Elvander, Filip

Estimating Inharmonic Signals with Optimal Transport Priors

Published in:
ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

DOI:
10.1109/ICASSP49357.2023.10095082

Published: 10/06/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Elvander, F. (2023). Estimating Inharmonic Signals with Optimal Transport Priors. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). Article 10095082 (Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095082

ESTIMATING INHARMONIC SIGNALS WITH OPTIMAL TRANSPORT PRIORS

Filip Elvander

Dept. of Information and Communications Engineering
Aalto University, Finland
email: filip.elvander@aalto.fi

ABSTRACT

In this work, we consider the problem of estimating the frequency content of inharmonic signals, i.e., sinusoidal mixtures whose components are close to forming a harmonic set. Intuitively, exploiting this closeness should lead to increased estimation performance as compared to unstructured estimation. Earlier approaches to this problem have relied on parametric descriptions of the inharmonicity, stochastic representations, or have resorted to misspecified estimation by ignoring the inharmonicity. Herein, we propose to use a penalized maximum-likelihood framework, where the regularizer is constructed based on optimal mass transport theory, promoting estimates that are close-to-harmonic in a spectral sense. This leads to an estimator that forms a smooth path between the unstructured maximum-likelihood estimator (MLE) and a misspecified MLE (MMLE), as determined by a regularization parameter. In numerical illustrations, we show that the proposed estimator worst-case dominates the MLE and MMLE, thereby allowing for robust estimation for cases when the inharmonicity level is unknown.

***Index Terms*—** Inharmonicity, spectral estimation, frequency estimation, optimal mass transport

1. INTRODUCTION

Signals that may be well-modeled as sinousidal mixtures whose frequencies obey a harmonic relationship arise in a multitude of signal processing applications, ranging from speech processing [1] to fault detection in industrial machinery [2] and monitoring of power grids [3]. In these applications, the fundamental frequency constitutes the main feature of interest and considerable work has been directed towards forming statistically and computationally efficient estimators of it, see, e.g., [4, 5]. However, in some situations the harmonic relationship between the sinusoidal components is not exact, i.e., the frequencies display small deviations from being perfect integer multiples of a fundamental frequency. This is for example the case for the sound produced by string instruments [6], as well as to some degree for the human voice [7]. Such close-to-harmonic signals are often referred to as being inharmonic. The estimation of the fundamental frequency for inharmonic signals has been considered for parametric models describing string instruments [8] as well as for unstructured inharmonicity [9–11]. Furthermore, in our recent work [12], we propose mathematically rigorous definitions of fundamental frequency for inharmonic signals, which in a strict sense lack such a concept, and explore the estimation theoretical consequences of these definitions. Common for the estimators proposed in these works is that they seek to estimate a single frequency, i.e., the nominal fundamental frequency, from noisy observations of sinusoidal mixtures. However, less attention is given to the frequencies of the individual components: the estimates of these are simply integer multiples of the fundamental frequency, unless a complete parametric description of the inharmonicity is available [8]. For stochastic inharmonicity with known distribution, an estimator allowing for forming estimates of the sinusoidal frequencies was proposed in [12], whereas, to the best of the author's knowledge, there are no widely available estimators for the case of deterministic but unstructured inharmonicity.

In this work, we propose precisely that: an estimator applicable to inharmonic signals wherein the inharmonicity is seen as deterministic but without (known) structure. Intuitively, estimating the frequency content of such signals should be easier, in an estimation theoretical sense, as compared to the case of estimating completely unrelated sinusoids. In order to leverage this close-to-harmonic property without relying on explicit models, one needs a reliable way of quantifying the closeness. Herein, we propose to do this by measuring distances between harmonic and inharmonic signals using the Monge-Kantorovich problem of optimal mass transport (OMT) [13]. OMT has earlier been used for introducing weakly continuous metrics on the space of power spectra [14] as well as in tracking, interpolation, and clustering problems [15–19]. In this work, we use a spectral OMT distance in order to construct a regularized maximum-likelihood estimator (MLE) as to promote estimates that are close-to-harmonic. As is shown, the proposed estimator allows for constructing a smooth path, as determined by a regularization parameter, between the MLE of an unstructured sinusoidal model and the misspecified MLE (MMLE) formed under the (erroneous) assumption of a perfectly harmonic signal. In numerical simulations, we show that the proposed estimator provides a robust alternative to the MLE and MMLE, allowing for smoothly incorporating varying knowledge of the size of the inharmonicity.

