EXACT ACOUSTIC SIGNAL RECONSTRUCTION FROM STFT MAGNITUDE WITH SMALL HOP SIZE

Yuliang Ji* Jian Wu† Yuanzhe Xi‡

* Individual Contributor
† School of Engineering, Tokyo Institute of Technology
‡Department of Mathematics, Emory University

ABSTRACT

Reconstructing acoustic signals from the Short-Time Fourier Transform (STFT) magnitude spectrogram is a pivotal challenge in acoustic signal analysis. While various techniques utilizing neural networks and other approximation methods have emerged recently to translate the STFT magnitude spectrogram of an audio signal into waveforms, the intricate mathematical structures inherent in the problem have often been overlooked. In this paper, we introduce a polynomial-time deterministic algorithm for the reconstruction of acoustic signals from their STFT magnitude when the hop size is small, and under certain conditions. Furthermore, we extend our methodology to another variant of STFT magnitude, offering another deterministic polynomial-time algorithm. The efficacy of our proposed techniques is underpinned by both theoretical analysis and empirical experiments.

Index Terms— STFT magnitude, signal reconstruction.

1. INTRODUCTION

Recovering acoustic signals from spectrograms is a central challenge in acoustic signal processing. This problem, also termed as 'phase retrieval' or 'phase reconstruction' [1], utilizes methods commonly known as vocoders to transform the spectrogram back into an audio signal [2]. Vocoders find extensive applications in areas such as Speech Synthesis [3, 4] and Audio Generation [5].

Introduced in 1984, the renowned Griffin-Lim algorithm [6] aims to approximate the original signal from its STFT spectrogram using a mean-square-error approach. Subsequent research has refined this algorithm, optimizing both its speed [7, 8, 9] and accuracy [10].

In the recent era, the prevailing techniques for acoustic signal recovery from spectrograms have predominantly leaned on convex programming [11, 12, 13, 14] and neural networks [15, 16], or a fusion of both [17]. Yet, a majority of these techniques are approximative in nature, implying they often fall short of precisely reconstructing the acoustic signal. Additionally, many neural network-based methodologies predominantly utilize the mel spectrogram as input [3, 4]. Dimensional analysis reveals that in most instances, the mel spectrogram is insufficient for exact acoustic signal reconstruction.

Motivated by these insights, our research endeavors to uncover efficient mathematical formulations that precisely reconstruct the acoustic signal from the STFT magnitude spectrogram for successful signal recovery.

In this paper, we present two deterministic algorithms, both polynomial-time, for the precise reconstruction of acoustic signals from two distinct types of STFT magnitude functions with minimal hop sizes. These functions are separately defined in [11] and [18].

Our paper is organized as follows: Section 2 offers a mathematical formulation of the STFT signal recovery problem. Section 3 details our deterministic polynomial-time algorithms, along with their supporting proofs. Section 4 presents the numerical experiments, while Section 5 delves into potential directions for future research in signal reconstruction.

2. PROBLEM SETUP AND NOTATIONS

In this section, we give the formal definition of the STFT problem. Here, we adopt the notations presented in [18] and incorporate certain descriptions from [11].

Let $x = (x(0), x(1), ..., x(T-1))$ be a discrete-time 1-D real-valued signal of length T and $w = (w(0), w(1), ..., w(N-1))$ be a given real-valued window of length N. One type of STFT with respect to the window w, denoted by Y_w, can be defined as follows[11]:

$$Y_w(m,n) = \sum_{t=0}^{T-1} x(t)w(mH-t)e^{-i\frac{2\pi nt}{T}}, \qquad (1)$$

for $0 \le n \le T-1$ and $0 \le m \le \lfloor \frac{T+N-1}{H} \rfloor - 1$, where H is the separation in time between adjacent short-time intervals. In some papers, H is called hop-size, and we

use the same term in our paper. We also need to state that when $t < 0$ or $t \geq N$, $w(t)$ is set to be 0.

