2019 27th European Signal Processing Conference (EUSIPCO)

A Least Squares Narrowband DOA Estimator with Robustness Against Phase Wrapping

Tobias Kabzinski[†]
Institute of Communication Systems
RWTH Aachen University, 52056 Aachen, Germany

Emanuël A. P. Habets
International Audio Laboratories Erlangen[*]
Am Wolfsmantel 33, 91058 Erlangen, Germany

***Abstract*—Narrowband direction-of-arrival (DOA) estimates are commonly used for source localization, parametric spatial audio coding, and directional filtering. As previously shown, a linear least squares direction estimate can be obtained by minimizing the difference of expected and observed inter-microphone phase differences. In this work, it is shown that phase wrapping induces severe estimation errors especially at frequencies just below spatial aliasing frequencies and in low signal-to-noise ratios. A cost function to mitigate the influence of phase wrapping errors on the DOA estimation is proposed. Even though the proposed cost function is nonlinear, it is shown that one iteration of a gradient descent method with proper initialization provides a large improvement when compared to the linear least squares solution.**

***Index Terms*—Direction-of-arrival estimation, narrowband, microphone arrays, phase wrapping, source localization**

I. Introduction

Direction-of-arrival (DOA) estimation is a commonly required step in microphone array processing. Especially, narrowband DOA estimators are used for source localization and counting [1], parametric spatial audio coding and processing [2], [3] and informed spatial filtering [4].

A multitude of DOA estimation approaches has been proposed (c.f. [5] and the references therein) based on signal subspaces [6], [7], time-differences-of-arrival [8], [9], inter-microphone phase differences [10], [11], and steered response power (SRP) [12], [13]. In [11] a phase-differences-based method has been proposed which supports arbitrary array geometries and requires low computational complexity. It involves minimizing the squared error between observed and expected phases of cross power spectral densities (CPSDs) of microphone pairs. Phase values computed from the observed complex-valued CPSDs are limited to the range $(-\pi, \pi]$. So-called phase wrapping occurs when disturbances, caused by reverberation, noise or estimation errors, change the original phase near $\pm\pi$ to a value close to $\mp\pi$. As the errors between observed and expected phase in [11] are measured using the Euclidean distance, severe DOA estimation errors can occur in case of phase wrapping.

In this contribution, we first analyze when phase wrapping can lead to large DOA estimation errors. We then propose a method to find a DOA estimate which is more robust against phase wrapping. Therefore, the least squares problem from [11] is reformulated taking into account that the DOAs corresponding to phase values of $+\pi$ and $-\pi$ are close.

The remainder of this paper is organized as follows. In Section II the DOA estimator proposed in [11] is reviewed and its limitations are analyzed in Section III. To mitigate the consequences of phase wrapping in the DOA estimation, an improved method is proposed in Section IV, and its performance is evaluated and discussed in Section V. Finally, conclusions are provided in Section VI.

II. Problem Description

In Sections II-A and II-B, we present the signal model and review the weighted least squares (WLS) DOA estimator proposed in [11], respectively.

A. Signal Model

We assume a microphone array of M microphones at positions $\mathbf{r}_i = [r_{\mathrm{x},i}, r_{\mathrm{y},i}, r_{\mathrm{z},i}]^{\mathrm{T}}, i = 1, \ldots, M$. The short-time Fourier transform (STFT) representation of the received microphone signals at time index λ and frequency index k are combined into a vector

$$\begin{aligned}\mathbf{x}(k,\lambda) &= [X_1(k,\lambda), X_2(k,\lambda), \ldots, X_M(k,\lambda)]^{\mathrm{T}} \\ &= \mathbf{x}_{\mathrm{s}}(k,\lambda) + \mathbf{x}_{\nu}(k,\lambda). \end{aligned} \quad (1)$$

