# Memory-Reduced DOA Estimators While Using Cyclic-Symmetrical Array

Ofer Schwartz
*Sound Business Unit CEVA Inc.*
Herzelia-Pituah, Israel
Ofer.Schwartz@ceva-dsp.com

***Abstract*—Home-assistant devices usually use microphone arrays for voice capture and enhancement via beamforming. A key parameter for beamforming is the direction of arrival (DOA) of the relevant speaker. The main disadvantage of DOA estimation algorithms is high memory usage due to the need for a steering vector to be prepared and saved for each candidate DOA. Frequently, a cyclic-symmetrical array is used within these devices. In this paper, a technique for reducing the memory usage for such arrays is shown. Moreover, a technique for blind estimation of the fundamental angle is described. Finally, as an experimental example, a cyclic-symmetrical array with 17 microphones is used for DOA estimation using the generalized cross correlation (GCC) technique and the proposed memory reduction technique is demonstrated.**

## I. Introduction

Online speaker localization is required in many applications, including beamforming, camera steering, multi-speaker separation, navigation, and target acquisition. In the audio-processing community, baseline DOA estimators are based on the GCC [1] or the multiple signals classification (MUSIC) algorithm [2]. Algorithms for estimating multiple DOAs of concurrent speakers were derived based on maximum likelihood (ML) criterion [3], the sparsity assumption of the speech nature [4]–[9], and more. All of these algorithms rely on the array geometry, and in particular on the phase difference between the microphones w.r.t. each possible DOA. The phase difference set for each DOA is fixed in time and therefore may be prepared and saved prior to the algorithm activation to spare online calculations. For end-point devices, the memory usage capabilities may not suffice for such memory demands.

In this paper, a technique for memory reduction while using a cyclic-symmetrical array is presented. A cyclic-symmetrical array can be rotated around the angular coordinates with some fundamental angle $\theta_F$, and results in the same microphone positions. It is later shown that, for these arrays, the phase difference set for two DOAs (with a difference of the fundamental angle) are equal up to the inner cyclic rotation of the order of the phase-difference set. Thus, the phase-difference needs to be pre-prepared and saved only for DOAs up to $\theta_F$, and the other sets can be obtained within the algorithm activation by only changing the order of the phases. As a result, memory usage is reduced by $\frac{360}{\theta_F}$.

In the following sections, the DOA estimation problem and a baseline GCC technique is described. Finally, the proposed memory-reduction technique is elaborated.

## II. Signal Model

Consider $N$ microphone observations consisting of speech and additive noise. The speaker beams from DOA $\theta_S$, which can be chosen from a set of predefined DOA candidates with the required resolution $\theta_S \in [0^o : \beta^o : 360^o]$ and overall $J = \frac{360}{\beta}$ DOA candidates. The $i$-th microphone observation can then be expressed in the short-time Fourier transform (STFT) domain as:

$$Y_i(m,k) = X_i(m,k) + V_i(m,k), \tag{1}$$

where $Y_i(m,k)$ denotes the $i$-th microphone observation with time-index $m$ and frequency index $k$, $X_i(m,k)$ denotes the speech as observed at the $i$-th microphone, and $V_i(m,k)$ denotes the ambient noise. Here $X_i(m,k)$ is modeled as a multiplication of the speech $X_1(m,k)$ (as received by the first microphone that was arbitrarily chosen as the reference microphone) and the relative steering response of the $i$-th microphone, i.e.:

$$X_i(m,k) = G_i(\theta_S,k) X_{j,1}(m,k). \tag{2}$$

Neglecting the reverberation phenomena, the steering response $G_i(\theta_S,k)$ is a pure phase depending on the time difference of arrival between the $i$-th microphone and the reference microphone:

$$G_i(\theta_S,k) = \exp\left(-\iota \frac{2\pi k}{K} \frac{\tau_i(\theta_S)}{T_\text{s}}\right), \tag{3}$$

where $\tau_i(\theta_S)$ is the time difference of arrival (TDOA) between the $i$-th microphone and reference microphone of the acoustic wave that comes from DOA $\theta_S$, $T_\text{s}$ is the sampling time, and $K$ is the number of frequency bins. Considering only the horizontal plane, and given the two-dimensional positions of the microphones, the TDOA $\tau_i(\theta)$ for each DOA $\theta$ is given by:

