

MULTI-SCALE MULTI-LAG CHANNEL ESTIMATION USING LOW RANK STRUCTURE OF RECEIVED SIGNAL

Sajjad Beygi, Urbashi Mitra

School of Electrical Engineering, University of Southern California

Emails: {beygi, ubli}@usc.edu

ABSTRACT

Underwater acoustic channels are wideband time-varying channels, which can be well-described by a multi-scale multi-lag channel model. In this paper, a robust method to estimate the channel parameters from noisy measurements is proposed. The proposed method computes the multiple Doppler scales, delays, and channel attenuation gains corresponding to different propagation paths. In this work, we adapt a spectral line estimation algorithm with a low number of measurements and modest complexity to compute the unknown channel parameters. The performance of the proposed estimation strategy is investigated via numerical simulation and shows that our method has at least 5 to 10 dB improvement in signal-to-noise ratio over previously proposed methods.

Index Terms— Underwater acoustic channels, Doppler scaling, Multiscale-Multilag channel, OFDM.

1. INTRODUCTION

The need to for high performance underwater acoustic (UWA) communication is motivated by both commercial and military applications. Achieving this goal is challenged by the low speed of propagation of sound in water, as well as the relatively high Doppler induced by mobility. In narrowband communication channels, the Doppler effect can be modeled as a frequency offset. Hence, when the maximum-likelihood (ML) approach is applied to estimate the parameters, it results in a simple correlation-based algorithm to estimate the frequency offset [1]; delays are computed using cyclic-prefix properties, and channel gains can be estimated using a least squares estimator [2].

In wideband signaling environments such as underwater acoustic communications, the Doppler distortion on each path results in a time scaling (compression or dilation) of the signal [3,4]. As such, the effective channel can be well described by a MSML channel model [4,5]. In contrast, for multi-scale, multi-lag (MSML) channels, the corresponding maximum-likelihood (ML) approach requires solving a multi-dimensional non-linear least-squares problem, incurring high complexity. In [6], the MSML channel model is considered, and subspace algorithms from the array processing literature, namely Root-MUSIC [7] and ESPRIT [7], are applied to estimate the MSML channel. Herein, we will compare our proposed method to those of [6] over which we get significant performance improvement. In this paper, we introduce a new method to jointly estimate the Doppler scaling factors, delays and channel attenuation gains

given that the number of different propagation paths is known. The main contributions of our work are: (i) we adapt and robustify the spectral estimation problem to estimate the MSML channel from noisy measurements; (ii) the proposed method is able to estimate the closely spaced frequencies generated by the scaling effect in the MSML channel; (iii) capability of high resolution estimation with low number of measurements with help of the structure of received signal.

The rest of this paper is organized as follows. The signal and MSML channels model are presented in Section 2. Section 3 presents the proposed channel estimation algorithm. Section 4 presents the numerical simulations to verify performance of the proposed algorithm, and Section 5 concludes the paper.

2. COMMUNICATION OVER MSML CHANNEL

The transmitted passband OFDM signal is given by

$$x(t) = \sum_{k=1}^K s_k e^{j2\pi f_k t}, \quad (0 \leq t \leq T), \quad (1)$$

where T is the OFDM symbol duration, K is the number of sub-carriers, s_k is the data modulated onto the k^{th} subcarrier; f_k is the k^{th} subcarrier frequency, where $f_k = f_{\min} + k\Delta f$; $\Delta f = \frac{1}{T}$ is the sub-carrier spacing; f_{\min} is minimum carrier frequency; and $B = K\Delta f$ is the bandwidth of the system. A rectangular pulse shape over the interval $t \in [0, T]$ is employed. The signal, after passing through a linear time-varying (LTV) channel, can be written as,

$$y(t) = \int_{-\infty}^{+\infty} h(t, \tau) x(t - \tau) d\tau + n(t), \quad (2)$$

where $n(t)$ is assumed to be additive, white Gaussian noise. The received signal is an aggregation of several scaled copies of the delayed and attenuated transmitted signal. The MSML channel model can be represented by $h(t, \tau) = \sum_{n=1}^M h_n(t) \delta(\tau - \tau_n(t))$, where $h_n(t)$ is the path amplitude, $\tau_n(t)$ is the time-varying path delay, and M is the number of dominant propagation paths. The continuously time varying delays are caused by motion of the transmitter/receiver as well as scattering off of the moving sea surface or refraction due to sound speed variations. The path amplitudes change with the delays as, the attenuation is related to the distance traveled as well as the physics of the scattering and propagation processes. For the duration of an OFDM symbol, the time variation of the path delays $\tau_n(t)$ evolves linearly as a function of time, namely $\tau_n(t) = \tau_n - a_n t$ where a_n is Doppler scaling factor. Thus the channel impulse response can