The second type of STFT considered in this paper has the following form [18]:

$$S(m,n) = \sum_{k=0}^{N-1} w(k)x(k+Hm)e^{-i2\pi\frac{kn}{N}}, \qquad (2)$$

for $0 \leq n \leq N-1$ and $0 \leq m \leq \lfloor \frac{T-N}{H} \rfloor$.

The STFT magnitude spectrogram can be defined as either $|Y_w(m,n)|$ or $|S(m,n)|$, which is the absolute value of the complex number, and the signal reconstruction problem is to reconstruct $x(t)$ from the STFT magnitude. Because we focus on acoustic signals, which is real-valued, and the given window w is also real-valued, we have $Y_w(m,n) = Y_w^*(m,T-n)$ and $S(m,n) = S^*(m,N-n)$. As a result, the number of the equations is only half of the range of n.

The two signal reconstructions from STFT magnitude problems can be mathematically stated as:

Problem 1. Reconstruct discrete-time 1-D real-valued signal $x(t)$ from the STFT magnitude $|Y_w(m,n)|$, where $0 \leq n \leq C = \lceil \frac{T}{2} \rceil$ and $0 \leq m \leq \lfloor \frac{T+N-1}{H} \rfloor - 1$. Without loss of generality, we assume that $T/2$ and $(T+N-1)/H$ are all integers.

Problem 2. Reconstruct discrete-time 1-D real-valued signal $x(t)$ from the STFT magnitude $|S(m,n)|$, where $0 \leq n \leq C = \lceil \frac{N}{2} \rceil$ and $0 \leq m \leq \lfloor \frac{T-N}{H} \rfloor$. Without loss of generality, we assume that $N/2$ and $(T-N)/H$ are all integers.

If there exists a polynomial-time (for the variable T) algorithm to solve Problem 2, we can reconstruct a long acoustic signal sequence by applying such an algorithm on the different intervals of the signal sequence simultaneously and efficiently. Also, we can reconstruct the acoustic signal at any start time t, without the constraint of starting from 0.

3. PROPOSED RECONSTRUCTION ALGORITHMS

In this section, we outline our deterministic polynomial-time algorithms designed to precisely reconstruct the acoustic signal from its STFT magnitude, subject to certain mild constraints. Subsection 3.1 offers a reconstruction algorithm addressing Problem 1 for values of $H = 1, 2, 3, 4$ and subsection 3.2 introduces a reconstruction solution for Problem 2 when $H = 1, 2$.

3.1. Proposed algorithm and its analysis for Problem 1

The main theorem in this subsection is shown below:

Theorem 3.1. Assume that $H \in \{1,2,3,4\}$, a real-valued $w(n)$ is given and $w(n) \neq 0$ for each n, $T/2$ and $\frac{T+N-1}{H}$ are all integers for $T > N \geq 4H$. Under these assumptions, when given all $|Y_w(m,n)|$, where $0 \leq n \leq T/2$ and $0 \leq m \leq \frac{T+N-1}{H} - 1$, almost all acoustic signals (1-D real-valued) x can be exactly reconstructed up to a global sign in polynomial time.

Proof. Denote $\hat{x}_{m,t} = x(t)w(mH-t)$ to simplify our notations in the proof. Obviously, when $t > mH$ or $mH - N \geq t$, $\hat{x}_{m,t} = 0$. Also, all $\hat{x}_{m,t}$ are real numbers. In our proof, we only consider non-vanishing signals.