Here, $\mathbf{x}_{\mathrm{s}}(k,\lambda) = [X_{\mathrm{s},1}(k,\lambda), \ldots, X_{\mathrm{s},M}(k,\lambda)]^{\mathrm{T}}$ represents a signal corresponding to a single free-field and far-field sound source. It should be noted that at each time and frequency a different source can be active. This assumption is known as W-disjoint orthogonality and commonly holds for mixtures of speech signals [14], [15]. The second term $\mathbf{x}_{\nu}(k,\lambda) = [X_{\nu,1}(k,\lambda), \ldots, X_{\nu,M}(k,\lambda)]^{\mathrm{T}}$ models microphone self-noise and/or a diffuse sound component. Therefore, $\mathbf{x}_{\nu}(k,\lambda)$ is either uncorrelated or exhibits frequency-dependent spatial coherence, e.g., corresponding to an isotropic diffuse sound field [16]. As each time instance and discrete frequency can be processed independently, the dependence of time and frequency is dropped for brevity where possible.

B. Weighted Least Squares DOA Estimator

The DOA estimator in [11] is based on minimizing the error between observed and expected phase differences of (a subset

[†] The work was conducted while the author was at the International Audio Laboratories Erlangen.

[*] A joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS

of) microphone pairs. Precisely, the phase difference between the two microphones corresponds to the phase of their CPSD. For M microphones, let

$$\begin{aligned}\mathcal{B} &= \{\beta_1, \beta_2, \ldots, \beta_B\} \\ &= \{(1,2), \ldots, (1,M), \ldots, (M-1,M)\} \\ &= \left\{(i,j) \in \mathbb{N}^2 \mid 1 \leq i \leq M-1, i < j \leq M\right\}\end{aligned} \quad (2)$$

denote the ordered set of all $B = |\mathcal{B}| = \frac{M}{2}(M-1)$ unique pairs of microphones.

Let $\hat{\Phi}_{\beta_b}$, $b = 1, \ldots, B$, denote a CPSD estimate for microphone pair β_b. Then, all B observed phases can be summarized in a vector

$$\hat{\boldsymbol{\mu}} = \left[\angle\hat{\Phi}_{\beta_1}, \angle\hat{\Phi}_{\beta_2}, \ldots, \angle\hat{\Phi}_{\beta_B}\right]^{\mathrm{T}}. \quad (3)$$

Given a source direction of azimuth φ and elevation ϑ,

$$\mathbf{n}(\varphi, \vartheta) = \left[\cos(\varphi)\cos(\vartheta),\ \sin(\varphi)\cos(\vartheta),\ \sin(\vartheta)\right]^{\mathrm{T}} \quad (4)$$

defines a unit-norm vector pointing from a reference point, e.g., the coordinate system origin, to the source position. The expected path difference of sound traveling from the source to i-th and j-th microphones of microphone pair β_b is obtained by projecting the direction vector $\mathbf{n}$ onto the position difference

$$\mathbf{d}_{\beta_b} = \mathbf{r}_j - \mathbf{r}_i \quad (5)$$

of microphone pair β_b. Path difference and phase difference are related by a physical constant, the wavenumber

$$\kappa(k) = 2\pi \frac{c}{f(k)} = 2\pi \frac{2Lc}{f_{\mathrm{s}} k}. \quad (6)$$

Here, c denotes the speed of sound in meters per second, f_{s} is the sampling frequency and L is the STFT frame length. The vector of all expected CPSD phases $\boldsymbol{\mu}$ is calculated by $\mathbf{Q}\,\mathbf{n}$, where

$$\mathbf{Q} = \kappa\left[\mathbf{d}_{\beta_1}, \mathbf{d}_{\beta_2}, \ldots, \mathbf{d}_{\beta_B}\right]^{\mathrm{T}}. \quad (7)$$

Depending on the frequency $f(k)$ microphone pair β_b is excluded from the estimation if it operates above its spatial aliasing frequency

$$f_{\mathrm{A},b} = \frac{c}{2\left\|\mathbf{d}_{\beta_b}\right\|}. \quad (8)$$

This is achieved by pre-multiplying both the observed CPSD phases $\hat{\boldsymbol{\mu}}$ and $\mathbf{Q}\,\mathbf{n}$ with a diagonal weighting matrix $\mathbf{W} = \mathrm{diag}(W_{11}, \ldots,\ W_{bb}, \ldots, W_{BB})$ with entries

$$W_{bb}(k) = \begin{cases} 1 & \text{if } f(k) \leq f_{\mathrm{A},b} \\ 0 & \text{otherwise} \end{cases}. \quad (9)$$