This research was funded in part by one or all of these grants: ONR N00014-09-1-0700, NSF CNS-0832186, NSF CCF-1117896, CCF-0917343, CNS-1213128, AFOSR FA9550-12-1-0215, and DOT CA-26-7084-00.

be simplified to

$$h(t, \tau) = \sum_{n=1}^M h_n \delta(\tau - [\tau_n - a_n t]). \quad (3)$$

Then the received signal, $y(t)$, can be represented as

$$y(t) = \sum_{n=1}^M \sum_{k=1}^K h_n s_k e^{j2\pi f_k(t - (\tau_n - a_n t))} + n(t), \quad (4)$$

where $n(t)$ is additive noise.

Remark 1. Suppose the frequencies f_k lie in $[-W, W]$, namely $B = 2W = K\Delta f$, and $d(t)$ is a continuous signal of the form: $d(t) = \sum_{n=1}^M \sum_{k=1}^K h_n s_k e^{j2\pi f_k(t - (\tau_n - a_n t))}$. By taking regularly spaced Nyquist samples at $t \in \{\frac{i}{2W} | i \in \mathcal{I}_s\}$, where $\mathcal{I}_s = \{0, 1, 2, \dots, N_s - 1\}$, we observe

$$d[i] = \sum_{n=1}^M \sum_{k=1}^K h_n s_k e^{-j2\pi f_k \tau_n} e^{j[2\pi \frac{f_k}{2W}(a_n + 1)]i},$$

where $\frac{f_k}{2W} \in [0, 1]$. Therefore after a trivial translation of the frequency domain, we can map $\frac{f_k}{2W} = \frac{f_{k \text{ in}}}{K} + \frac{k}{K}$ to the new normalized frequency for $d[i]$ as $\hat{f}_k = \frac{k}{K}$. Thus, any mixture of sinusoids after appropriate normalization, can be assumed to have frequencies in $[0, 1]$.

We can express the sampled signal as $y[i] = d[i] + n[i]$, where the index i denotes the sample time. We rewrite $d[i]$ as

$$d[i] = \sum_{l=1}^{MK} c_l z_l^i, \quad (5)$$

where $l = (k - 1)M + n$ with $1 \leq k \leq K$ and $1 \leq n \leq M$,

$$c_{(k-1)M+n} = h_n e^{-j2\pi \hat{f}_k \tau_n} s_k, \quad (6)$$

$$z_{(k-1)M+n} = e^{j2\pi f_k(1+a_n)}. \quad (7)$$

From Equation (7), we see that the z_k are clustered around each subcarrier frequency due to small value of a_n .

3. CHANNEL ESTIMATION

A natural approach to solve the parametric estimation problem in (4) is maximum likelihood estimation (MLE). We remark that MLE can take the form of a spectral estimation problem considering the representation form in (5), which consists of retrieving the parameters of a sum of complex exponentials from noisy samples. When the unknown extant frequencies in the signal are not too close to each other, classical spectral estimation techniques like MUSIC [7], and ESPRIT [7], or greedy strategies, can be used; they are fast, but statistically suboptimal. Here, due to the typical values of a_n , frequencies in each cluster are too close to each other, resulting in poor performance of the aforementioned methods. In the sequel, we suggest an approximation method for ML which is based on standard line spectral estimation methods and its robust extensions which we specialize to the MSML channel case. The resulting method is of modest complexity, and is effective in separating the closely spaced frequencies in the noisy signal.

