the 4-th order discretized Trispectrum $C_{ijkm}^4(k_1, k_2, k_3, k_4)$ defined as the three dimensional DFT of the fourth order cross-cumulants, and equals [10]:

$$\sum_{p=1}^{N_i} \gamma_{ip}^4 H_{ip}(-k_3 - (k_1 + k_2)) H_{ip}^*(-k_1) H_{ip}(k_2) H_{mp}^*(-k_3)$$  (12)

Let $k_1 = -\delta, k_2 = 0, k_3 = -m + r \delta$, and $C^4((-\delta, 0, -m + r \delta)$ denote the four-way tensor constructed by elements $C_{ijkm}^4((-\delta, 0, -m + r \delta)$. Let us define:

$$A_r \triangleq \hat{H}(m - (r - 1) \delta) \Gamma^4 \quad B_r \triangleq \hat{H}^*(\delta) \quad C_r \triangleq \hat{H}(0) \quad D_r \triangleq \hat{H}^*(\delta)$$  (13)

For fixed $i, j$, let us stack the $N_o^2$ matrices $(N_o \times N_o) C_r^4((-\delta, 0, -m + r \delta), i, j = 1, ..., N_o)$ to form a $N_o^2 \times N_o$ tall matrix $U_{B_i}(r)$. It holds [2]:

$$U_{B_i}(r) = (A_r \otimes B_r \otimes C_r) D_r^T$$  (13)

**Lemma 2:** The Khatri-Rao Product $(A_r \otimes B_r \otimes C_r)$ has a left inverse for all $r = 0, ..., N - 1$ under (A5).
Proof: It is easy to shown that $k_B + k_C \geq N_i + 1$ from (A5), thus $(B_i \odot C_i)$ has full column rank, also full k-rank $N_i$. By noting that $k_A = k_H \geq 1$, we get $k_A + k_B(\beta_{BC}) \geq N_i + 1$, and thus $(A_o \odot B_o \odot C_o)$ has full k-rank $N_i$. B.

Under assumption (A5), the tensor $C^4(-\delta, 0, -m)$ can be decomposed by the PARAFAC algorithm into [2]:

$$\hat{A}_0 = A_0 \Delta P_{10}, \quad \hat{B}_0 = B_0 \Delta P_{20},$$

$$\hat{C}_0 = C_0 \Delta P_{30}, \quad \hat{D}_0 = D_0 \Delta P_{40} \quad (18)$$

where $A_{10} A_{20} A_{30} A_{40} = I$.

For $r = 0, 1, ..., N - 1$, let us define the iteration:

$$\hat{D}^4(r) = ((\hat{D}^4(r - 1) \odot \hat{B}_0 \odot \hat{C}_0)^{-1} U_B^0(r))^T$$

$$\hat{D}^4(0) = \hat{D}_0 \quad (19)$$

Similar to the third-order case, it holds:

$$\hat{D}^4(r) = H(m - r \delta)P_{(r)} \odot e^{i(-\Phi_1 + (r+1) \Phi_2)} \quad (20)$$

where $\Phi_1, \Phi_2$ are constant diagonal matrices, and $S_1, S_0$ are diagonal matrices with positive elements.

Similar to (11), equation (20) provides the even- or odd-indexed samples of the system frequency response within a phase and constant permutation and scalar ambiguities. We need to take $\delta, N$ comrime, and $m$ as integer multiple of $\delta$ to facilitate solving the trivial phase ambiguity. The phase ambiguity can be handled in exactly the same manner as in (11).

3.3. MIMO systems with more inputs than outputs

Even if $N_i > N_o$, it is still possible to apply the proposed ISPD method for certain $N_i$ and $N_o$ combinations as long as the relevant assumptions are satisfied, i.e., (A3) for the $3^{rd}$ order case and (A5) for the $4^{th}$ order case.

For the $3^{rd}$ order case, if the channel matrix elements are random and independent, then the $k$-rank of $H(k)$ equals $k_H = \min\{N_o, N_i\} = N_i$ and assumption (A4) becomes:

$$3k_H = 3N_o \geq 2N_i + 2, \quad (21)$$

Thus, the third-order based ISPD method can be applied to system in which $N_o = 4, N_i = 5$, or $N_o = 6, N_i = 8$.