Based on $\cos(-\frac{2\pi n}{T}c) = \cos(-\frac{2\pi n}{T}(T-c))$, we have

$$\begin{aligned} |Y_w(m,n)|^2 &= |\sum_{t=0}^{T-1} \hat{x}_{m,t} e^{-i\frac{2\pi nt}{T}}|^2 \\ &= \sum_{t_1=0}^{T-1}\sum_{t_2=0}^{T-1} \cos(-\frac{2\pi n}{T}(t_1-t_2))\hat{x}_{m,t_1}\hat{x}_{m,t_2} \qquad (3) \\ &= \sum_{c=0}^{T-1} \cos(-\frac{2\pi n}{T}c)[\sum_{t=0}^{T-1} \hat{x}_{m,t}\hat{x}_{m,(t+c)\%(T)}], \end{aligned}$$

for $0 \leq n \leq T/2$ and n is an integer, and we can calculate the auto-correlation function as

$$R(m,n) = \sum_{t=0}^{T-1} \hat{x}_{m,t}\hat{x}_{m,(t+n)\%(T)} \qquad (4)$$

for each $0 \leq n \leq T/2$ by solving the linear system in $O(T^3)$ time for a fixed m. Note that we need to use the fact that $R(m,n) = R(m,T-n)$. Here % represents the mod function, which calculates the remainder after the number is divided by the divisor.

First, we consider $m = 0$ to compute $x(0)$. Recall the definition $w(t) \equiv 0$ for $t < 0$ or $t > N-1$, we have $R(0,0) = \hat{x}_{0,0}^2$. Because $x' = -x$ has the same STFT magnitude as x, we can manually set the sign of $x(0)$ as positive or negative. Hence, we get $x(0) = \frac{\sqrt{R(0,0)}}{w(0)}$ and all $\hat{x}_{m,0} = x(0)w(mH)$.

Second, we consider $m = 1$ to compute $x(t)$ for $1 \leq t \leq H$. We propose the following variants of the proposed algorithm as H varies:

$H = 1$: In this case, $R(1,1) = \hat{x}_{1,0}\hat{x}_{1,1}$. We have $x(1) = \frac{R(1,1)}{\hat{x}_{1,0}w(0)}$.

$H = 2$: In this case, $R(1,1) = \hat{x}_{1,0}\hat{x}_{1,1} + \hat{x}_{1,1}\hat{x}_{1,2}$ and $R(1,2) = \hat{x}_{1,0}\hat{x}_{1,2}$. Hence, $\hat{x}_{1,2} = \frac{R(1,2)}{\hat{x}_{1,0}}$, $\hat{x}_{1,1} = \frac{R(1,1)}{\hat{x}_{1,0}+\hat{x}_{1,2}}$, and we can get $x(t) = \hat{x}_{1,t}/w(H-t)$.

$H = 3$: Let $C = \hat{x}_{1,0} = x(0)w(3)$. We have $\hat{x}_{1,3} = R(1,3)/C$. Then, $R(1,2) = C\hat{x}_{1,2} + \hat{x}_{1,1}R(1,3)/C$, $R(1,1) = C\hat{x}_{1,1} + \hat{x}_{1,1}\hat{x}_{1,2} + \hat{x}_{1,2}R(1,3)/C$ and $R(1,0) = C^2 + R(1,3)^2/C^2 + \hat{x}_{1,1}^2 + \hat{x}_{1,2}^2$. Plug the first equation into the last two equations, by solving two quadratic

polynomials we can get $\hat{x}_{1,1}$ and $\hat{x}_{1,2}$ exactly with only one solution in almost all cases.

$H = 4$: Let $C = \hat{x}_{1,0} = x(0)w(4)$. We have $\hat{x}_{1,4} = \frac{R(1,4)}{C}$, $\hat{x}_{1,3} = \frac{R(1,3)}{C} - \frac{R(1,4)}{C^2}\hat{x}_{1,1}$, and $\hat{x}_{1,2} = (R(1,2) - \frac{R(1,3)}{C}\hat{x}_{1,1} + \frac{R(1,4)}{C^2}\hat{x}_{1,1}^2)/(C + \frac{R(1,4)}{C})$. Plug these formulas into $R(1,1) = \sum_{t=0}^{3} \hat{x}_{1,t}\hat{x}_{1,t+1}$, we will have a cubic polynomial in one variable $\hat{x}_{1,1}$. After solving the cubic polynomial algebraically (in polynomial time), we find three possible solutions. Consider the formula $R(1,0) = \sum_{t=0}^{3} \hat{x}_{1,t}^2$, we can check all the possible solutions to find the exact correct one in almost all cases. After that, we will get $\hat{x}_{1,t}$ for $1 \le t \le 4$ to compute $x(t)$.