3.1. Standard line spectral estimation algorithm

We determine the unknown parameters $c_l \in \mathbb{C}$ and $z_l \in \mathbb{C}$ for $1 \leq l \leq MK$ in the following model using N_s available measurements of the signal,

$$d[i] = \sum_{l=1}^{MK} c_l z_l^i \quad \text{for } \forall i \in \mathcal{I}_s. \quad (8)$$

The basic idea is to recognize that Equation (8) is the solution to a homogeneous difference equation whose characteristic equation has roots equal to the poles in (8), namely

$$\sum_{k=0}^{MK} q[k] d[m - k] = 0, \quad q[MK] = 1, \quad (9)$$

where the $q[k]$ are the unknown coefficients in the difference equation, *i.e.*, $Q(z) = \sum_{k=0}^{MK} q[k] z^{-k}$ is an annihilator filter for the measurement signal $d[i]$, $(q * d)[i] = 0$ for $i \in \mathcal{I}_s$. Note that hereafter we will denote the coefficients of a polynomial, say $Q(z)$ as a vector using bold lowercase, *i.e.*, \mathbf{q} . To compute the coefficients, c_l , and poles, z_l , we adopt the following steps: first, we determine the annihilating filter; this involves solving a linear system of equations; second, we find the roots of the z -transform of the annihilating filter, which is a nonlinear function; and third, we solve another linear system of equations to determine the weights. We next elaborate upon these three steps.

Finding the filter coefficient $Q(z)$: We can rewrite the Eq. (9) in matrix/vector form as $\mathbf{D}\mathbf{q} = -\mathbf{d}$ where $\mathbf{D} \in \mathbb{C}^{(N_s - MK + 1) \times MK}$ is Hankel matrix, *i.e.* $[\mathbf{D}]_{i,j} = d[j - i]$, and $[\mathbf{d}]_i = d[i + M_p - 1]$

Finding the z_k : Once the vector \mathbf{q} has been computed, the pole locations are estimated as the roots of the polynomial $Q(z) = \sum_{k=0}^{MK} q[k] z^{-k} = 0$, $q[MK] = 1$.

Finding the c_k : The final step is to solve for the vector \mathbf{c} in (8).

Remark 2. Since the data are noisy, the solution to Hankel equations (9) will produce perturbed linear prediction coefficients. The algorithm then finds the poles by rooting the perturbed polynomial. The algorithm uses the perturbed pole locations to generate the Vandermonde system of equations to determine \mathbf{c} . Hence, errors caused by noise in the data propagate and amplify through the algorithm. In prior work, [8], Least- and Total Least-squares methods were employed to improve robustness to noise. The main drawback for the Least-Squares (LS) family of approaches is that they need a large number of measurements ($N_s \gg 2MK$). Increasing the size of the data matrix also increases the complexity of these methods.

3.2. Structural Algorithm for MSML channels Estimation

In this section, we design an algorithm which requires fewer number of measurements, while being robust to noise. In the sequel, we show that the data matrix \mathbf{D} for the case of OFDM transmission over an MSML channel is a *low rank* Hankel matrix with rank M . This structural feature enables data denoising to reduce the perturbation error in low SNR. Furthermore, since the data matrix is low rank, we need much fewer measurements to estimate the channel parameters than the methods

in [6]. In particular, for a data matrix D with rank M , we just need $N_s = MK + M - 1$ measurements to estimate the channel parameters, which is almost half of the minimum number of required measurements for the LS family of methods and other prior methods. As mentioned in Section I, the newly generated frequencies due to scaling effect in MSML channel, namely $\{f_k a_n : \text{for } n \in \{1, 2, \dots, M\}\}$ are close to each other and the corresponding carrier frequency f_k for $k \in \{1, 2, \dots, K\}$. In the following theorem, we show that because of this property, the vector of coefficients associated to polynomial $Q(z)$ has sparsity (compressibility) of order M , which dictates the (near) low rank property to the data matrix D .

Theorem 1. *Suppose that the measurement data follow (5) with defined parameters in (6) and (7). Let us define $P(z) = \prod_{n=1}^M (1 - z_{a_n} z^{-K})$, where $z_{a_n} = e^{j2\pi a_n}$. Then, (a) $Q(z)$ can be approximated by $P(z)$, i.e.*

$$Q(z) \approx \prod_{n=1}^M (1 - z_{a_n} z^{-K}),$$

(b) the approximation error is bounded as,

$$\frac{\|\mathbf{q} - \mathbf{p}\|_2^2}{L} \leq \frac{2c_0\pi^2}{M} \sum_{n=1}^M a_n^2,$$

where $c_0 \leq 1$ is a constant and $L = MK$ is the length vector \mathbf{q} .