$$\tau_i(\theta) = \frac{1}{c} \cdot [\cos(\theta) \quad \sin(\theta)] (\mathbf{p}_i - \mathbf{p}_1), \tag{4}$$

where $c$ is the sound velocity and $\mathbf{p}_i = [x_i, y_i]^T$ is the horizontal position of microphone $i$. The $N$ microphone signals can be concatenated in a vector form:

$$\mathbf{y}(m,k) = \mathbf{g}(\theta_S,k) X_{j,1}(m,k) + \mathbf{v}(m,k) \tag{5}$$

ISBN: 978-1-6654-6798-8 

where:

$$\mathbf{y}(m,k) = \left[\begin{array}{ccc} Y_1(m,k) & \dots & Y_N(m,k) \end{array}\right]^{\mathrm{T}},$$
$$\mathbf{g}(\theta_S,k) = \left[\begin{array}{ccc} G_1(\theta_S,k) & \dots & G_N(\theta_S,k) \end{array}\right]^{\mathrm{T}},$$
$$\mathbf{x}(m,k) = \left[\begin{array}{ccc} X_{1,1}(m,k) & \dots & X_{J,1}(m,k) \end{array}\right]^{\mathrm{T}},$$
$$\mathbf{v}(m,k) = \left[\begin{array}{ccc} V_1(m,k) & \dots & V_N(m,k) \end{array}\right]^{\mathrm{T}}.$$

In the next section, a baseline DOA estimator is described.

## III. DOA Estimation

In the sequel, the indexes $m$ or $k$ are omitted for brevity. Common DOA estimators usually prepare steering vector $\mathbf{g}(\theta)$ for each possible DOA $\theta$ and determine the DOA by maximizing a cost-function:

$$\widehat{\theta}_S = \underset{\theta}{\operatorname{argmax}}\, J\left[\mathbf{y}, \mathbf{g}(\theta)\right]. \tag{6}$$

As an example, the GCC-Hanan-Thompson (HT) [1] technique is adopted in this paper. Because the GCC-HT is originally designed only for dual-microphone cases, and this paper assumes any microphone array configuration, the GCCs for each possible pair of microphones are summed. Accordingly, The GCC-HT is given by:

$$J\left[\mathbf{y}, \mathbf{g}(\theta)\right] = \sum_k \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\widehat{\Phi}_{\mathbf{y},i,j}}{\left|\widehat{\Phi}_{\mathbf{y},i,j}\right|} \frac{\psi_{i,j}}{1-\psi_{i,j}} \frac{G_i(\theta)}{G_j(\theta)} \tag{7}$$

where $\widehat{\Phi}_{\mathbf{y},i,j}$ is the cross-spectrum between the $i$-th and $j$-th microphone signals and $\psi_{i,j}$ is the coherence between the microphone signals calculated by $\psi_{i,j} = \frac{\left|\widehat{\Phi}_{\mathbf{y},i.j}\right|^2}{\widehat{\Phi}_{\mathbf{y},i,j}\widehat{\Phi}_{\mathbf{y},i,j}}$. The cross spectrum is usually estimated by a recursion on time:

$$\widehat{\Phi}_{\mathbf{y},i,j}(m) = \alpha\widehat{\Phi}_{\mathbf{y},i,j}(m-1) + (1-\alpha)Y_i(m)Y_j^*(m).$$

In the next section, the main contribution of this paper (memory reduction for cyclic-symmetric arrays) is presented.

