Signal Inpainting from Fourier Magnitudes
Louis Bahrman, Marina Krémé, Paul Magron, Antoine Deleforge

To cite this version:
Louis Bahrman, Marina Krémé, Paul Magron, Antoine Deleforge. Signal Inpainting from Fourier Magnitudes. EUSIPCO 2023, Sep 2023, Helsinki, Finland. 10.23919/EUSIPCO58844.2023.10289727 . hal-03832480v3

HAL Id: hal-03832480
https://hal.science/hal-03832480v3
Submitted on 22 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Signal Inpainting from Fourier Magnitudes

Louis Bahrman\textsuperscript{1,2} \hspace{1cm} Marina Krémé\textsuperscript{2}, Paul Magron\textsuperscript{2}, Antoine Deleforge\textsuperscript{2}

\textsuperscript{1}LTCI, Télécom Paris, Institut Polytechnique de Paris, Paris, France
louis.bahrman@telecom-paris.fr
\textsuperscript{2}Université de Lorraine, CNRS, Inria, LORIA, Nancy, France
\{ama-marina.kreme, paul.magron, antoine.deleforge\}@inria.fr

Abstract—Signal inpainting is the task of restoring degraded or missing samples in a signal. In this paper we address signal inpainting when discrete Fourier magnitudes are observed. We propose a mathematical formulation of the problem that highlights its connection with phase retrieval, and we introduce two methods for solving it. First, we derive an alternating minimization scheme, which shares similarities with the Gerchberg-Saxton algorithm, a classical phase retrieval method. Second, we propose a convex relaxation of the problem, which is inspired by recent approaches that reformulate phase retrieval into a semidefinite program. We assess the potential of these methods for the task of inpainting gaps in speech signals. Our methods exhibit both a high probability of recovering the original signals and robustness to magnitude noise.

Index Terms—Signal inpainting, phase retrieval, audio restoration, convex relaxation, alternating minimization.

I. Introduction

Signal inpainting [1] is an inverse problem that consists in restoring signals degraded by sample loss. Such a problem typically arises as a result of degradation during signal transmission (packet loss concealment [2]) or in digitization of physically degraded media. Inpainting can also be used to restore signal samples subject to a degradation so heavy that the information about the samples can be considered lost (e.g., signal clipping [3] or impulsive noises [4]). More specifically, in this paper we focus on inpainting compact gaps, which occurs, e.g., when an audio signal is corrupted with clicks [5].

Approaches that tackle this issue can be divided into two categories depending on the number of missing samples or duration of the gaps. When considering short gaps (less than 100 ms), approaches based on autoregressive modeling [6], convex optimization [7], sparse modeling [8], or Bayesian estimation [9] have shown promising results. Conversely, approaches based on sinusoidal modeling [10] or graphs [11] are more suitable for longer gaps (more than 100 ms). However, signal inpainting remains a challenging problem, and these approaches are not adapted to scenarios where some additional information about the signal is available.

We focus here on a setting where the missing gap is contained in a short context window, while the \textit{discrete Fourier magnitudes} of the complete signal on that segment are observed. A similar magnitude-informed setting was notably studied in [12] and [13] in the different context of source separation. Beyond the fact that this problem is open and has not received specific attention in the literature, the motivation behind studying it is the existence of a vast literature dedicated to modeling and processing short-term Fourier magnitudes in the audio literature, e.g., using nonnegative matrix factorization [14] or more recently variational auto-encoders [15]. This is because the Fourier magnitudes of natural signals tend to exhibit smoother and hence more predictable evolution than their respective Fourier phases. In this work, we assume that such a magnitude model has been leveraged beforehand, and we focus on the latest part of the inpainting problem, thus considering that some magnitude estimates are available, possibly up to errors.

The problem then shares a close connection with phase retrieval [16], the task of retrieving a signal from nonnegative measurements (usually magnitudes of a set of inner products). From the seminal early works of Gerchberg and Saxton [17], this task has been revived over the last decade with the development of novel optimization approaches based on gradient descent [18] or convex relaxations [19], [20]. While phase retrieval has experienced considerable progress in recent years, its connection to signal inpainting is largely left to explore, from the theoretical, methodological and applicative standpoints.