The proof of Theorem 1 is in Appendix A. Based on Theorem 1, we see that for data generated by OFDM signaling and passed through MSML channel, $Q(z)$ can be well approximated by a polynomial $P(z)$ with M non-zero coefficients. Since the vector \mathbf{d} is approximated well by M specific columns of D (which is specified by non-zero coefficients in \mathbf{q}), we can conclude that $\text{rank}\{D\} \leq M$. Now, since we have this prior structural information about matrix D that it is a low-rank Hankel matrix, we can use this information to denoise the data matrix. To this end, we seek the best Hankel approximation to D_z (noisy data matrix, i.e., $D_z = D + Z$, where Z is a matrix whose elements are Gaussian noise) with rank M . We can recast this statement to the following matrix approximation problem,

$$\hat{D} = \underset{Y}{\text{argmin}} \|Y - D_z\|_F^2 \text{ s.t. } \text{rank}(Y) = M, \quad (P1)$$

where Y is a Hankel matrix. The structured low rank approximation (SLRA) problem (P_1), which consists of projecting a matrix onto the intersection of a linear subspace and a non-convex manifold, is an NP-hard problem [9]. A possible approach to circumvent the general NP-hardness of low-rank approximation, is to replace the rank by its convex surrogate: the nuclear norm [10]. Thus the SLRA problem in (P_1), is converted to

$$\hat{D} = \underset{Y}{\text{argmin}} \|Y - D_z\|_F^2 + \lambda \|Y\|_*, \quad (P2)$$

where Y is a Hankel matrix and the nuclear norm of a matrix, $\|\cdot\|_*$, is the sum of its singular values. The Lagrangian parameter λ controls the tradeoff between the two terms. This is a convex optimization problem; in particular, it can be recast as a *semidefinite program* (SDP). There are a variety of solution methods for solving the SDP in (3.2) such as the *alternating direction* method and the *augmented Lagrange multiplier* (ALM)

method [10, 11]. In the following, we state a theorem to show that the solution computed by ($P2$) provides a close approximation to the noiseless data matrix.

Theorem 2. *Assume $D_z = D + Z$ where Z is noise matrix. To recover D using ($P2$), for $\lambda \geq 2\|Z\|_2$, we have*

$$\|D^* - D\|_F \leq 64\lambda\sqrt{M} \quad (10)$$

where D^* is the solution found by solving (P_2).

The proof of Theorem 2 is in Appendix B. Based on the derived bound for remaining error in Theorem 2, we can conclude that the denoising algorithm performs quite well. Note that after computing the parameters c_l and z_l by the proposed algorithm, we need just 2 pilots per symbol to compute the channel parameters, $a_n = \frac{\angle z^{(k-1)M+n}}{2\pi f_k} - 1$, $|h_n| = \left| \frac{c^{(k-1)M+n}}{s_k} \right|$, and $\angle h_n - 2\pi f_k \tau_n = \angle \frac{c^{(k-1)M+n}}{s_k}$.

In the following, we state the proposed MSML channel estimation algorithm, which is explained in detail earlier in this section.

Structural MSML channel estimation algorithm:

- | |
|---|
| <p>Step 0: Construct augmented data matrix $D_a = [D, \mathbf{d}]$.</p> <p>Step 1: Perform denoising on D_a by (P_2), using any SDP algorithm, say by ALM [11].</p> <p>Step 2: Construct matrix D_p by eliminating the columns associated with zero coefficients in vector \mathbf{p} from \hat{D}.</p> <p>Step 3: Compute the non-zero elements of \mathbf{p} as,
$\mathbf{p} = -D_p^{-1} \mathbf{d}$.</p> <p>Step 4: Compute the roots of $Q(z) \approx P(z)$, namely z_l.</p> <p>Step 5: Find c_l using (8).</p> <p>Step 6: Compute the channel parameters using (6), (7), and pilot subcarriers information.</p> |
|---|