## IV. Memory-Efficient DOA Estimator

In cyclic-symmetric arrays, a clockwise rotation of the array with some fundamental angle $\theta_F$ recovers the same microphone positions while only the microphone indexes are exchanged. Mathematically, this concept can be described using a clockwise rotation matrix. Assuming a cyclic-symmetric array with fundamental angle $\theta_F$, the rotated positions of the microphones are equal to the initial positions with some injective map:

$$\mathbf{R}(\theta_F)\cdot\mathbf{p}_i = \mathbf{p}_{\mathcal{F}(i)} \tag{8}$$

Where $\mathbf{R}(\theta_F)$ is a clockwise rotation matrix: [1]

$$\mathbf{R}(\theta_F) = \left(\begin{array}{cc} \cos(\theta_F) & \sin(\theta_F) \\ -\sin(\theta_F) & \cos(\theta_F) \end{array}\right) \tag{10}$$

[1]Note that the clockwise rotation matrix upholds:

$$\mathbf{R}(\theta_F)\begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix} = \begin{bmatrix}\cos(\theta-\theta_F)\\ \sin(\theta-\theta_F)\end{bmatrix} \tag{9}$$

and $\mathcal{F}: \{1,2,..,N\} \to \{1,2,..,N\}$ is an injective map between the indexes of the initial array positions and the positions after the rotation. For example, for a uniform and circular array with $N$ microphones, the fundamental angle is $\theta_F = \frac{360^o}{N}$ and the injective function is $\mathcal{F} = \mathrm{mod}(i-1,N)$, where $\mathrm{mod}$ is the modulus function.

Using the concepts described so far, the memory reduction offered in this paper can now be presented. Note that, in executing the GCC-HT estimator in (7), the steering responses $G_i(\theta,k)$ need to be calculated for each microphone, each candidate $\theta$, and each frequency $k$. Because the steering responses are time-independent, they may be pre-calculated and saved. Using the cyclic-symmetrical concept, memory usage can be reduced by a factor of $\frac{360}{\theta_F}$. In particular, the steering responses only for angles $0 < \theta < \theta_F$ can be calculated and saved, and the other steering responses can be obtained from the prepared steering responses using the aforementioned map.

For any DOA $\theta$, it can be expressed by:

$$\theta = \bar{\theta} + n\theta_F \tag{11}$$

where $\bar{\theta} = \mod(\theta,\theta_F)$ and $n = \left\lfloor\frac{\theta}{\theta_F}\right\rfloor$. Note that $\bar{\theta}$ is confined to $[0..\theta_F]$. Looking at the expression for the TDOA in (4), the TDOA for any angle can be obtained by:

$$\begin{aligned}
\tau_i(\theta) &= \tau_i(\bar{\theta}+n\theta_F) &(12)\\
&= \frac{1}{c}\cdot\left[\cos\left(\bar{\theta}+n\theta_F\right) \quad \sin\left(\theta+n\theta_F\right)\right]\mathbf{p}_i\\
&= \frac{1}{c}\cdot\left[\cos\left(\bar{\theta}\right) \quad \sin\left(\bar{\theta}\right)\right]\mathbf{R}^n(\theta_F)\mathbf{p}_i\\
&= \frac{1}{c}\cdot\left[\cos\left(\bar{\theta}\right) \quad \sin\left(\bar{\theta}\right)\right]\mathbf{p}_{\mathcal{F}^n(i)}\\
&= \tau_{\mathcal{F}^n(i)}(\bar{\theta}). &(13)
\end{aligned}$$

Using this property of the DOA and using (3), each steering response $G_i(\theta)$ can be obtained by:

$$G_i(\theta) = G_i(\bar{\theta}+n\theta_F,k) = G_{\mathcal{F}^n(i)}(\bar{\theta},k). \tag{14}$$

Finally, the GCC-HT estimator can be calculated using only the steering responses prepared for angles in the range $[0..\theta_F]$:

$$J(\theta) = \sum_k \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\widehat{\Phi}_{\mathbf{y},i,j}}{\left|\widehat{\Phi}_{\mathbf{y},i,j}\right|} \frac{\psi_{i,j}}{1-\psi_{i,j}} \frac{G_{\mathcal{F}^n(i)}\left(\bar{\theta}\right)}{G_{\mathcal{F}^n(j)}\left(\bar{\theta}\right)}. \tag{15}$$

In conclusion, the GCC-HT estimator as expressed above has the same outputs as compared to the regular expression in (7). The only modification is the saving in memory usage by preparing only steering responses for angles up to $\theta_F$. It should be noted that this saving can be added to other similar DOA estimators when using cyclic-symmetric arrays.