Finally, consider $m \ge 2$ to get $x(t)$ for $t > H$. Use the same technique in [19], we can solve $x(t)$ for $(m-1)H + 1 \le t \le mH$ from $|Y_w(m,\cdot)|$ incrementally as following:

For a fixed $m_0 \ge 2$, suppose that we know all $x(t)$ for $t \le (m_0 - 1)H$. For each $n \ge H$, Equation (4) will become a linear equation with H unknown variables $\hat{x}_{m_0,t}$ for $(m_0-1)H+1 \le t \le m_0 H$. Because we assume $T > N \ge 4H$ and $n \in [0, T/2]$, the range of n is from 0 to at least $2H$. Hence, we have at least $2H - H + 1 = H + 1$ linear equations for the H unknown variables. As a result, we can solve the linear system to get $x(t)$, where $(m_0 - 1)H + 1 \le t \le m_0 H$, in $O(H^3 + HT)$ time.

Our algorithm will fail only for the cases that encounter a linear system with a singular matrix during our process, and the set of these cases is a 0-measurement set in space $\mathbb{R}^T$.

As a result, after the whole process, we will reconstruct almost all x_k in $O((H^3 + HT)T/H) = O(H^2T + T^2)$ time. □

3.2. Proposed algorithm and its analysis for Problem 2

The main theorem in this subsection is shown below:

Theorem 3.2. Assume that $H = 1$ or $H = 2$, $w(n) \equiv 1$(or other non-zero constant), $N/2$ and $(T-N)/H \ge 2$ are all integers, and $N \ge 4H$. Under these assumptions, when given all $|S(m,n)|$, where $0 \le n \le N/2$ and $0 \le m \le \frac{T-N}{H}$, almost all acoustic signals (1-D real-valued) x can be exactly reconstructed up to a global sign in polynomial time.

Proof. We will use the notation x_t to represent $x(t)$ in some formulas. First, we prove the theorem for $(T - N)/H = 2$.

When $H = 1$, because $|S(m,0)| = |\sum_{k=0}^{N-1} x(k+m)|$, $\sum_{k=0}^{N-1} x(k+m)$ has 2 choices $\pm\sqrt{|S(m,0)|^2}$. Because $m = 2$ when $(T-N)/H = 2$, there are $2^{(2+1)} = 8$ sign choices. Assume we know the signs for the sum, then we can calculate $y_m = x(m+N) - x(m) = \sum_{k=0}^{N-1} x(k + (m+1)) - \sum_{k=0}^{N-1} x(k+m)$ for $m = 0, 1$.

When $H = 2$, consider $\sum_{k=0}^{N-1} x(k+2m) = S^+(m) = \pm\sqrt{|S(m,0)|^2}$, $\sum_{k=0}^{N-1} x(k+2m)(-1)^k = S^-(m) = \pm\sqrt{|S(m,N/2)|^2}$. Because $m = 0, 1, 2$ when $(T - N)/H = 2$, there are $2^{2(2+1)} = 64$ sign choices. Assume we know the signs for the sum, then we can calculate $y_{2m} = x(2m+N) - x(2m) = (S^+(m+1) + S^-(m+1) - S^+(m) - S^-(m))/2$ and $y_{2m+1} = x(2m+N+1) - x(2m+1) = (S^+(m+1) - S^-(m+1) - S^+(m) + S^-(m))/2$ for $m = 0, 1$. Hence, for $H = 1, 2$, we can calculate all $y_k = x(k+N) - x(k)$ for $0 \le k \le 2H - 1$.