4. NUMERICAL RESULTS

In our simulations, we consider the OFDM signaling with four subcarriers ($N = 4$), subcarrier space is $\Delta f = 250$ Hz, and carrier frequency $f_c = 10$ kHz. We assume that the number of multipath, M , is equal to 5; and channel parameters are given as, channel gains = [1 0.625 0.455 0.372 0.155], channel delays = [5, 7.5, 10.5, 15.8, 23] msec., and Doppler scales = [1.3, 5.3, 9.7, 16.7, 23.4] $\times 10^{-3}$. In Fig. 1, the performance of our proposed algorithm is compared with the Root-MUSIC [7] and ESPRIT [7] algorithms suggested in [6] for MSML underwater channel estimation and TLS-Prony [8]. We see that our proposed algorithm has better performance. Thus, exploiting the signal structure provides measurably improved performance. Both Root-MUSIC and ESPRIT are implemented based on the methods in [6] which also take advantage of a sparse model, but for the channel versus the effective data matrix [6]. As mentioned in [6], the reason that ESPRIT has worst performance is due to model mismatch and the existence of closely frequencies. To compute the MSE, we consider that $\text{MSE} = \|\hat{\mathbf{h}} - \mathbf{h}\|_2^2$ where $\|\cdot\|_2$ represents l_2 -norm of a vector and element in $[\mathbf{h}]_l = c_l z_l$ with $s_k = 1$ and indexing is similar to (5) and (6). As observed from Fig. 1, our methods achieve the same MSE with 5 to 10 dB less SNR than the methods of [6].

5. CONCLUSIONS

In this paper, we considered the estimation of multi-scale, multi-lag channels. We adapted a spectral line estimation approach

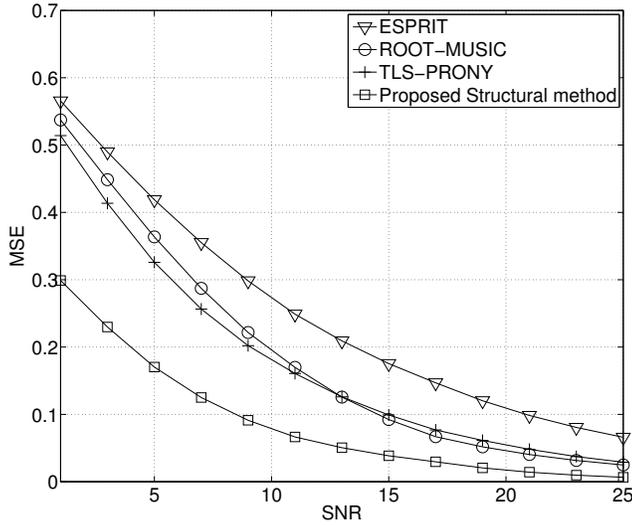


Fig. 1: Mean-square-error (MSE) versus signal-to-noise ratio (SNR) comparison.

to estimate the MSML channel, based on the path-based model. Taking advantage of the low rank structure of the data matrix, our proposed method is robust to noise and also requires a low number of measurements to estimate the closely spaced frequencies with high resolution. Finally, in the simulation results, we show that the proposed algorithm provides the performance (in SNR sense) with a 5 to 10 dB improvement compared to prior approaches.

A. PROOF OF THEOREM 1

(a). We start the proof by definition of $Q(z)$. We know that

$$Q(z) = \sum_{l=0}^{MK} q[l]z^{-l} = \prod_{l=1}^{MK} (1 - z_l z^{-1}), \quad (11)$$

where z_l are defined in (7) as $z_{(k-1)M+n} = e^{j2\pi f_k} e^{j2\pi f_k a_n}$. Lets rewrite the product in (11) as following:

$$Q(z) = \prod_{n=1}^M \prod_{k=1}^K (1 - z_{(k-1)M+n} z^{-1}). \quad (12)$$

We know that $\hat{f}_k = \frac{k}{K} \leq 1$ and $a_n \ll 1$ ($|a_n| < 10^{-2}$), therefore it is reasonable to say that $\hat{f}_k a_n = k \frac{a_n}{K} \approx \frac{a_n}{K}$. Thus we can conclude that $z_{(k-1)M+n} \approx e^{j2\pi \hat{f}_k} e^{j2\pi \frac{a_n}{K}}$. Then plugging this value in (12), we have

$$Q(z) \approx \prod_{n=1}^M \prod_{k=1}^K (1 - e^{j2\pi \hat{f}_k} z_{a_n} z^{-1}). \quad (13)$$