An estimate of the source direction vector $\hat{\mathbf{n}} = [\hat{n}_{\mathrm{x}}, \hat{n}_{\mathrm{y}}, \hat{n}_{\mathrm{z}}]^{\mathrm{T}}$ can be obtained by solving the weighted least squares (WLS) problem

$$\arg\min_{\mathbf{n}} \left\|\mathbf{W}\hat{\boldsymbol{\mu}} - \mathbf{W}\mathbf{Q}\mathbf{n}\right\|^2. \quad (10)$$

Its solution is given by

$$\hat{\mathbf{n}} = \left[\mathbf{Q}^{\mathrm{T}}\mathbf{W}\mathbf{Q}\right]^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{W}\hat{\boldsymbol{\mu}}. \quad (11)$$

Note that $\hat{\mathbf{n}}$ is not necessarily of unit norm. Estimates of φ and ϑ can be obtained from a normalized version of $\hat{\mathbf{n}}$:

$$\hat{\varphi} = \mathrm{atan2}\left(\frac{\hat{n}_{\mathrm{y}}}{\|\hat{\mathbf{n}}\|}, \frac{\hat{n}_{\mathrm{x}}}{\|\hat{\mathbf{n}}\|}\right) = \mathrm{atan2}\left(\hat{n}_{\mathrm{y}}, \hat{n}_{\mathrm{x}}\right), \quad (12)$$

where $\mathrm{atan2}(\cdot)$ is the four-quadrant inverse tangent, and

$$\hat{\vartheta} = \arcsin\left(\frac{\hat{n}_{\mathrm{z}}}{\|\hat{\mathbf{n}}\|}\right). \quad (13)$$

III. Analysis and Oracle Unwrapping

For a microphone pair β_b which is operated near its spatial aliasing frequency, the expected phase of the CPSD, $\kappa\mathbf{d}_{\beta_b}^{\mathrm{T}}\mathbf{n}$, is close to $\pm\pi$ if the sound source direction points in a similar direction as the axis defined by the position difference of the microphone pair (cf. (6), (8)). The estimated CPSD

$$\begin{aligned}\hat{\Phi}_{\beta_b} &= \Phi_{\beta_b}^{\mathrm{s}} + \Phi_{\beta_b}^{\nu} + \tilde{\Delta}\Phi_{\beta_b} \\ &= \Phi_{\beta_b}^{\mathrm{s}} + \Delta\Phi_{\beta_b}\end{aligned} \quad (14)$$

is comprised of the true CPSD $\Phi_{\beta_b}^{\mathrm{s}} = \mathrm{E}\left\{X_{\mathrm{s},i}X_{\mathrm{s},j}^{*}\right\}$ and an error term $\Delta\Phi_{\beta_b}$. If the noise components $X_{\nu,m}$ are mutually uncorrelated, $\Phi_{\beta_b}^{\nu}$ vanishes. To also consider the case in which the noise components exhibit spatial coherence, e.g., modeling a diffuse component of reverberation, the error term $\Delta\Phi_{\beta_b}$ represents both estimation error and a coherence contribution. Due to the error term the observed phase of a CPSD might change from $\pm\pi \mp \epsilon_1$ to $\mp\pi \pm \epsilon_2$. For this case of phase wrapping, the corresponding phasor representation of (14) is illustrated in Figure 1.