## V. Fundamental Angle Estimation

Sometimes, the fundamental cyclic-symmetrical angle is not simply visible. The fundamental angle of array $\theta_F$ can be automatically detected by the microphone positions using a string-matching algorithm [10]. The microphone positions are first centered around the origin and represented by polar

**Algorithm 1:** Automatic Fundamental Angle Estimation

Normalize the microphone positions around the origin by $\mathbf{p}_i \leftarrow \mathbf{p}_i - \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_i$.
Represent the microphone positions with polar coordinates by $\mathbf{r}_i = \sqrt{x_i^2 + y_i^2}$ and $\boldsymbol{\theta}_i = \text{Atan2}\left(x_i, y_i\right)$.
Erase microphone positions if $\mathbf{r}_i = 0$.
Resort $\boldsymbol{\theta}$ in an increasing order.
Rearrange $\mathbf{r}$ according to the rearrangement of $\boldsymbol{\theta}$.
Normalize $\boldsymbol{\theta}$ by $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \min(\boldsymbol{\theta})$.
**for** *each* $\theta_i$ $i = 2..N$ *or up to* $180^o$ **do**
  Initialize error function $J(\theta_i) = 0$
  **for** $j = 1..N$ **do**
    Set cyclic rotation of $i$ by $j$ using:

$$q = \mod(j - i, N)$$

    Accumulate the error function by:

$$J(\theta_i) = J(\theta_i) + (\mathbf{r}_j\cos(\theta_j - \theta_i) - \mathbf{r}_q\cos(\theta_q))^2 + (\mathbf{r}_j\sin(\theta_j - \theta_i) - \mathbf{r}_q\sin(\theta_q))^2$$

  **end**
  **if** $J(\theta_i) < \eta$ **then**
    $\theta_F = \theta_i$
    Break for-loop
  **end**
**end**

**Algorithm 2:** Automatic Cyclic-Symmetric Map Estimation

Rotate the microphone positions by $\mathbf{q}_i = \mathbf{R}(\theta_F) \cdot \mathbf{p}_i$.
For each possible microphone pair, calculate $N_{i,j} = ||\mathbf{p}_i - \mathbf{q}_j||^2$.
**for** $i = 1..N$ **do**
  Determine $\mathcal{F}(i) = \text{argmin}_j\, N_{i,j}$
  To ensure injectivity, set for all $j$ $N_{j,\mathcal{F}(i)} = \infty$
**end**

coordinates (distances and angles). Each obtained angle is a candidate to be the fundamental angle (where its maximum value is $180^o$). Then, the distance and angle pairs are rearranged in increasing order of the angles. For each candidate angle, two options are compared: 1) The rotated positions obtained by adding the candidate angle to the overall angles and 2) the corresponding cyclic rotation of the microphone positions. The squared distance in Cartesian coordinates between these two options is then serially calculated for each candidate angle (from the smaller angle up to $180^o$). When the squared distance is lower than a predefined threshold $\eta$, the process is terminated and a fundamental angle is then found. For further elaboration, the algorithm for estimating the fundamental angle is summarized in Algorithm 1.

## VI. Automatic Mapping Estimation

In this section, a way to automatically determine the mapping $\mathcal{F}(i)$ given the fundamental angle $\theta_F$ is described using the algorithm at [11]. First, the positions are rotated using the fundamental angle and the rotation matrix. Then, the rotated positions are compared to the original positions in terms of squared distance. The mapping $\mathcal{F}(i)$ is determined by the pair of positions (rotated and original) with the minimal squared distance. To ensure injectivity of $\mathcal{F}(i)$ when a map between microphone $i$ and microphone $j$ is determined, the $i+1..N$ microphone can no longer be mapped to mic $j$. For further elaboration, the algorithm for mapping the rotated microphones is summarized in Algorithm 2.

## VII. Experimental Example

The concept of the proposed memory-reduction technique is presented using real recording using CEVA's "smart and connected" development platform.[2]. Fig. 1 shows a picture of the platform, and Fig. 2 shows a picture of the recording room.