2. SIGNAL MODEL

Consider the signal[1]

$$y_t = x_t + v_t = \sum_{k=1}^{K} \alpha_k e^{i\omega_k t} + v_t, \qquad (1)$$

for $t = 1, 2, \ldots, N$, with $N \in \mathbb{N}$, where $\alpha_k \in \mathbb{C}$, for $k = 1, \ldots, K$, are the complex amplitudes and $\omega_k \in \mathbb{T} \triangleq [0, 2\pi)$ are the frequencies. Furthermore, let v_t be well-modeled as circularly symmetric Gaussian white noise with variance σ^2. We will here assume that the noise-free waveform x_t in (1) is an inharmonic signal. Specifically, this means that the frequencies of the sinusoidal components satisfy

$$\omega_k = k\omega_0 + \Delta_k \ , \ k = 1, \ldots, K, \qquad (2)$$

for some $\omega_0 \in \mathbb{T}$. Here, $\Delta_k \in \mathbb{R}$ are referred to as inharmonicity parameters and are assumed to be small in the sense that $|\Delta_k| \ll \omega_0$. Even though x_t is not harmonic, ω_0 is often referred to as the fundamental frequency. It should here be noted that given x_t, it is not self-evident what value ω_0 should take as the relation (2) offers considerable freedom in choosing ω_0 and Δ_k. In our recent work [12], we addressed this ambiguity and offered three different definitions of fundamental frequency for an inharmonic signal, emanating from three different views of the model in (1): one viewing y_t as a misspecified harmonic model; one seeing it as a perturbed harmonic spectrum; and one viewing the inharmonicity parameters as zero-mean stochastic variables. All three definitions allow for constructing information theoretical lower bounds for estimates of ω_0 as well as for formulating optimal estimators of ω_0. However, only the stochastic view-point led to an estimator of also the inharmonicity parameters Δ_k. That is, no estimator able to leverage the close-to-harmonic structure when estimating the frequency vector $\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & \ldots & \omega_K \end{bmatrix}^T$, as opposed to only the fundamental frequency ω_0, was offered for the case when Δ_k were deterministic.

Herein, we propose an estimator of the unknown (deterministic) parameter $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\omega}^T & \boldsymbol{\alpha}^T \end{bmatrix}^T$, where $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_K \end{bmatrix}^T$, that is able to use the information that the frequencies of vector $\boldsymbol{\omega}$ almost, but not quite, obeys a harmonic relation. In particular, we will here consider an estimator formulated as a penalized MLE, where the regularizer is constructed based on the optimal mass transport (OMT) induced definition of fundamental frequency from [12].

[1] Here, for generality, we will consider the complex-valued representation, noting that this can easily be formed as the discrete-time analytical version of a real-valued signal [20].

3. OPTIMAL MASS TRANSPORT

For the waveform x_t in (1), the spectral representation[2] is

$$\Phi_{\boldsymbol{\theta}}(\omega) = 2\pi \sum_{k=1}^{K} |\alpha_k|^2 \, \delta(\omega - \omega_k), \qquad (3)$$

where δ is the Dirac delta function. Here, the notation $\Phi_{\boldsymbol{\theta}}$ emphasizes that the spectrum of x_t is parametrized by $\boldsymbol{\theta}$. In [12], we proposed a definition of fundamental frequency ω_0 based on this spectral representation codifying the intuitive idea that $\Phi_{\boldsymbol{\theta}}$ could be seen as a perturbation (in frequency) of a perfectly harmonic spectrum Φ_μ, i.e.,

$$\Phi_\mu(\omega) = 2\pi \sum_{\ell=1}^{L} \gamma_\ell \delta(\omega - \ell\omega_0), \qquad (4)$$

