Accelerated Griffin-Lim algorithm: A fast and provably converging numerical method for phase retrieval

Rossen Nenov†, Dang-Khoa Nguyen⋆, Peter Balazs† and Radu Ioan Boţ⋆
† Acoustics Research Institute, Vienna, Austria
⋆ University of Vienna, Austria

***Abstract*—The recovery of a signal from the magnitudes of its transformation, like the Fourier transform, is known as the phase retrieval problem and is of big relevance in various fields of engineering and applied physics. In this paper, we present a fast inertial/momentum based algorithm for the phase retrieval problem. Our method can be seen as an extended algorithm of the Fast Griffin-Lim Algorithm, a method originally designed for phase retrieval in acoustics. The new numerical algorithm can be applied to a more general framework than acoustics, and as a main result of this paper, we prove a convergence guarantee of the new scheme. Consequently, we also provide an affirmative answer for the convergence of its ancestor Fast Griffin-Lim Algorithm, whose convergence remained unproven in the past decade. In the final chapter, we complement our theoretical findings with numerical experiments for the Short Time Fourier Transform phase retrieval and compare the new scheme with the Griffin-Lim Algorithm, the Fast Griffin-Lim Algorithm, and two other iterative algorithms typically used in acoustics.**

***Index Terms*—phase retrieval, inertial algorithm, Griffin-Lim algorithm, Fast Griffin-Lim algorithm, convergence guarantee.**

I. Introduction

Reconstructing the phase of a signal from phaseless measurements of its (Short Time) Fourier transform is a pervasive challenge, called the phase retrieval problem. It arises in a great amount of areas of applications, most prominently in audio processing [1, 2], imaging [3, 4], and electromagnetic theory [5, 6]. Therefore it is, to this day, an active research topic with numerous algorithms designed to find satisfactory solutions to the phase retrieval problem. These algorithms can be divided into two classes: non-iterative and iterative algorithms. For the *Short Time Fourier Transform* (STFT), it was analysed in [7] that the performance of non-iterative algorithms strongly depends on the redundancy of the STFT. They perform better than iterative algorithms in terms of quality and speed for high redundancies, which are rarely considered in practice. For the lower, more common, redundancies the iterative methods gave more qualitative results than the non-iterative ones, which motivates their study.

The *Griffin-Lim algorithm* (GLA) [8] is a well known and widely used iterative algorithm based on the method of alternating projections and applied to the phase retrieval problem in the time-frequency setting. In optics, this algorithm is also known as the Gerchberg-Saxton algorithm [9]. Inspired by the *Fast iterative shrinkage threshold algorithm* (FISTA [10]), the authors of [11] proposed to introduce an inertial/momentum step to GLA, which formed the *Fast Griffin-Lim algorithm* (FGLA). As a result, they obtained a method that converges faster than GLA in practice and recovers signals with lower reconstruction error, but the convergence guarantee remained an open question. Since its introduction, FGLA gained considerable traction for phase retrieval in audio processing [12, 13] and machine learning [14, 15], but still, a convergence result for this iterative algorithm was pending.

In this work, we will present the *Accelerated Griffin-Lim algorithm* (AGLA), a new iterative method, and prove a convergence result for it in a generalized setting covering a wide range of areas, where the phase retrieval problem arises, beyond the field of audio processing. This method is an extension of FGLA, and therefore our results cover FGLA as well.

Furthermore, we will round out the theoretical results with the comparison of numerical simulations of AGLA against its predecessors and other iterative algorithms to highlight the improved numerical performance of our algorithm.

This manuscript is an extended version of [29] providing theoretical convergence results and more extensive numerical experiments. This work was supported by the Austrian Science Fund FWF-project NoMASP ("Nonsmooth nonconvex optimization methods for acoustic signal processing"; P 34922-N). The authors would like to express their gratitude to Dr. Nicki Holighaus (Austria Academy of Sciences) for his significant input and ideas in the numerical experiments section, and to Prof. Radu Balan (University of Maryland) for his valuable insight into the properties of the problem.