Taking into account the phase of $\Phi_{\beta_b}^{\mathrm{s}}$, we define that the observed phase is wrapped if the following condition applies:

$$\left|\angle\hat{\Phi}_{\beta_b} - \angle\Phi_{\beta_b}^{\mathrm{s}}\right| > \pi. \quad (15)$$

If the additive error term in (14) is known, e.g., in a simulation environment, the phase wrapping can be corrected. This correction is referred to as oracle unwrapping in the remainder of this paper. The unwrapped phase for the b-th microphone pair is given by

$$\tilde{\mu}_{\beta_b} = \begin{cases} \angle\hat{\Phi}_{\beta_b} + 2\pi & \text{if } \angle\hat{\Phi}_{\beta_b} - \angle\Phi_{\beta_b}^{\mathrm{s}} < -\pi \\ \angle\hat{\Phi}_{\beta_b} - 2\pi & \text{if } \angle\hat{\Phi}_{\beta_b} - \angle\Phi_{\beta_b}^{\mathrm{s}} > \pi \\ \angle\hat{\Phi}_{\beta_b} & \text{otherwise} \end{cases}. \quad (16)$$

Phase wrapping of $\angle\hat{\Phi}_{\beta_b}$ can significantly enlarge the error contribution of pair β_b since $\angle\hat{\Phi}_{\beta_b}$ and $\kappa\mathbf{d}_{\beta_b}^{\mathrm{T}}\mathbf{n}$ are of opposite sign. The phase error, as depicted in Figure 1, is $\epsilon_1 + \epsilon_2$, not $2\pi - (\epsilon_1 + \epsilon_2)$. The WLS cost function which is minimized in (10) does not reflect this. Hence, the least squares solution, strongly influenced by outliers, can lead to a severe DOA estimation error. To overcome the limitations of the cost function based on the Euclidean distance, a more suitable cost function is introduced in the following section.

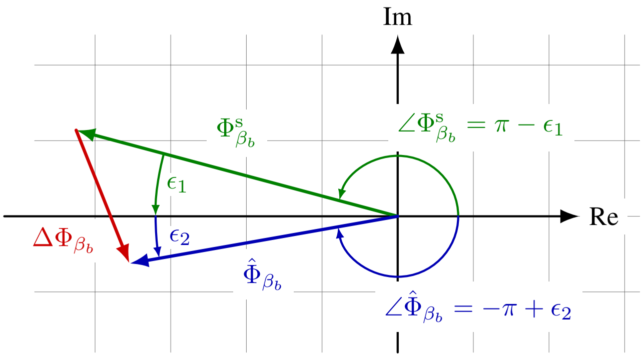

Fig. 1. A phasor representation of (14) with phase wrapping.

IV. Proposed Solution

As seen in Section II, the cost function in (10) is advantageous because the solution is given in closed-form and thus yields low computational complexity. However, as discussed in Section III, it is largely influenced by errors of those microphone pairs suffering from phase wrapping. This is undesirable as it could lead to erroneous DOA estimates. Instead of measuring the squared Euclidean distance of the observed and expected CPSD phases, i.e., $\left|\angle\hat{\Phi}_{\beta_b} - \kappa \mathbf{d}^{\mathrm{T}}_{\beta_b}\mathbf{n}\right|^2$, the squared Euclidean distance of the corresponding phasors in the complex plane yields a more suitable error measure. It takes into account that phase values of $\pm\pi$ are close and can be written as

$$\left\|\exp\left(\jmath\angle\hat{\Phi}_{\beta_b}\right) - \exp\left(\jmath\kappa \mathbf{d}^{\mathrm{T}}_{\beta_b}\mathbf{n}\right)\right\|^2, \qquad (17)$$

where $\jmath = \sqrt{-1}$. Separating real and imaginary parts, the combined cost function with error contributions from all selected microphone pairs then reads

$$\xi(\varphi,\vartheta) = \left\|\begin{bmatrix}\cos(\mathbf{W}\hat{\boldsymbol{\mu}})\\ \sin(\mathbf{W}\hat{\boldsymbol{\mu}})\end{bmatrix} - \begin{bmatrix}\cos(\mathbf{WQn}(\varphi,\vartheta))\\ \sin(\mathbf{WQn}(\varphi,\vartheta))\end{bmatrix}\right\|^2, \qquad (18)$$

where $\cos(\cdot)$ and $\sin(\cdot)$ are applied elementwise. It can be shown that minimizing (18) is equivalent to minimizing the narrowband PHAT-weighted version of the SRP cost function. Azimuth and elevation estimates can then be obtained as

$$\left(\hat{\varphi},\hat{\vartheta}\right) = \underset{\varphi,\vartheta}{\arg\min}\ \xi(\varphi,\vartheta). \qquad (19)$$