where ω_0 is the fundamental frequency and $\gamma_\ell > 0$ are the powers of the harmonics, and with L not necessarily equal to K. The closest harmonic spectrum Φ_μ, or equivalently the values of ω_0 and γ_ℓ, were then determined as the one minimizing the spectral OMT distance $S(\Phi_{\boldsymbol{\theta}}, \Phi_\mu)$ with respect to Φ_μ over the set of harmonic spectra. Here, S is defined via the Monge-Kantorovich problem of OMT [13] according to

$$S(\Phi_{\boldsymbol{\theta}}, \Phi_\mu) \triangleq \min_{M \in \Omega(\Phi_{\boldsymbol{\theta}}, \Phi_\mu)} \int_{\mathbb{T}^2} |\omega_1 - \omega_2|^2 \, dM(\omega_1, \omega_2), \qquad (5)$$

where $\mathbb{T}^2 \triangleq \mathbb{T} \times \mathbb{T}$, and with

$$\Omega(\Phi_{\boldsymbol{\theta}}, \Phi_\mu) \triangleq \Big\{ M \in \mathcal{M}_+(\mathbb{T}^2) \mid \int_{\mathbb{T}} dM(\cdot, \omega_2) = \Phi_{\boldsymbol{\theta}} \ , \int_{\mathbb{T}} dM(\omega_1, \cdot) = \Phi_\mu \Big\},$$

where $\mathcal{M}_+(\mathbb{T}^2)$ is the set of non-negative measures on $\mathbb{T}^2$. It may here be noted that $\Omega(\Phi_{\boldsymbol{\theta}}, \Phi_\mu)$ contains all measures M whose so-called margins coincide with $\Phi_{\boldsymbol{\theta}}$ and Φ_μ. Elements of $\Omega(\Phi_{\boldsymbol{\theta}}, \Phi_\mu)$ are referred to as transport plans, as $dM(\omega_1, \omega_2)$ is the amount of power, or mass, that is transported from ω_1 to ω_2. Then, $S(\Phi_{\boldsymbol{\theta}}, \Phi_\mu)$ corresponds to the minimal possible cost of rearranging the power in $\Phi_{\boldsymbol{\theta}}$ as to match that of Φ_μ, where the cost of moving one unit of power from ω_1 to ω_2 is defined as $|\omega_1 - \omega_2|^2$. Minimizing S over the set of harmonic spectra then yields the harmonic spectrum Φ_μ closest to $\Phi_{\boldsymbol{\theta}}$ in the OMT sense. It should here be noted that S defines a metric on the space of power spectra [14]. Whereas such minimizers were used in [12] as to define a concept of fundamental frequency ω_0 for inharmonic signals, we will here use them as regularizers in inference. In particular, we will here consider the regularizing function

$$\Psi(\boldsymbol{\theta}) \triangleq \min_{\Phi_\mu} S(\Phi_{\boldsymbol{\theta}}, \Phi_\mu), \qquad (6)$$

where the minimization is understood as being over harmonic spectra of the form (4).

[2] This corresponds to letting the initial phases be uniform random variables on $[-\pi, \pi)$ and seeing them as nuisance parameters in estimation.

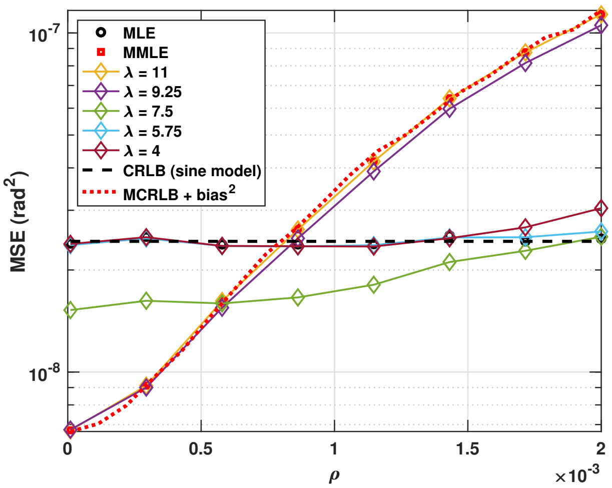

Fig. 1. MSE as a function of the inharmonicity level ρ, with SNR 10 dB, for varying values of the parameter λ.