This paper aims at contributing to bridging this gap by formulating signal inpainting from Fourier magnitudes as a constrained phase retrieval problem. Inspired by phase retrieval algorithms, we derive two methods to tackle it, based on alternating minimization (AM) and convex relaxation (CR). Experiments conducted on speech signals reveal the potential of these techniques for audio inpainting, since they exhibit a high probability of recovering the original signal, as well as some robustness to magnitude errors. The approach compares favorably to the sate-of-the-art sparsity-based inpainting method in [8] as long as magnitudes are observed with sufficient accuracy.

The rest of this paper is organized as follows. Section II formulates the problem and introduces two methods for solving it. Experiments on speech signals are conducted in Section III. Finally, Section IV draws some concluding remarks.

II. Methods

A. From inpainting to phase retrieval

Let $x^0 \in \mathbb{R}^L$ denote a signal. We partition its support $\{0, \ldots, L - 1\}$ into two sets $v$ and $\bar{v}$ such that $x^0_v \in \mathbb{R}^{L-d}$
Amplitude measures for phase retrieval. We focus on the Euclidean norm in this study. The entries are given by the vector \( b \). This highlights the need for efficient algorithms that provide a stable and efficient manner to recover the complete signal.

As pointed out in the review [24], convex and can be viewed as an instance of quadratic programming, which is known to be NP-hard in general [27]. Notwithstanding our uniqueness result, the problem is non-convex and can be viewed as an instance of quadratic programming, which is known to be NP-hard in general [27].

Let us formulate (1) as the following optimization problem:

\[
\min_{x \in \mathbb{R}^L} \| \Phi x - b \|^2 \quad \text{s.t.} \quad x_v = x_v^3,
\]

where \( \| \cdot \| \) denotes the Euclidean norm. Problem (2) reads as a constrained phase retrieval problem. If the whole signal is missing \( (d = L) \), the constraint vanishes and it reduces to Fourier phase retrieval. We theoretically show in a supplementary material that for almost all \( x^3 \in \mathbb{R}^L \), this problem admits a unique solution if at most 33% of the signal is missing [23]. Note that Fourier phase retrieval has been extensively investigated from a theoretical perspective; we refer the interested reader to the book chapter [24] for a comprehensive review on uniqueness guarantees. The cases where one sample or half of the samples at the end of the signal are observed were respectively covered some time ago in [25] and in [26].

However, both studies considered DFT with zero-padding. Our supplemental theorem complements these results by treating the more challenging unpadded DFT case, using a different proof technique.

Notwithstanding our uniqueness result, the problem is non-convex and can be viewed as an instance of quadratic programming, which is known to be NP-hard in general [27]. As pointed out in the review [24], “there is [currently] no algorithm that knows how to exploit the given entries to recover the complete signal in a stable and efficient manner”. This highlights the need for efficient algorithms that provide a solution. The cornerstone of our approaches lies in introducing an auxiliary phase variable \( u \in \mathbb{C}^L \) such that \( |u| = 1 \). Then, \( b = |\text{diag}(b)u| \), where \( \text{diag}(b) \) is the diagonal matrix whose entries are given by the vector \( b \). We then turn our attention to minimizing the following auxiliary function, which is exact with respect to (2) in the sense of [28]:

\[
\min_{x \in \mathbb{R}^L, u \in \mathbb{C}^L} \| \Phi x - \text{diag}(b)u \|^2 \quad \text{s.t.} \quad x_v = x_v^3, \quad |u| = 1.
\]