Similar arguments apply to the $4^{th}$ order case. MIMO system with $N_o = 3, N_i = 4$ or $N_o = 4, N_i = 6$ can be solved by the fourth-order based ISPD method.

The main constraint to get the system estimate is the assumptions to guarantee the uniqueness of the initial PARAFAC decomposition, i.e., (A3) for $3^{rd}$ order case and (A5) for $4^{th}$ order case.

Next we show for the $3^{rd}$ order case how to combine two tensors $C^3(-m, \delta)$ and $C^3(-m + \delta, \delta)$ to construct a $2N_o \times N_o \times N_o$ tensor, which can give us unique initial PARAFAC decomposition under the more relaxed condition $2N_o \geq N_i + 2$ (instead of (A3)).

Along the lines of the definition of $U_B(r)$ (7), let us define $U_B(r)$ by fixing index $j$ and stacking the matrices corresponding to $j = 1, ..., N_o$. Also, let us define $U_C(r)$ by fixing index $l$ and stacking the matrices corresponding to $l = 1, ..., N_o$.

Similar to (8), and noting that $B_1 = A_o$ and $C_0 = C_1$, we can get two $N_o^2 \times N_o$ equations for $B_0$ and $A_1$ as:

$$B_0 = ((C_0 \odot A_0)^{-1} U_B(0))^T \quad (22)$$

$$A_1 = ((A_o \odot C_0)^{-1} U_A(1))^T \quad (23)$$

Noting that $B_1 = A_o$ and $C_0 = C_1$, we can also get two $2N_o^2 \times N_o$ equations for $C_0$ and $A_0$ as:

$$C_0 = ((A_o \odot B_0)^{-1} (U_C(0))^T \quad (24)$$

$$A_0 = ((B_0 \odot C_0)^{-1} (U_A(0))^T \quad (25)$$

Now we have four equations (22-25) for the four unknowns: $A_o, B_0, C_0$ and $A_1$. we can solve them in an Alternative Least Square (ALS) [9] manner as: every step update the four unknowns according to the four equations, and compute the fitting of the decomposed matrices and the tensors, stop when the convergence is slow. This is equivalent to solving a PARAFAC decomposition problem for a 3-way tensor of dimensions $2N_o \times N_o \times N_o$. As long as $2N_o \geq N_i$, we can reduce the assumption (A3) to $2N_o \geq N_i + 2$.

Thus, by combining two tensors, the third-order based ISPD can be applied to more systems with more inputs than outputs, i.e., $N_o = 3, N_i = 4$ or $N_o = 4, N_i = 6$.

The same idea can be applied to the $4^{th}$ order case as well. By combining two consecutive tensors, we can have three $2N_o^2 \times N_o$ updating equations for $B_0, C_0$ and $D_0$ (which is only different with $A_o$ in a diagonal matrix $T^*$), and two $N_o^3 \times N_o$ updating equations for $A_0$ and $D_1$. This is equal to solving a PARAFAC decomposition for a 4-way tensor of dimension $2N_o \times N_o \times N_o \times N_o$. As long as $2N_o \geq N_i$, we can reduce the assumption (A5) to $3N_o \geq N_i + 3$, which can be satisfied with $N_o = 2, N_i = 3$ or $N_o = 3, N_i = 5$.

3.4. Simulations

We next demonstrate the performance of the proposed approach.

We still consider the $2 \times 2$ bandpass MIMO channels as in [1]: $h_{ij}(n) = r_1 \cdot c(0.25(n - 10)) + r_2 \cdot c(0.25(n - 6)) + r_3 \cdot c(0.25(n - 8))$ where $c(n)$ is a sinc function with delay $n$ and the $r_i$’s are zero-mean Gaussian random variables. By varying the $r_i$’s we can generate multiple bandpass channels. We tested the performance for the proposed $3^{rd}$ order ISPD method for $50 \times 2 \times 2$ channels of length $L = 6$, generated by varying the $r_i$’s randomly for each channel. We also provide comparison results between the proposed method against the SPD method of [1], the methods of [4], and [7]. The method of [7] is a deflation-type approach, where the input sequences are extracted and removed one by one and then the system is estimated based on the system output and the estimated input.