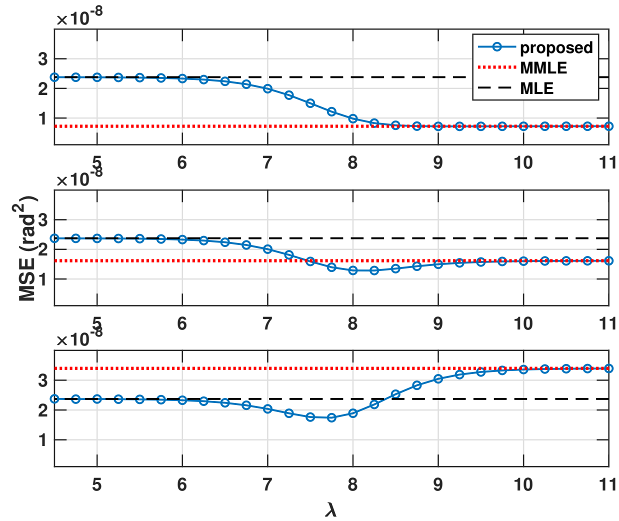

Fig. 2. MSE as a function of the regularization parameter λ for SNR 10 dB for three different levels of inharmonicity. Top: $\rho = 2\times10^{-4}$. Middle: $\rho = 6\times10^{-4}$. Bottom: $\rho = 10^{-3}$.

4. PROPOSED ESTIMATOR

For the signal in (1), the log-likelihood of the sample $\mathbf{y} = \begin{bmatrix} y_1 & \ldots & y_N \end{bmatrix}^T$, excluding additive constants, is[3]

$$\log p(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) = -N \log \sigma^2 - \frac{1}{\sigma^2} \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\alpha} \right\|_2^2, \quad (7)$$

where p is the (Gaussian) probability density function of $\mathbf{y}$, with $\mathbf{A}(\boldsymbol{\omega})$ denoting the dictionary matrix constructed from the Fourier basis vectors corresponding to the frequencies $\boldsymbol{\omega}$. Herein, we propose to form estimates of $\boldsymbol{\theta}$ and the unknown noise variance σ^2 as the solution to

$$\underset{\boldsymbol{\theta},\sigma^2}{\arg\min}\ \mathcal{L}(\boldsymbol{\theta}, \sigma^2) = -\log p(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) + 10^{\lambda} \Psi(\boldsymbol{\theta}),$$

where $\lambda \in \mathbb{R}$ is the regularization parameter[4] determining the trade-off between data-fit and the reliance on the OMT prior. Clearly, for any $\boldsymbol{\theta}$, $\mathcal{L}$ is minimized with respect to σ^2 by $\sigma^2 = \frac{1}{N} \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\alpha} \right\|_2^2$, implying that minimizers $\boldsymbol{\theta}$ can be found as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\min}\ N \log \left(\left\| \mathbf{y} - \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\alpha} \right\|_2^2 \right) + 10^{\lambda} \Psi(\boldsymbol{\theta}). \quad (8)$$

From this, we can deduce a connection between $\hat{\boldsymbol{\theta}}$ and the MLE and MMLE, depending on the value of λ. These estimators are given as (see [12] for the MMLE)

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \underset{\boldsymbol{\theta}}{\arg\min}\ \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\alpha} \right\|_2^2, \quad (9)$$

and

$$\hat{\boldsymbol{\theta}}_{\text{MMLE}} = \underset{\boldsymbol{\theta}}{\arg\min}\ \left\| \mathbf{y} - \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\alpha} \right\|_2^2, \text{ s.t. } \omega_k = k\omega_1\ \forall k, \quad (10)$$

[3]This may be extended to the non-white case by incorporating a weighting with the noise covariance matrix or an estimate thereof, see, e.g., [21].