**B. Alternating minimization**

The first approach we propose to solve (3) is an AM scheme. Let us first fix \( u \) and derive the update for \( x \), for which we propose to incorporate the constraint \( x_v = x_v^3 \) within the loss function. To that end, let us reorder \( x \) as \( x = [x_\circ, x_v] \) (and similarly for \( x^3 \)), and split \( \Phi \) accordingly as \( \Phi = [\Phi_\circ, \Phi_v] \), with \( \Phi_\circ \in \mathbb{C}^{L \times d} \) and \( \Phi_v \in \mathbb{C}^{L \times (L-d)} \). Using these notations, we have

\[
\min_{x_\circ \in \mathbb{R}^L, \Phi_v, x_v} \| \Phi_\circ x_\circ + \Phi_v x_v^3 - \text{diag}(b)u \|^2.
\]

Since \( \Phi_\circ \) is full-rank, it has a left inverse which is its Hermitian transpose \( \Phi_\circ^H \). Besides, recall that since the DFT is an orthogonal transform, then \( \Phi_\circ^H \Phi_v = 0 \). Altogether this yields the following solution to (4):

\[
x_\circ = \Phi_\circ^H \text{diag}(b)u.
\]

Let us now derive the update for \( u \) when \( x \) is fixed, for which (3) rewrites:

\[
\min_{u \in \mathbb{C}^L} \| \Phi x - \text{diag}(b)u \|^2 \quad \text{s.t.} \quad |u| = 1,
\]

which is straightforward to solve:

\[
u = \frac{\Phi x}{|\Phi x|}.
\]

Alternating (5) and (7) yields a procedure summarized in Algorithm 1. We discuss the initialization strategy in Section III-A. Note that at line 5 we apply the real part function to minimizing the following auxiliary function, which is exact with respect to (2) in the sense of [28]:

\[
\min_{x \in \mathbb{R}^L, u \in \mathbb{C}^L} \| \Phi x - \text{diag}(b)u \|^2 \quad \text{s.t.} \quad x_v = x_v^3, \quad |u| = 1.
\]

**Remark:** Algorithm 1 consists in computing the DFT of a signal, setting its magnitude to a target value, inverting the DFT, putting back the observed samples, and repeating. As such, it is similar to the Gerchberg-Saxton algorithm [17] except that the signal-domain magnitude constraint is here replaced with a projection onto the partially observed samples. Besides, Algorithm 1 can be seen as a member of the general family of constrained phase retrieval algorithms reviewed in [24]. While such algorithms are generally heuristically derived, the proposed alternate minimization scheme on an exact auxiliary function is guaranteed to converge - though not necessarily to a global minimum - by construction.

**C. Convex relaxation**

Let us now derive a method inspired by the PhaseCut algorithm [20], which is based on a CR of the problem. This approach consists in reformulating phase retrieval as a constrained trace minimization problem by lifting it to a higher dimensional space and relaxing the rank-one constraint. The
Algorithm 1 AM for signal inpainting

Input: \( b \in \mathbb{R}_+^L \) : observations, \( x_v^i \) : known signal, \( \Phi \) : Fourier matrix,
1: Initialize \( x_v^0 \) and \( x^0 \) ← \([x_v^0 \ x_v^0]\)
2: \( i \leftarrow 0 \)
3: while convergence not reached do
4: \( u^{i+1} \leftarrow \Phi x_v^{i+1} \)
5: \( x_v^{i+1} \leftarrow \Re(\Phi H \text{diag}(b) u^{i+1}) \)
6: \( x^{i+1} \leftarrow \frac{x_v^{i+1}}{\|x_v^{i+1}\|} \)
7: \( i \leftarrow i + 1 \)
8: end while
Output: Reconstructed signal \( x^{(i)} \)

resulting problem can then be efficiently solved via semi-definite programming. We can then introduce the expression of \( x \) given by (5):

\[
\min_{u \in \mathbb{C}^L} \| (\Phi_v \Phi_v^H - I) \text{diag}(b) + \Phi_v x_v \|^2 \text{ s.t. } |u| = 1. \tag{8}
\]