Let us define $f(z) = \prod_{k=1}^K (1 - e^{j2\pi \hat{f}_k} z^{-1})$. Since $e^{-j2\pi \hat{f}_k}$ are zeros of unity for $1 \leq k \leq K$, we can conclude that $f(z) = 1 - z^{-K}$, then we have $\prod_{k=1}^K (1 - e^{j2\pi \hat{f}_k} z_{a_n} z^{-1}) = f\left(\frac{z}{z_{a_n}}\right) = 1 - z_{a_n}^K z^{-K}$. if we plug this value in (13), we

¹we recognize that this approximation may seem loose; however, the resulting sparse structure significantly increases the robustness of our algorithm.

will obtain the desire result for (a).

(b) We first provide a lemma regarding the relationship between the perturbation of a polynomial root and its coefficients perturbation.

Lemma 1. (see [12]) Suppose that z_l is the a root with multiplicity α_l of polynomial $p(z)$, i.e. $p(z) = (z - z_l)^{\alpha_l} \hat{p}(z)$ where $\hat{p}(z_l) \neq 0$. Furthermore, consider that $p'(z) = (z - z'_l)^{\alpha_l} \hat{p}(z)$ where $z'_l = z_l + \delta z_l$. Then, $|\delta z_l| \approx \|\mathbf{p} - \mathbf{p}'\|_2^{\frac{1}{\alpha_l}}$, where $|\delta z_l|$ is the absolute value of δz_l and \mathbf{p} and \mathbf{p}' are the coefficient vectors associated with polynomials $p(z)$ and $p'(z)$, respectively.

Using Lemma 1, we know that if we change the root $z_l = e^{j2\pi \hat{f}_k} e^{j2\pi \frac{k}{K} a_n}$ to $z'_l = e^{j2\pi \hat{f}_k} e^{j2\pi \frac{a_n}{K}}$, where $l = (k-1)M + n$, and design new polynomial $Q_l(z) = (1 - z'_l z^{-1}) \frac{Q_{l-1}(z)}{(1 - z_l z^{-1})}$, then we have

$$\begin{aligned} \|\mathbf{q}_l - \mathbf{q}_{l-1}\|_2^2 &\approx |z_l - z'_l|^2 = \left| e^{j2\pi \frac{k}{K} a_n} - e^{j2\pi \frac{a_n}{K}} \right|^2 \\ &\approx \left| (1 + j2\pi \frac{k}{K} a_n) - (1 + j2\pi \frac{a_n}{K}) \right|^2 = 4\pi^2 a_n^2 \frac{k^2 - 1}{K^2} \end{aligned}$$

We assume that all roots are perturbed similarly, thus we can bound the difference between the coefficient in the resultant polynomial and the original polynomial as,

$$\|\mathbf{q}_0 - \mathbf{q}_{MK}\|_2^2 = \left\| \sum_{l=0}^{MK-1} (\mathbf{q}_l - \mathbf{q}_{l+1}) \right\|_2^2 \leq \sum_{l=0}^{MK-1} \|(\mathbf{q}_l - \mathbf{q}_{l+1})\|_2^2$$

where $\mathbf{q}_0 = \mathbf{q}$ and $\mathbf{q}_{MK} = \mathbf{p}$. Therefore, we can conclude that

$$\frac{1}{L} \|\mathbf{q} - \mathbf{p}\|_2^2 \leq \sum_{n=1}^M \sum_{k=1}^K 4\pi^2 a_n^2 \frac{k^2 - 1}{MK^3} = c_0 \left(\frac{4\pi^2}{M} \sum_{n=1}^M a_n^2 \right)$$

where $c_0 = \frac{K+3}{3K} < 1$ for $K \geq 2$.