For each channel we performed 50 Monte Carlo runs. We took $L_e = 10, T = 8000, SNR = 20dB$ in all methods. In all methods the output cross-cumulants were estimated using the same parameters. We use $N = 128$, $m = \delta$ for the SPD and ISPD methods, and $\delta$ is chosen from the peak of the power-spectra. Table 1 shows the mean, variance and the
cumulative distribution of the ONMSE for both the proposed method and also the lower bound. The ONMSE of the 4th order ISPD is less than 0.08 with probability higher than 0.9.

In Table 2, we showed the performance for the proposed 3rd order ISPD method for 50 \( N_o = 3, N_t = 4 \) channels of length \( L = 6 \). For each channel we performed 50 Monte Carlo runs. We took \( L_o = 10, T = 8000, N = 128, S N R = 20 d B, m = \delta, \) and \( \delta \) is chosen from the peak of the power-spectra.

**Table 1.** ONMSE Dist. for 4th-order ISPD for 50 2 x 2 channels

<table>
<thead>
<tr>
<th>ONMSE</th>
<th>mean</th>
<th>std</th>
<th>50%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISPD 4th</td>
<td>0.0354</td>
<td>0.0012</td>
<td>0.019</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>lower bound</td>
<td>0.0016</td>
<td>0.4 \times 10^{-6}</td>
<td>0.0016</td>
<td>0.0018</td>
<td>0.002</td>
</tr>
</tbody>
</table>

**Table 2.** ONMSE Dist. for 3rd-order ISPD for 50 3 x 4 channels

<table>
<thead>
<tr>
<th>ONMSE</th>
<th>mean</th>
<th>std</th>
<th>50%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISPD 3rd</td>
<td>0.214</td>
<td>0.009</td>
<td>0.18</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td>lower bound</td>
<td>0.005</td>
<td>0.6 \times 10^{-6}</td>
<td>0.005</td>
<td>0.0055</td>
<td>0.006</td>
</tr>
</tbody>
</table>

4. CONCLUSION

We presented a robust iterative scheme (ISPD) under the frequency domain framework of [1] for the identification of a multiple-input multiple-output (MIMO) system driven by white, mutually independent unobservable inputs. The proposed ISPD method, can achieve lower ONMSE value compared with the SPD method of [1], the methods of [4] and [7]. And the ISPD can be applied to certain systems with more inputs than outputs. Also we extended the proposed ISPD method to the 4th order case for communication scenario.

Appendix I

\[
A(1) = ((A^*(0) \odot \hat{C}_0)^{-1}U_A(1))^T
= ((A_A^*P_0A_{10}^* \odot C_0P_0A_{30})^{-1}U_A(1))^T
= ((B_1P_0^*(T_0)A_{10}^* \odot C_1P_0A_{30})^{-1}U_A(1))^T
= ((B_1 \odot C_1P_0^*(T_0)A_{10}^* \odot C_1P_0A_{30})^{-1}U_A(1))^T
= ((B_1 \odot C_1P_0^*(T_0)A_{30})^{-1}U_A(1))^T
= A_1P_0^*(A_{20})^{-1}U_A(1)
\]

Similarly, based on \( C_i^j(\cdot - m + r \delta, \delta) \) we can get:

\[
A(r) = H(m - r \delta - \delta)P_0A_1e^{j(\Phi_{10} + r(\Phi_{13} - \Phi_{30}))}, \quad (26)
\]

where

\[
K_r = \begin{cases} 
|A_{20}|G_0^3 & \text{for } r \text{ odd} \\
|A_{10}| & \text{for } r \text{ even}
\end{cases}
\]

A. REFERENCES


Fig. 1. Comparison of Cumulative distribution of the ONMSEs of 50 channels with 2 sources and 2 sensors (3rd order)