Now, let us introduce the following:

\[
\tilde{m} := [(\Phi_v \Phi_v^H - I) \text{diag}(b), \Phi_v x_v] \text{ and } \tilde{u} := \begin{bmatrix} u \\ 1 \end{bmatrix}, \tag{9}
\]

from which we can rewrite (8) as:

\[
\min_{\tilde{u} \in \mathbb{C}^{L+1}} \| \tilde{m} \tilde{u} \|^2 \text{ s.t. } |\tilde{u}| = 1 \text{ and } |\tilde{u}| = 1. \tag{10}
\]

Drawing on [20], we lift and relax (10) to the following convex problem:

\[
\min_{\tilde{U} \in \mathbb{C}^{(L+1) \times (L+1)}} \text{trace} (\tilde{M} \tilde{U}) \text{ s.t. } \text{diag}(\tilde{U}) = 1, \ \tilde{U} \succeq 0, \tag{11}
\]

where \( \tilde{M} = \tilde{m}^H \tilde{m} \in \mathbb{C}^{(L+1) \times (L+1)} \) and \( \tilde{U} = \tilde{u} \tilde{u}^H \). As in [20] and [12], problem (11) can be solved by means of a block coordinate descent algorithm, which we summarize in Algorithm 2 (lines 1 to 11). It consists of a nested loop where at each iteration \( i \), the columns of \( \tilde{U} \) are updated sequentially using the notation:

\[
k^c = \{0, \ldots, k - 1, k + 1, \ldots, L - 1 \}. \tag{12}
\]

This yields a global solution \( \tilde{U} \) to problem (11). If this solution is of rank 1, i.e., \( \tilde{U} = \tilde{u} \tilde{u}^H \), then \( \tilde{u}[0, \ldots, L - 1]/|\tilde{u}[L]| \) is guaranteed to globally solve (8). We then obtain a global solution of (3) via:

\[
x_v = \Phi_v^H \text{diag}(b) \tilde{u}[0, \ldots, L - 1]/|\tilde{u}[L]|. \tag{13}
\]

However, \( \tilde{U} \) needs not be rank-1 in general. Hence, as commonly done in semi-definite relaxations, its closest rank-1 approximation is used in practice, which by the Eckart-Young-Mirsky theorem amounts to setting \( \tilde{u} \) to the eigenvector associated to the largest eigenvalue of \( \tilde{U} \).

III. EXPERIMENTS

In this section we assess the potential of our methods for inpainting gaps in audio signals. Our code is available online.\(^3\)

A. Experimental setting

a) Data: We consider 100 speech signals from the Librispeech dataset [29]. Signals are sampled at 16 kHz. For each signal, we extract a non-silent sub-signal of variable length \( L \) at a random location, where we create a gap of \( d \) samples.

b) Proposed methods: For the AM algorithm, iterations stop when the the maximum loss variation over the previous 5 iterations does not exceed \( 10^{-10} \), or when a maximum number of 1000 iterations is reached. Preliminary experiments revealed no significant difference between various basic initialization schemes, e.g., using a random or zero phase in the Fourier domain, or using a random or zero missing signal in the time domain. We also derived an initialization scheme inspired by the spectral initialization method in [30], but this did not yield any significant improvement either. Therefore, the results displayed hereafter use \( x_v^0 \) ← 0. The CR algorithm uses a fixed amount \( n_{\text{iter}} = 10 \) since the performance did not show further improvement beyond in our experiments. We simply set the barrier parameter \( \nu \) to 0 as in [12]. We also consider a combined CR+AM algorithm. This technique consists in first estimating the signal with CR, and then using this estimate as an initialization for AM, with the same stopping criterion as above. Our methods are fed with the ground truth magnitudes \( (b = |\Phi x|^2) \), except in the last experiment where noisy magnitudes are considered.