Unfortunately, (19) is a non-convex in (φ,ϑ). Therefore, it is suggested to obtain a (locally optimal) solution by applying an iterative method, such as a quasi-Newton method with cubic line search [17], [18]. The gradient of (18) can be derived analytically and the Hessian is updated according to the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [19]–[22]. Iterations are repeated until the algorithm has converged, i.e., either all components of the gradient are less than a specified tolerance γ_{grad} or the norm of the Newton step falls below the tolerance γ_{step}. Note that this new problem formulation does not pose any restriction on the array geometry. The proposed method is referred to as the phase wrapping robust (PWR) DOA estimation method.

V. Performance Evaluation

To compare the performance of the proposed solution to the WLS solution [11], several simulations were conducted. Due to space constraints, the evaluation is limited to one- and two-dimensional arrays and noise excitation.

A. Description of Simulations

A uniform linear array (ULA) of $M = 5$ microphones with a spacing of $3.4\,\mathrm{cm}$ and a uniform circular array (UCA) of radius $3.4\,\mathrm{cm}$ with $M = 8$ microphones serve as examples. For up to two-dimensional arrays only one value, the azimuth angle, is estimated. Therefore, the DOA estimation performance can be characterized by mean and standard deviation of the absolute angular (azimuth) deviation defined by

$$\varepsilon = \mathrm{E}\left\{\left|\arccos\left(\frac{\hat{\mathbf{n}}^{\mathrm{T}}\mathbf{n}}{\|\hat{\mathbf{n}}\|}\right)\right|\right\} \qquad (20)$$

and

$$\sigma = \sqrt{\mathrm{E}\left\{\left|\arccos\left(\frac{\hat{\mathbf{n}}^{\mathrm{T}}\mathbf{n}}{\|\hat{\mathbf{n}}\|}\right)\right|^2\right\} - \varepsilon^2}. \qquad (21)$$

Additionally, for the one-dimensional case, the angle between $\hat{\mathbf{n}}$ and $\mathbf{n}$ is computed considering the ambiguity of the azimuth angle w.r.t. the array axis.

In total, 20 000 realizations with azimuth angles distributed uniformly in $[0, 2\pi]$ were computed and averaged to approximate the expectation operation. The source positions were simulated to be on a circle of $2\,\mathrm{m}$ radius around the microphone array center, which coincides with the coordinate system origin. White Gaussian noise was used as a source signal. The microphone signals were obtained by filtering the source signal with an anechoic room impulse response describing the transfer function between source position and microphone. Additionally, independent white Gaussian noise of variance σ^2_ν was added to each microphone signal to achieve a signal-to-noise ratio (SNR) of

$$\mathrm{SNR} = 10\log_{10}\left(S_{\text{ref}}/\sigma^2_\nu\right), \qquad (22)$$

where S_{ref} denotes the signal power received at one of the microphones (i.e., the reference microphone).

The STFT uses a Hann window [23] and a frame length of $L = 2048$ with 50% overlap. To estimate the short-term CPSDs, five consecutive frames are averaged. The sampling frequency is set to $48\,\mathrm{kHz}$. The tolerances for convergence are set to $\gamma_{\text{grad}} = \gamma_{\text{step}} = 10^{-6}$ and the quasi-Newton implementation from MATLAB [24] is used.

The following methods are compared:

- WLS-OU: weighted least squares [11] with oracle unwrapping using (16),
- WLS: as proposed in [11],
- PWR-Rand-Con: PWR method randomly initialized and repeated until convergence is achieved,
- PWR-WLS-Con: PWR method initialized by WLS solution and repeated until convergence is achieved,
- PWR-WLS-One: PWR method initialized by WLS solution and conducting one iteration (only in Figures 3 and 4).