Fig. 1. Platform with 17 microphones (marked with red circles)

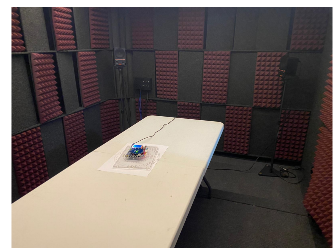

Fig. 2. Recording room

The platform has 17 digital microphones in a two-dimensional array. Twelve microphones are uniformly placed in an outer circle, four microphones are placed in an inner circle, and a single microphone is placed at the origin. When using all of the microphones, the cyclic-symmetric fundamental angle is $90^o$ (due to the inner circle); however, using only the outer circle of microphones (and perhaps the microphone at the origin), the cyclic-symmetric fundamental frequency is $30^o$.

A speaker from DOA $150^o$ was recorded in a low-reverberant and silent room. The sampling frequency was 16 kHz and the frame length of the STFT was 32 ms (512 sample lengths of the analysis window) with 8 ms between successive time frames (i.e., 75% overlap). Only frequencies between 300-3000Hz were inputted to the DOA estimator.

### A. Blind fundamental angle and mapping estimation

Four options were assumed when using Algorithm 1: (a) six microphones in the outer circle (uniformly spaced) plus the microphone at the origin ($\theta_F = 60^o$), (b) twelve microphones in the outer circle plus the microphone at the origin ($\theta_F = 30^o$), (c) twelve microphones in the outer circle, two microphones in

[2]For more details, see https://www.ceva-dsp.com/product/smart-connected-development-platform/

| Mic indexes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{F}(i)$ | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 16 | 13 | 14 | 15 | 17 |

TABLE I
THE MAPPING $\mathcal{F}(i)$ FOR 17 MICROPHONES

the inner circle (uniformly spaced), and the microphone at the origin ($\theta_F = 180^o$) and (d) All 17 microphones ($\theta_F = 90^o$). The error function $J(\theta_i)$ for each option is shown in Figures 3-6. The threshold for determining the fundamental angle was set as $\eta = 10^-6$. It can be verified that, for options (a-d), the error function has a value lower than $\eta$ only when $\theta$ is higher than $60^o$, $30^o$, $180^o$, and $90^o$ correspondingly (as expected). As an example of the blind mapping technique in 2 when using all 17 microphones and the fundamental angle $\theta_F = 90^o$, the output mapping $\mathcal{F}(i)$ is depicted in Table I. The order of microphones in $\vec{p}$ is first the microphones in the outer circle, then the microphones in the inner circle, and then the microphone at the origin.

It is should be noted that the overall array may be split into sub-arrays and a fundamental angle may be found separately for each sub-array. In this way, memory usage may be reduced for each array separately (and thus the overall memory usage may be more reduced). However, to exhibit the concept of this work, a single fundamental angle is assumed for the overall array.

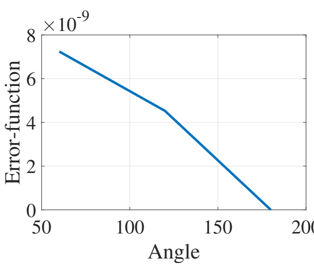


Fig. 3. $J(\theta_i)$ for option (a)

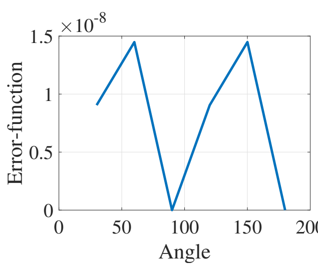


Fig. 4. $J(\theta_i)$ for option (b)

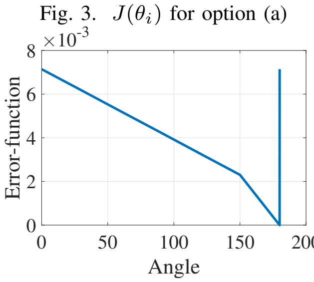


Fig. 5. $J(\theta_i)$ for option (c)