[4]The representation 10^{λ} is utilized for scaling convenience, as shown in the numerical section.

respectively. It may here be noted that the MMLE is the MLE formed under the assumption that the signal is perfectly harmonic. In [12], it was shown that for small inharmonic perturbations, Ψ is locally smooth in a neighbourhood of its minimizer, meaning that (8) is locally smooth. Furthermore, Ψ is only zero on the set of harmonic spectra. Noting that S, or technically $S^{1/2}$, in (5) is a metric on the set of spectra, we have that $\hat{\boldsymbol{\theta}}$ tends to $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$ and $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ as $\lambda \to \pm\infty$, respectively. From the smoothness of Ψ, we expect this convergence behavior to be smooth. However, it should be noted that this does not imply that the performance of the estimator in (8) interpolates that of $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ and $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$ in a monotone fashion. Instead, what we observe in practice is that for any λ, the performance of $\hat{\boldsymbol{\theta}}$, in terms of mean-squared-error (MSE), is at least as good as the worst of $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ and $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$. That is, the proposed estimator offers a robust alternative to the MLE and MMLE for situations in which it is not known which of the two is expected to perform best (c.f., [22]). This empirical observation is demonstrated in the numerical section.

Remark 1. *Apart from the intuitive appeal of regularizing the MLE with the OMT cost Ψ and thereby seeking signals being close to harmonic in the sense of small frequency perturbations, the specific choice has an additional motivation based on estimation theory. Specifically, it was in [12] shown that for small inharmonicity parameters, the OMT cost function in* (6) *can be written as*

$$\Psi(\boldsymbol{\theta}) = \min_{\omega_0} 2\pi \sum_{k=1}^{K} \left| \alpha_k \right|^2 \left(\omega_k - k\omega_0 \right)^2, \quad (11)$$

where ω_0 was the corresponding definition of fundamental frequency. Furthermore, for the MLE of the unstructured model, the frequency estimates are asymptotically indepen-

dent of the amplitude estimates, with the variance of estimates of ω_k being proportional to $1/|\alpha_k|^2$ (see, e.g., [23]). Thus, each deviation $\omega_k - k\omega_0$ is weighted by (an estimate of) its inverse (MLE) variance. This principle has also been used for estimating the fundamental frequency using the EXIP framework [24] for the case of a perfectly harmonic signal [25].

It may be noted that the optimization criterion in (8) defining the proposed estimator is non-convex in $\boldsymbol{\theta}$. Thus, a non-linear search requires a good initial point in order to find the global minimum. Due to the connection to the MLE and MMLE, two possible candidates are given by $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ and $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$. We here propose to use $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$ due to its considerably lower variance, and due to the fact that it can be found by a one-dimensional search [12]. Also, for small inharmonicity parameters, the OMT cost in (11) is explicitly given by

$$\Psi(\boldsymbol{\theta}) = 2\pi \sum_{k=1}^{K} |\alpha_k|^2 \left(\omega_k - k \frac{\sum_{\ell=1}^{K} |\alpha_\ell|^2 \ell \omega_\ell}{\sum_{\ell=1}^{K} |\alpha_\ell|^2 \ell^2} \right)^2,$$

meaning that the estimator under this assumption can be computed using, e.g., a Newton scheme.

5. NUMERICAL EXAMPLES

Herein, we illustrate the behavior of the proposed estimator, and in particular its connection to the MLE and the MMLE. This is performed using a simulation setting in which we vary the signal to noise ratio (SNR) and the inharmonicity level, i.e., the size of Δ_k. Throughout, consider a signal with harmonic order $K = 5$, where the fourth component is missing, i.e., $\alpha_4 = 0$, and the remaining amplitudes are $|\alpha_k| = e^{-0.2(k-K/2)^2}$ with uniform random phase. The SNR (in dB) is here defined as SNR $= 10\log_{10} \frac{\sum_k |\alpha_k|^2}{\sigma^2}$, where σ^2 is the variance of the additive noise. Letting $\omega_0 = \pi/10$ be the nominal fundamental frequency, we in each simulation draw $\Delta_k \sim \mathcal{U}(\rho\omega_0[-1,1])$, for $k = 1,\ldots,K$, independently, where $\mathcal{U}$ is the uniform distribution and where the factor $\rho : 0 \leq \rho \ll 1$ controls the level of inharmonicity. For each value of ρ and SNR, we perform 1000 Monte Carlo simulations in which we collect $N = 500$ samples of the signal and compute the mean-squared-error (MSE) for the estimates of the frequencies ω_k, averaged over the four components. This is done for the proposed estimator $\hat{\boldsymbol{\theta}}$ in (8) for varying values of the regularization parameter λ, as well as for the MLE $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ and MMLE $\hat{\boldsymbol{\theta}}_{\text{MMLE}}$ in (9) and (10), respectively.