B. PROOF OF THEOREM 2

Assume \mathbf{D}^* is a feasible solution for (P2) and $\hat{\mathbf{D}}$ is the optimal solution of (P2), then we have,

$$\|\hat{\mathbf{D}} - \mathbf{D}_z\|_F^2 + \lambda \|\hat{\mathbf{D}}\|_* \leq \|\mathbf{D}^* - \mathbf{D}_z\|_F^2 + \lambda \|\mathbf{D}^*\|_*,$$

which can be rewritten as, $\|\hat{\mathbf{D}} - \mathbf{D}_z\|_F^2 - \|\mathbf{D}^* - \mathbf{D}_z\|_F^2 \leq \lambda \left\{ \|\mathbf{D}^*\|_* - \|\hat{\mathbf{D}}\|_* \right\}$. Let us define $\Delta = \mathbf{D}^* - \hat{\mathbf{D}}$. Now performing some algebra yields the inequality

$$\|\Delta\|_F^2 \leq 2|\langle \text{vec}\{\mathbf{Z}\}, \text{vec}\{\Delta\} \rangle| + \lambda \left\{ \|\hat{\mathbf{D}} + \Delta\|_* - \|\hat{\mathbf{D}}\|_* \right\},$$

where $\text{vec}\{\cdot\}$ represents the vector form of a matrix, and $\langle \cdot, \cdot \rangle$ denotes the vectors inner product (dot product). We know that $|\langle \text{vec}\{\mathbf{Z}\}, \text{vec}\{\Delta\} \rangle| = |\text{trace}\{\mathbf{Z}^H \Delta\}|$. Now using the Holder inequality for matrix product [14] and norm inequalities, i.e. $\|A\|_2 \leq \|A\|_F \leq \sqrt{r} \|A\|_2$ where r is the rank of matrix A , we have $\|\Delta\|_F^2 \leq 2\|\mathbf{Z}\|_2 \|\Delta\|_* + \lambda \|\Delta\|_*$. By our choice of λ in the Theorem, we have $\|\Delta\|_F^2 \leq 2\lambda \|\Delta\|_*$. Now we need to find a maximum bound for the right hand side, namely $\|\Delta\|_*$ using following lemma.

Lemma 2. (see, e.g., [13]) There exists a matrix decomposition for error matrix Δ as $\Delta = \Delta_1 + \Delta_2$ such that (i) $\text{rank}(\Delta_1) \leq 2M$ and (ii) $\|\Delta_2\|_* \leq 3\|\Delta_1\|_*$.

Using Lemma 2 and norm inequality we have, (i) $\|\Delta_1\|_* \leq \sqrt{2M} \|\Delta_1\|_F \leq 2\sqrt{M} \|\Delta\|_F$ and (ii) $\|\Delta\|_* \leq \|\Delta_1\|_* + \|\Delta_2\|_* \leq 4\|\Delta_1\|_*$. Putting together these pieces, we have $\|\Delta\|_F \leq 64\lambda\sqrt{M}$.

C. REFERENCES

- [1] J.-J. Van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1800–1805, 1997.
- [2] H. Arslan *et al.*, "Channel estimation for wireless OFDM systems," *IEEE Surveys and Tutorials*, vol. 9, no. 2, pp. 18–48, 2007.
- [3] S. Yerramalli and U. Mitra, "Optimal resampling of OFDM signals for multiscale-multilag underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 36, no. 1, pp. 126–138, 2011.
- [4] S. Beygi and U. Mitra, "Optimal bayesian resampling for OFDM signaling over multi-scale multi-lag channels," *IEEE Signal Process. Letters.*, vol. 20, no. 11, pp. 1118–1121, 2013.
- [5] Y. Jiang and A. Papandreou-Suppappola, "Discrete time-scale characterization of wideband time-varying systems," *Signal Processing, IEEE Transactions on*, vol. 54, no. 4, pp. 1364–1375, 2006.
- [6] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, 2010.
- [7] H. Van Trees, *Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory, (Part IV), 1st ed. New York: Wiley, 2002.
- [8] W. M. Steedly, C.-H. J. Ying, and R. L. Moses, "Statistical analysis of TLS-based prony techniques," *Automatica*, vol. 30, no. 1, pp. 115–129, 1994.
- [9] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is NP-hard," *SIAM J. Matrix Anal. Applicat.*, vol. 32, no. 4, pp. 1149–1165, 2011.
- [10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.
- [11] M. Fazel, T. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications in system identification and realization," *SIAM J. Matrix Anal. Applicat.*, vol. 2, pp. 123–144, 2012.
- [12] J. H. Wilkinson, *Rounding errors in algebraic processes*, Dover, New York, 1994.
- [13] D. S. Bernstein, "Matrix mathematics," 2005.
- [14] H. Albadawi, "Matrix inequalities related to holder inequality," *Banach J. Math. Analysis*, vol. 7, pp. 162–171, 2013.