\(^3\)https://github.com/Louis-Bahrman/Inpainting-Fourier
c) **Baselines:** Let us note that typical phase retrieval methods [17], [20] are not appropriate comparison baselines since they are agnostic to the observed samples. Besides, more recent phase retrieval approaches proposed in audio [22], [31] are tailored to time-frequency processing, and are therefore not suitable for the setup considered in this paper. On the other hand, traditional audio restoration techniques such as [6] do not leverage magnitude knowledge, making comparisons unfair. Nevertheless, results obtained using the Fourier-sparsity-based method SPAIN [8] are shown, as an indication of the potential of exploiting magnitude models beyond sparsity. The same parameters as in [8] are used, except for the relaxation parameter $r$ which is set to 32.

d) **Metric:** To assess the quality of the reconstruction, we resort to the signal-to-error ratio (SER) expressed in dB:

$$
\text{SER}(\hat{x}_v, x^-_v) = 10 \log_{10} \left( \frac{\| \hat{x}_v \|^2}{\| \hat{x}_v - x^-_v \|^2} \right),
$$

where $\hat{x}$ denotes the estimated signal (higher is better). Let us outline that the SER is only calculated over the set of missing samples $\hat{v}$ since $\hat{x}_v = x^-_v$ elsewhere. Note that AM’s initialization (replacing the missing signal with zeros) corresponds to an SER of 0. Based on preliminary listening tests, we consider that perfect reconstruction is achieved when the SER is greater than 20 dB.

### B. Results

First, we compare the recovery rate of the three proposed methods with respect to the missing signal fraction. The results corresponding to $L = 1024$ (64 ms) are displayed in Fig. 2. Our methods significantly outperform SPAIN in this setting, regardless of the missing fraction. This demonstrates the potential of exploiting Fourier magnitudes when these are available. CR+AM achieves the best results and consistently outperforms the other approaches. This is explained by the ability of CR to provide a solution that is more likely to converge to a global minimum than AM’s basic initialization. While all the methods exhibit a performance drop when the missing fraction increases, CR+AM and AM still yield perfect reconstruction in at least 80 % of the cases when less than 30 % of the signal is missing. The results of CR+AM have to be weighted against the higher memory and time complexity of CR, which processes matrices of size $L \times L$ instead of vectors of size $L$. Accordingly, a general guideline might be to resort to the AM method when very few samples are missing (5 to 10 %), and to turn to CR+AM in all remaining scenarios, at a higher computational cost.

Then, Fig. 3 shows the recovery rate achieved by the best performing method CR+AM as a function of the signal length (in samples) and the missing signal fraction. The red line represents the 33 % theoretical limit under which perfect reconstruction is achievable for almost all signals [23].

![Fig. 2. Influence of the missing fraction onto performance.](image2)

![Fig. 3. Probability of achieving perfect reconstruction for CR+AM, as a function of the signal length (in samples) and the missing signal fraction. The red line represents the 33 % theoretical limit under which perfect reconstruction is achievable for almost all signals [23].](image3)

![Fig. 4. Influence of the magnitude noise onto performance.](image4)
estimation errors. To simulate such a scenario, we consider noisy magnitudes \( \mathbf{b} = \max (0, |\Phi \mathbf{x}| + \mathbf{n}) \), where \( \mathbf{n} \) is a white Gaussian noise whose variance is adjusted to fit a given magnitude signal-to-noise (SNR) value. The results are presented in Fig. 4, where the signal length is fixed at \( L = 1024 \) and the fraction of missing signal at 25\%. Overall, we observe that all of our methods exhibit a similar linear decay in performance in the log-log plot when the SNR falls below 20 dB, suggesting an encouraging robustness to errors. Interestingly, while CR performs worse than the other methods at higher SNRs, it still yields solutions that provide a better initialization than zero to an encouraging robustness to errors. Interestingly, while CR

IV. CONCLUSION

We have investigated the problem of signal inpainting from observed Fourier magnitude measurements. After formulating the associated optimization problem, we proposed two estimation methods based on alternating minimization and convex relaxation. Experiments in the noiseless regime highlighted that combining the two methods yields near-exact reconstruction most of the time when less than 33\% of the signal is missing, closely following our theoretical limit. Further experiments demonstrated a relative robustness of the methods to magnitude errors. Future work will therefore focus on applying such techniques to more realistic settings where magnitudes are estimated, e.g., via light magnitude interpolation methods in the short-time Fourier domain or via data-driven models such as deep neural networks.

REFERENCES