For a fixed SNR of 10 dB, Figure 1 shows the MSE for varying values of the inharmonicity level ρ. As can be seen, the performance of the MLE does not depend on the inharmonicity level, as the components are well-separated. In contrast, the MMLE, which is formed under the assumption of a perfectly harmonic signal, has an increasing MSE as a function of ρ, mainly due to increasing bias. As reference, we also include the Cramér-Rao lower bound (CRLB) [23] for the unstructured signal as well as the theoretical MSE of the MMLE incorporating the misspecified CRLB (MCRLB) and the squared bias (see [12] and [26, 27]). Furthermore, it may be noted that the MSE of the proposed estimator for all values[5] of λ is bounded from above by the maximum of the MSEs corresponding to the MLE and MMLE. Note also that for some levels of inharmonicity, there are values of λ such that the proposed estimator outperforms both the MLE and MMLE. This is highlighted in Figure 2, showing corresponding results for three different values of ρ and a finer grid of the regularization parameter λ. As can be seen, for the smallest ρ, the performance of the proposed estimator lies between that of the MLE and MMLE, whereas for the two larger values it outperforms the benchmarks for some values of the regularization parameter (note that the MSE is not monotone in λ). For both cases, it may be noted that as $\lambda \to \pm\infty$, the MSE smoothly tends to that of the MMLE and MLE, respectively. Figure 3 shows corresponding results for fixed inharmonicity level $\rho = 5 \times 10^{-4}$ and varying SNR. As can be seen, as the SNR increases, the MMLE is outperformed by the MLE as the former is biased. It may here be noted that the proposed estimator worst-case dominates the MLE and MMLE for all values of λ. Note that for SNR -10 dB the MLE and the proposed estimator with $\lambda = 4$ fail to provide accurate estimates.

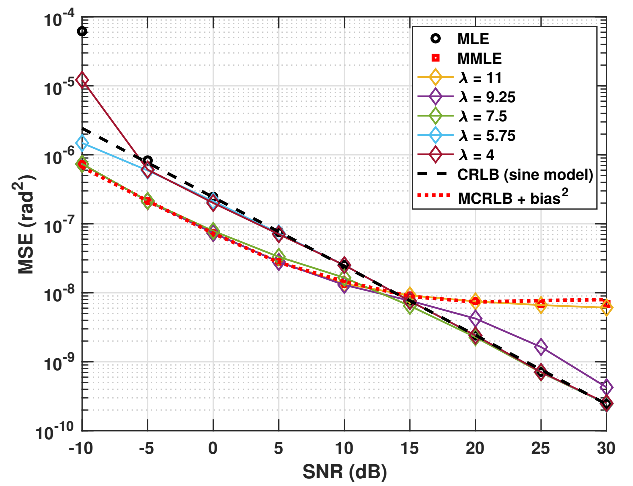

Fig. 3. MSE as a function of the SNR, with inharmonicity level $\rho = 5 \times 10^{-4}$, for varying values of the parameter λ.

6. CONCLUSIONS

We have herein presented a parameter estimator for inharmonic signals. As demonstrated, the proposed estimator allows for incorporating the information of close-to-harmonic components in a smooth manner, and is robust in the sense of not performing worse than the worst of the MLE for unstructured sinusoidal mixtures and the MMLE formed under a perfectly harmonic assumption, in terms of MSE.

[5] For values of λ outside this range, the proposed estimator numerically starts to coincide with the MLE or MMLE.

7. REFERENCES

[1] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Instantaneous Fundamental Frequency Estimation With Optimal Segmentation for Nonstationary Voiced Speech," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 12, pp. 2354–2367, 2016.