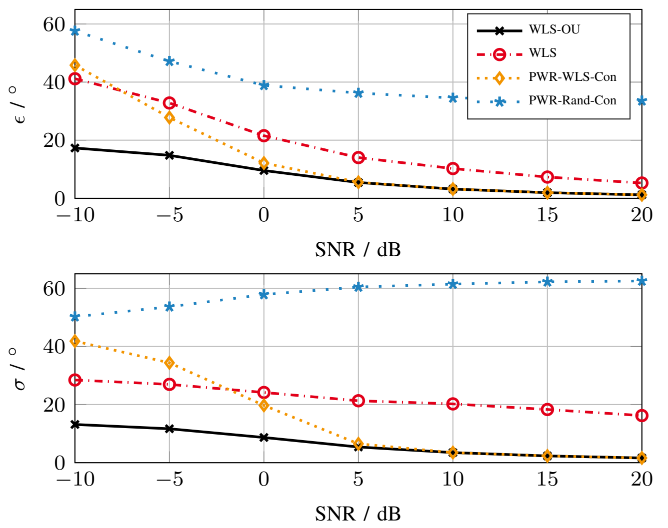

Fig. 2. Performance measures in dependence of SNR at fixed frequency $f(k) = 1664$ Hz close to spatial aliasing frequency $f_{\mathrm{A}} = 1667$ Hz of ULA.

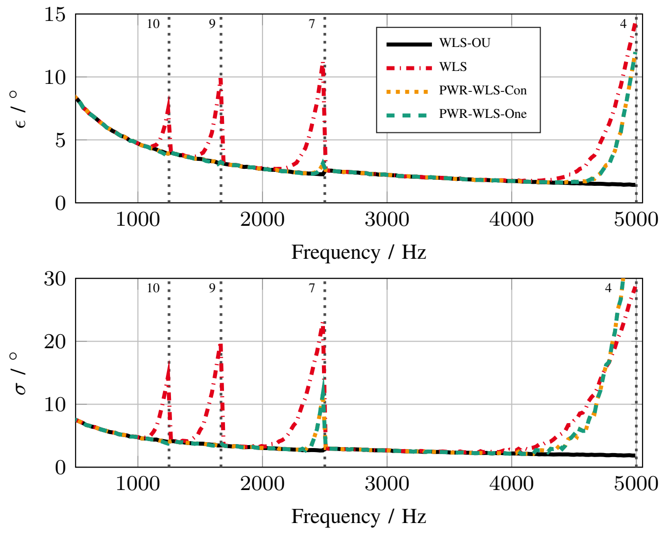

Fig. 3. Performance measures at 10 dB SNR for ULA.

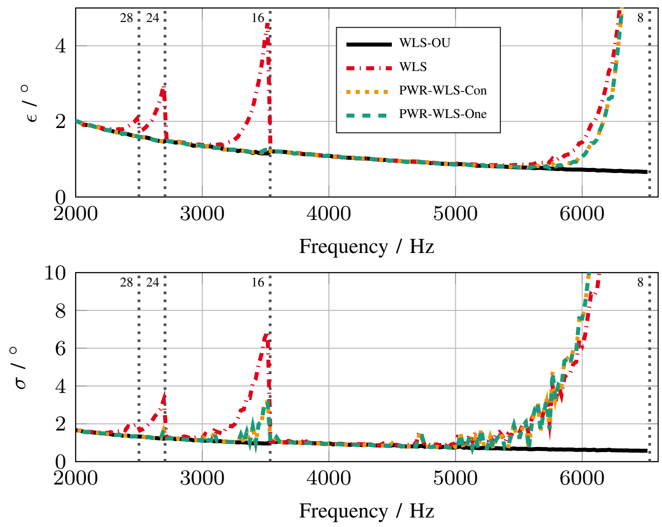

Fig. 4. Performance measures at 10 dB SNR for UCA.