Based on $\cos(-\frac{2\pi n}{N}c) = \cos(-\frac{2\pi n}{N}(N-c))$, we have

$$\begin{aligned} |S(m,n)|^2 &= |\sum_{k=0}^{N-1} x(k+Hm)e^{-i2\pi\frac{kn}{N}}|^2 \\ &= \sum_{k_1=0}^{N-1}\sum_{k_2=0}^{N-1} \cos(-\frac{2\pi n}{N}(k_1-k_2))x_{k_1+Hm}x_{k_2+Hm} \quad (5) \\ &= \sum_{c=0}^{N-1} \cos(-\frac{2\pi n}{N}c)[\sum_{k=0}^{N-1} x_{k+Hm}x_{(k+c)\%(N)+Hm}], \end{aligned}$$

for $0 \le n \le N/2$ and n is interger, we can calculate the auto-correlation function as

$$R(m,n) = \sum_{k=0}^{N-1} x_{k+Hm}x_{(k+n)\%(N)+Hm} \quad (6)$$

for each $0 \le n \le N/2$ and $m = 0, 1, 2$ by solving the linear system in $O(N^3)$ time for each m.

Hence, we have the following $3(N/2+1)$ equations

$$\begin{aligned} R(0,n) &= \sum_{k=0}^{N-1} x_k x_{(k+n)\%(N)} \\ R(1,n) - R(0,n) &= \sum_{k=0}^{H-1} (x_{N+k} - x_k)(x_{N+k-n} + x_{k+n}) \\ R(2,n) - R(1,n) &= \sum_{k=0}^{H-1} (x_{N+k+H} - x_{k+H})\cdot \\ &\qquad (x_{N+k-n+H} + x_{k+n+H}) \end{aligned} \quad (7)$$

Because we get $y_k = x_{k+N} - x_k$ for $0 \le k \le 2H - 1$, the last two formulas in Equation (7) will be $R(1,n) - R(0,n) = \sum_{k=0}^{H-1}(x_{N+k-n} + x_{k+n})y_k$ and $R(2,n) - R(1,n) = \sum_{k=0}^{H-1}(x_{N+k-n+H} + x_{k+n+H})y_{k+H}$. If the time index k of x is not smaller than N, we can use $x_k = y_{k-N} + x_{k-N}$ to replace x_k by a time index smaller than N.

As a result, by using y_k, the last two formulas can be transformed into linear equations with N unknown variables x_k for $0 \le k < N$. Because we have $2(N/2+1) = N + 2$ linear equations, if the rank of the linear

system matrix is N, we can solve all x_k for $0 \leq k < N$ by solving the linear system in $O(N^3)$ time. For $N \leq k < T$, use $x_k = y_{k-N} + x_{k-N}$ to get the remaining x_k.

Because we do not know the sign of $S(m, 0)$ and $S(m, N/2)$, we need to repeat the previous steps for each possible sign (8 for $H = 1$ and 64 for $H = 2$), and then check whether the resulting x satisfies the Equations (7). We will only keep the resulting x which can satisfy Equations (7). Note that we always have pairs of results since $x' = -x$ has the same STFT magnitude as x.

Now, we have constructed a polynomial-time deterministic algorithm with time $O(N^3)$ for $T = N + 2H$.

Then we consider $T > N + 2H$. Because $N \geq 4H$ from the assumption, the auto-correlation formula Equation (6), will be a linear equation for the last H variables if we know $x_{Hm}, x_{1+Hm}, ..., x_{N-H-1+Hm}$ for $H \leq n \leq N/2$. Hence, we have $N/2 - H + 1$ linear equations and H unknown variables, we can solve the linear system in $O(N^3)$. This is done by first solving x_k for $0 \leq k < N + 2H$ and then solving $x_{k+(N+2H)+Hm}$ for $0 \leq k < H$ incrementally.