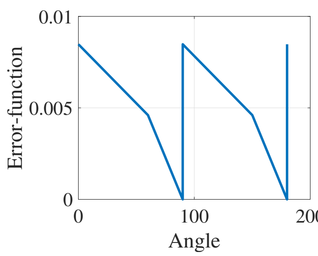


Fig. 6. $J(\theta_i)$ for option (d)

### B. DOA estimation and memory-saving analysis

The GCC-HT (7) was employed on the microphone signals using all 17 microphones. The DOA resolution is $1^o$ (360 DOAs). The DOA estimates along time and one of the signals are shown in Fig. 7 and Fig. 8. It can be seen that the outputs are $150^o$ in the time period that the speaker is active ($t > 3 \cdot 10^4$). Without using the concept of this work, to pre-prepare the steering responses $G_j(\theta, k)$, 360 (DOAs) $\times$ 87 (frequency bins) $\times$ N (microphones) $\times$ 2 (real and imaginary) memory cells are needed (which is overall 1002240 memory cells). Using the concept of this work, the number of needed memory cells is reduced by 4 (while using all 17 microphones).

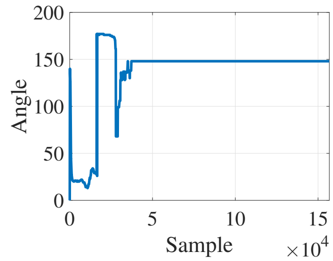


Fig. 7.

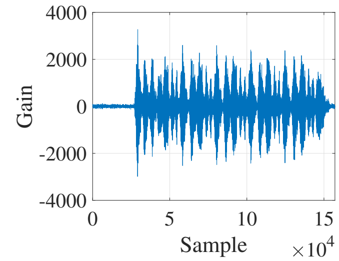


Fig. 8.

## VIII. CONCLUSIONS

DOA estimators usually need high memory usage due to the preparation and saving of the steering vector for each candidate DOA. In this paper, a technique for reducing the memory usage for cyclic-symmetrical arrays is shown. Moreover, a technique for blind estimation of the fundamental angle is described. Finally, as an experimental example, a cyclic-symmetrical array with 17 microphones was used for DOA estimation using the GCC technique and the proposed memory reduction is demonstrated.

## REFERENCES

[1] Charles Knapp and Glifford Carter, “The generalized correlation method for estimation of time delay,” *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, pp. 320–327, 1976.

[2] Ralph Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[3] Hao Ye and D DeGroat, “Maximum likelihood DOA estimation and asymptotic cramér-rao bounds for additive unknown colored noise,” *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 938–949, 1995.

[4] Ozgur Yilmaz and Scott Rickard, “Blind separation of speech mixtures via time-frequency masking,” *IEEE Transactions on signal processing*, vol. 52, no. 7, pp. 1830–1847, 2004.

[5] Michael I Mandel, Ron J Weiss, and Daniel PW Ellis, “Model-based expectation-maximization source separation and localization,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 382–394, 2010.

[6] Ofer Schwartz, Yuval Dorfan, Emanuël AP Habets, and Sharon Gannot, “Multi-speaker DOA estimation in reverberation conditions using expectation-maximization,” in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2016.

[7] Yuval Dorfan, Ofer Schwartz, Boaz Schwartz, Emanuël AP Habets, and Sharon Gannot, “Multiple DOA estimation and blind source separation using estimation-maximization,” in *IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 2016.

[8] Ofer Schwartz, Yuval Dorfan, Maja Taseska, Emanuël AP Habets, and Sharon Gannot, “DOA estimation in noisy environment with unknown noise power using the EM algorithm,” in *Hands-free Speech Communications and Microphone Arrays (HSCMA)*, 2017, pp. 86–90.

[9] Koby Weisberg, Sharon Gannot, and Ofer Schwartz, “An online multiple-speaker DOA tracking using the cappé-moulines recursive expectation-maximization algorithm,” in *ICASSP 2019*. IEEE, pp. 656–660.

[10] Peter T Highnam, “Optimal algorithms for finding the symmetries of a planar point set,” *Information Processing Letters*, vol. 22, no. 5, pp. 219–222, 1986.

[11] Olivier Collier and Arnak Dalalyan, “Permutation estimation and minimax matching thresholds,” 2013.