B. Results and Discussion

Firstly, the performance is studied for a fixed frequency close to a spatial aliasing frequency of one of the microphone pairs. Figure 2 shows the mean and the standard deviation of the absolute azimuth deviation for different SNRs for the ULA. As expected, the oracle unwrapping provides an upper performance bound as the phase errors in (10) are correctly quantified after unwrapping. The performance of WLS being worse than the oracle demonstrates that errors due to phase wrapping, as described in Section III, actually occur. The proposed method with random initialization performs worst. Due to the random initialization, only local minima of (18) are obtained, even in high SNR conditions. Therefore, the random initialization is not considered further. For positive SNRs, the PWR-WLS approach is better than the WLS algorithm, e.g., at 10 dB ϵ decreases from about $10\,°$ to $3\,°$ and σ decreases from $20\,°$ to $3\,°$. For SNRs greater than 5 dB the PWR-WLS performance is very close to the performance with oracle unwrapping. For lower SNRs, many microphone pairs suffer from phase wrapping which cannot be corrected by the proposed algorithm as the initial values from the WLS algorithm are too far from the optimum. Hence, only a local optimum of (18) is obtained which does not necessarily correspond to an improved direction estimate.

Secondly, the frequency-dependent DOA estimation performance is evaluated for the two example array configurations. Figures 3 and 4 display the means and the standard deviations of the absolute azimuth deviation at different frequencies. The SNR is selected to be 10 dB. Vertical dashed lines mark spatial aliasing frequencies and the numbers of active microphone pairs, i.e., non-zero entries in $\mathbf{W}$, are indicated alongside. For both arrays the mean and the standard deviation of the estimation error using the WLS algorithm increase for frequencies just below spatial aliasing frequencies. Hence, phase wrapping errors degrade the estimation performance of WLS. Again, applying the oracle unwrapping before, the performance does not degrade. It can be seen that mean and standard deviation increase a bit just after each spatial aliasing frequency. This is a result of using fewer microphone pairs as pairs above their spatial aliasing frequency are excluded. In the frequency region just below a spatial aliasing frequency the performance, in terms of both mean and standard deviation, is considerably increased using the proposed method. Either the performance is on par with the oracle (e.g., below 2300 Hz for ULA and below 3000 Hz for UCA) or considerably better than WLS (e.g., 2400 Hz for ULA and 3500 Hz for UCA).

Towards the highest spatial aliasing frequency of each array, the standard deviation of the estimation error for PWR-WLS exceeds the one of the WLS method. Significant errors are observed when more than one microphone pair suffers from phase wrapping. If these cases are excluded, the standard deviation for PWR-WLS and WLS are similar (not shown). As those pairs above their spatial aliasing frequencies are excluded, fewer microphone pairs are available towards the highest spatial aliasing frequency. When the portion of microphone pairs suffering from phase wrapping becomes too large, the robustness limit is exceeded, and the proposed solution fails. Obtaining an improvement with PWR-WLS cannot be guaranteed when the initialization by the WLS solution is severely wrong. Therefore, PWR processing is only useful when a sufficient number of microphone pairs is available.

The performances of PWR-WLS-Con and PWR-WLS-One are very close, which demonstrates that conducting one iteration is sufficient. Note that the BFGS quasi-Newton method reduces to steepest descent if only the first iteration is conducted and the Hessian is initialized with an identity matrix. Thus, the computational complexity is further reduced. Furthermore, the additional step of PWR is only necessary for certain frequency bands. It depends on the SNR how far the frequency region extends below a spatial aliasing frequency in which the WLS methods performs poorly. In large parts of the frequency range PWR processing is unnecessary, e.g., from $2600\,\mathrm{Hz}$ to $4000\,\mathrm{Hz}$ for the ULA, and from $3600\,\mathrm{Hz}$ to about $5000\,\mathrm{Hz}$ for the UCA under test.