As a result, after the whole process, we will reconstruct all x_k in $O(N^3 + N^3 \frac{T-2H}{H}) = O(N^3 T)$ time. Our algorithm will fail only for the cases that encounter a linear system with a singular matrix during our process. Together with the uniqueness property proved in [19], we conclude that our algorithm can reconstruct almost all 1-D real-valued signals up to a global sign. □

4. NUMERICAL EXPERIMENTS

In this section, we evaluate the Mean Absolute Error (MAE) between the original acoustic signal and the signal reconstructed by our proposed algorithms. Note that previously proposed methods, such as [6, 7], are approximate algorithms while our proposed methods can solve the problem exactly under mild conditions. As a result, we do not compare our methods with any approximation-based baselines.

The implementations of STFT functions are based on Pytorch-STFT. To simplify the experiments, we use the default setting $N = 4H$ from Pytorch-STFT in all experiments using double precision.

4.1. Experiment of Theorem 3.1

We test the algorithm on 5 randomly sampled real-valued 1-D signals with length $T = 8H$. We select such a small T to avoid the rounding error accumulation during the incremental process. The window functions are randomly sampled from $(0, 1)$, and the signals are sampled uniformly from $(-1, 1)$.

In Table 1, we can see that the MAE between the original signals and the reconstructed signals are small, which shows we reconstructed the signals almost exactly. When $H = 4$, our algorithm needs to solve a large linear system and a cubic polynomial, hence, MAE in $H = 4$ is larger than others.

Table 1. Mean Absolute Error (MAE) between input signal and signal reconstructed from $|Y_w(m, n)|$. Numbers in the bracket are standard deviations over five trials.

Random Samples	MAE
H=1	$1.348 \times 10^{-11}(4.836 \times 10^{-22})$
H=2	$1.845 \times 10^{-10}(1.360 \times 10^{-19})$
H=3	$4.584 \times 10^{-13}(4.085 \times 10^{-25})$
H=4	$3.210 \times 10^{-06}(3.607 \times 10^{-11})$

4.2. Experiment of Theorem 3.2

We use a voice recording from LibriTTS dataset [20] with length $T = 98560$ as the dataset. The average absolute value of the voice recording is 0.0449, and the range of the signal is $(-1, 1)$. We also test the algorithm on 5 randomly sampled real-valued 1-D signals with length $T = 1000$, which is sampled uniformly from $(-1, 1)$.

Table 2. Mean Absolute Error (MAE) between the input signal and the signal reconstructed from $|S(m, n)|$. Numbers in the bracket are standard deviations over five trials.

data	Voice Recording	Random Samples
$H = 1$	4.810×10^{-17}	$3.828 \times 10^{-15}(3.855 \times 10^{-30})$
$H = 2$	7.009×10^{-14}	$3.684 \times 10^{-14}(7.799 \times 10^{-28})$

In Table 2, we can see that the MAEs are smaller than 10^{-13}, which shows we reconstructed the signals almost exactly. The reason why MAE is not 0 is that, we use float-64 data type in our codes, which will cause the rounding error when we solve the linear system.

5. FUTURE WORK AND CONCLUSION

We always encounter a system of polynomials with large degrees for larger H, which means that the generalization of our methods may have no algebraic solutions. Because of this, we conjecture that the signal reconstruction problem from STFT magnitude with large H has no solution in radicals, like the statement in Abel–Ruffini theorem[21].

In this paper, we proposed polynomial-time deterministic algorithms for acoustic signal reconstruction from STFT magnitude when the hop size is small. Compared to previous work, our algorithms can solve the problem algebraically and exactly.

6. REFERENCES

[1] Kishore Jaganathan, Yonina C. Eldar, and Babak Hassibi, "Phase retrieval: An overview of recent developments," ArXiv, vol. abs/1510.07713, 2015.

[2] Ehab A. AlBadawy, Andrew Gibiansky, Qing He, Jilong Wu, Ming-Ching Chang, and Siwei Lyu, "Vocbench: A neural vocoder benchmark for speech synthesis," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 881–885.