[2] R. B. Randall, *Vibration-Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, John Wiley & Sons, Chichester, UK, 2011.

[3] J. Li, Z. Teng, Y. Wang, S. You, Y. Liu, and W. Yao, "A fast power grid frequency estimation approach using frequency-shift filtering," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2461 – 2464, 2019.

[4] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient," *Elsevier Signal Processing*, vol. 135, pp. 188–197, Jan 2017.

[5] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, San Rafael, Calif., 2009.

[6] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer-Verlag, New York, NY, 1988.

[7] E. B. George and M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 389–406, 1997.

[8] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "A Robust and Computationally Efficient Subspace-Based Fundamental Frequency Estimator," *IEEE Trans. Audio, Speech, Language Process*, vol. 18, no. 3, pp. 487–497, March 2010.

[9] F. Elvander, S. I. Adalbjörnsson, J. Karlsson, and A. Jakobsson, "Using Optimal Transport for Estimating Inharmonic Pitch Signals," in *Proc. 42nd IEEE Int. Conf. on Acoust. Speech, Signal Process.*, New Orleans, LA, USA, Mar 2017, pp. 331–335.

[10] R. Flamary, C. Févotte, N. Courty, and V. Emiya, "Optimal spectral transportation with application to music transcription," in *Advances in Neural Information Processing Systems 29*, 2016.

[11] N. R. Butt, S. I. Adalbjörnsson, S. D. Somasundaram, and A. Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities," in *38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[12] F. Elvander and A. Jakobsson, "Defining Fundamental Frequency for Almost Harmonic Signals," *IEEE Trans. Signal Process.*, vol. 68, pp. 6453–6466, 2020.

[13] C. Villani, *Optimal transport: old and new*, Springer Science & Business Media, 2008.

[14] T. T. Georgiou, J. Karlsson, and M. S. Takyar, "Metrics for power spectra: an axiomatic approach," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 859–867, Mar. 2009.

[15] F. Elvander, A. Jakobsson, and J. Karlsson, "Interpolation and Extrapolation of Toeplitz Matrices via Optimal Mass Transport," *IEEE Trans. Signal. Process*, vol. 66, no. 20, pp. 5285 – 5298, Oct. 2018.

[16] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal Transport Over a Linear Dynamical System," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2137–2152, May 2017.

[17] F. Elvander, I. Haasler, A. Jakobsson, and J. Karlsson, "Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion," *Signal Process.*, vol. 171, June 2020, Art. no. 107474.

[18] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal Mass Transport: Signal processing and machine-learning applications," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 43–59, July 2017.

[19] F. Elvander, J. Karlsson, and T. van Waterschoot, "Convex Clustering for Multistatic Active Sensing via Optimal Mass Transport," in *29th European Signal Processing Conference*, 2021.

[20] S. L. Marple, "Computing the discrete-time "analytic" signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, September 1999.

[21] A. E. Jaramillo, A. Jakobsson, J. K. Nielsen, and M. G. Christensen, "Robust Fundamental Frequency Estimation in Coloured Noise," in *45th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2019.

[22] F. Elvander, J. Ding, and A. Jakobsson, "On Harmonic Approximations of Inharmonic Signals," in *45th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*. 2019.

[23] P. Stoica, A. Jakobsson, and J. Li, "Cisoid Parameter Estimation in the Colored Noise Case: Asymptotic Cramér-Rao Bound, Maximum Likelihood and Nonlinear Least-Squares," *IEEE Trans. Signal Process.*, vol. 45, pp. 2048–2059, August 1997.

[24] P. Stoica and T. Soderstrom, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, pp. 383–387, 1989.

[25] H. Li, P. Stoica, and J. Li, "Computationally Efficient Parameter Estimation for Harmonic Sinusoidal Signals," *Signal Processing*, vol. 80, pp. 1937–1944, 2000.

[26] C. D. Richmond and L. L. Horowitz, "Parameter Bounds on Estimation Accuracy Under Model Misspecification," *IEEE Trans. Signal. Process*, vol. 63, no. 9, pp. 2263–2278, 2015.

[27] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 142–157, Nov 2017.