VI. Conclusion

The influence of phase wrapping errors on the weighted linear least squares narrowband DOA estimator from [11] was investigated theoretically and using simulations. It has been shown that phase wrapping especially occurs if a microphone pair is operated closely below its spatial aliasing frequency. When lowering the SNR conditions, the amount of phase wrapping errors increases. The Euclidean distance of CPSD phases used in the original WLS approach to quantify the phase errors leads to estimation errors. It is not suitable to quantify phase errors properly as it does not reflect that phase values of $\pm\pi$ are close. Therefore, measuring the phase errors in the complex plane is proposed and a corresponding cost function is introduced. It is shown to provide a more robust solution in simulations for different array geometries. The proposed solution builds on the result of [11] and achieved with only one iteration a significant improvement for the ULA and UCA under test. The additional complexity can be limited by computing the proposed solution only for selected frequency bands, which can be determined experimentally for a given array geometry and SNR conditions. Future work could include a comprehensive evaluation with real-world signals in reverberant environments. Moreover, the development and investigation of alternative strategies to increase the robustness just below the spatial aliasing frequency could be performed.

References

[1] A. Alexandridis and A. Mouchtaris, "Multiple Sound Source Location Estimation and Counting in a Wireless Acoustic Sensor Network," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.

[2] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[3] K. Kowalczyk, O. Thiergart, M. Taseska, G. Del Galdo, V. Pulkki, and E. A. P. Habets, "Parametric Spatial Sound Processing: A Flexible and Efficient Solution to Sound Scene Acquisition, Modification, and Reproduction," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 31–42, Mar 2015.

[4] O. Thiergart, M. Taseska, and E. A. P. Habets, "An Informed Parametric Spatial Filter Based on Instantaneous Direction-of-Arrival Estimates," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2182–2196, Dec 2014.

[5] N. Madhu and R. Martin, "Acoustic Source Localization with Microphone Arrays," in *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds., pp. 135–170. John Wiley & Sons, Ltd, 2008.

[6] B. D. Rao and K. V. S. Hari, "Performance Analysis of Root-Music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, Dec 1989.

[7] R. Roy and T. Kailath, "ESPRIT—Estimation of Signal Parameters Via Rotational Invariance Techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, Jul 1989.

[8] Y. T. Chan and K. C. Ho, "A Simple and Efficient Estimator for Hyperbolic Location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, Aug 1994.

[9] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereati, "Real-Time Passive Source Localization: A Practical Linear-Correction Least-Squares Approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, Nov 2001.

[10] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA Estimation for Multiple Sparse Sources with Arbitrarily Arranged Multiple Sensors," *Journal of Signal Processing Systems*, vol. 63, no. 3, pp. 265–275, Jun 2011.

[11] O. Thiergart, W. Huang, and E. A. P. Habets, "A Low Complexity Weighted Least Squares Narrowband DOA Estimator for Arbitrary Array Geometries," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2016, pp. 340–344.

[12] J. H. DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*, Ph.D. thesis, Brown University, Aug 2000.

[13] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust Localization in Reverberant Rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds., pp. 157–180. Springer, Berlin, Heidelberg, 2001.

[14] S. Rickard and O. Yilmaz, "On the Approximate W-Disjoint Orthogonality of Speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2002, vol. 1, pp. I–529–I–532.

[15] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures Via Time-Frequency Masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, Jul 2004.

[16] G. W. Elko, "Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds., pp. 61–85. Springer, Berlin, Heidelberg, 2001.

[17] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Ltd, 1980.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 3 edition, 2007.

[19] C. G. Broyden, "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[20] R. Fletcher, "A New Approach to Variable Metric Algorithms," *The Computer Journal*, vol. 13, no. 3, pp. 317–322, 1970.

[21] D. Goldfarb, "A Family of Variable-Metric Methods Derived by Variational Means," *Mathematics of Computation*, vol. 24, pp. 23–26, 1970.

[22] D. F. Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computation*, vol. 24, pp. 647–656, 1970.

[23] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Prentice-Hall, Inc., 2 edition, 1999.

[24] MATLAB, *Version 9.4.0 (R2018a)*, The MathWorks Inc., Natick, MA, USA, 2018.