[3] Jungil Kong, Jaehyeon Kim, and Jaekyoung Bae, "Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis," in Proceedings of the 34th International Conference on Neural Information Processing Systems(NeurIPS), 2020.

[4] Ryan J. Prenger, Rafael Valle, and Bryan Catanzaro, "Waveglow: A flow-based generative network for speech synthesis," 2018, pp. 3617–3621.

[5] Sang-gil Lee, Wei Ping, Boris Ginsburg, Bryan Catanzaro, and Sungroh Yoon, "Bigvgan: A universal neural vocoder with large-scale training," 2023.

[6] D. Griffin and Jae Lim, "Signal estimation from modified short-time fourier transform," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 2, pp. 236–243, April 1984.

[7] Rossen Nenov, Dang-Khoa Nguyen, and Peter Balazs, "Faster than fast: Accelerating the griffin-lim algorithm," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.

[8] Nathanaël Perraudin, Peter Balazs, and Peter L. Søndergaard, "A fast griffin-lim algorithm," in 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2013.

[9] Jonathan Le Roux, H. Kameoka, Nobutaka Ono, Shigeki Sagayama, and Morinosato Wakamiya, "Fast signal reconstruction from magnitude stft spectrogram based on spectrogram consistency," in International Conference on Digital Audio Effects (DAFx), 2010.

[10] Tal Peer, Simon Welker, and Timo Gerkmann, "Beyond griffin-lim: Improved iterative phase retrieval for speech," in International Workshop on Acoustic Signal Enhancement (IWAENC), 2022.

[11] Kishore Jaganathan, Yonina C. Eldar, and Babak Hassibi, "Stft phase retrieval: Uniqueness guarantees and recovery algorithms," IEEE Journal of Selected Topics in Signal Processing, 2014.

[12] Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," Communications on Pure and Applied Mathematics, vol. 66, no. 8, pp. 1241–1274, 2011.

[13] Wen Huang, K. A. Gallivan, and Xiangxiong Zhang, "Solving phaselift by low-rank riemannian optimization methods for complex semidefinite constraints," SIAM Journal on Scientific Computing, vol. 39, no. 5, pp. B840–B859, 2017.

[14] Arian Eamaz, Farhang Yeganegi, and Mojtaba Soltanalian, "One-bit phase retrieval: More samples means less complexity?," IEEE Transactions on Signal Processing, vol. 70, pp. 4618–4632, 2022.

[15] Kaihui Liu, Jiayi Wang, Zhengli Xing, Linxiao Yang, and Jun Fang, "Low-rank phase retrieval via variational bayesian learning," IEEE Access, vol. 7, pp. 5642–5648, 2018.

[16] Lars Thieling, Daniel Wilhelm, and Peter Jax, "Recurrent phase reconstruction using estimated phase derivatives from deep neural networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 7088–7092.

[17] Christopher A. Metzler, Philip Schniter, Ashok Veeraraghavan, and Richard Baraniuk, "prdeep: Robust phase retrieval with a flexible deep network," in International Conference on Machine Learning, 2018.

[18] Nicolas Sturmel and Laurent Daudet, "signal reconstruction from stft magnitude: a state of the art," in International Conference on Digital Audio Effects (DAFx), 2011.

[19] S. Nawab, T. Quatieri, and Jae Lim, "Signal reconstruction from short-time fourier transform magnitude," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 31, no. 4, pp. 986–998, 1983.

[20] Heiga Zen, Viet-Trung Dang, Robert A. J. Clark, Yu Zhang, Ron J. Weiss, Ye Jia, Z. Chen, and Yonghui Wu, "Libritts: A corpus derived from librispeech for text-to-speech," in Interspeech, 2019.

[21] Wikipedia contributors, "Abel–ruffini theorem," https://en.wikipedia.org/wiki/Abel-Ruffini_theorem, 2023.