II. The Phase Retrieval Problem

A linear and injective transformation from $\mathbb{C}^L \to \mathbb{C}^M$ with $M \geq L$ can be written as a transformation matrix T in $\mathbb{C}^{M\times L}$ with full column rank, for example, the discrete Fourier transform or the analysis operator of a finite frame [16]. The vector $s \in [0,+\infty)^M$ will denote the measured magnitudes of the coefficients of the transform. The phase retrieval problem can be expressed mathematically as finding the signal $x^* \in \mathbb{C}^L$, whose transform coefficients match the magnitudes s, that is $|Tx^*| = s$, where $|\cdot|$ is understood as the absolute value applied componentwise. The feasibility, uniqueness, and stability of the phase retrieval problem have been studied in many works, most notably in [17, 18]. The results in the present paper are actually independent of the solvability of the phase retrieval problem.

In practice, measurements can include some noise and inaccuracies and therefore we are usually looking for solutions x^* of the following minimization problem

$$\min_{x\in\mathbb{C}^L} \left\| |Tx| - s \right\|, \tag{1}$$

where $\|\cdot\|$ denotes the norm of $\mathbb{C}^M$. In [11] it was proposed to consider this problem, as the task in finding a vector c^* in the set of coefficients admitted by the transformation matrix T, namely its range, which is as close as possible to the set of coefficients, whose magnitude match with s. The range of the transformation matrix T will be denoted as

$$C_1 = \{c \in \mathbb{C}^M \mid \exists x \in \mathbb{C}^L : c = Tx\}. \tag{2}$$

This set is a linear subspace of $\mathbb{C}^L$. Let C_2 be the set of vectors whose magnitudes are equal to s, namely

$$C_2 = \{c \in \mathbb{C}^M \mid |c_i| = s_i \quad \forall i \in \{1,\dots,M\}\}. \tag{3}$$

The set C_2 is, by definition, compact. To formulate the problem (1) as finding the closest point between two sets, we introduce two distance functions. The indicator function δ_C of a set C is defined as

$$\delta_C(c) = \begin{cases} 0 & \text{if } c \in C, \\ +\infty & \text{else,} \end{cases}$$

and the distance function d_C to a compact set C is defined as

$$d_C(c) = \min_{v\in C} \|c - v\| \qquad \text{for } c \in \mathbb{C}^M. \tag{4}$$

Our aim is to solve the following optimization problem

$$\min_{c\in\mathbb{C}^M} f(c) := \delta_{C_1}(c) + \frac{1}{2} d^2_{C_2}(c), \tag{5}$$

which is nonconvex and has a close connection to the original problem (1), as can be seen later.

For finding the closest point between two sets iteratively, projection operators are a common tool. Since C_1 is a linear subspace spanned by T, we can write its orthogonal projection as

$$P_{C_1}(c) = TT^\dagger c,$$

where $T^\dagger$ denotes the pseudo-inverse of T, which is well-defined since T is assumed to have full column rank [16].

Since C_2 is a nonconvex closed set, the projection of a vector c onto C_2 is nonempty but not always unique. Therefore we will denote by $\overline{P}_{C_2} : \mathbb{C}^M \rightrightarrows \mathbb{C}^M$ the possibly set valued projection

$$\overline{P}_{C_2}(c) = \arg\min_{v\in C_2} \|c - v\|,$$

which maps c to the set of its closest points in C_2. As in [11] we will denote by $P_{C_2} : \mathbb{C}^M \to \mathbb{C}^M$ a closed form for a possible choice of the projection onto C_2, which we will use in the algorithms

$$(P_{C_2}(c))_i = \begin{cases} \frac{s_i c_i}{|c_i|} & \text{if } c_i \neq 0, \\ s_i & \text{if } c_i = 0. \end{cases} \tag{6}$$

This projection is equivalent to scaling the elements of c to have the magnitudes of s, without changing the phase. The following proposition, whose proof can be found in the appendix, gives another convenient way to write d_{C_2} and states that this choice of P_{C_2} indeed minimizes the distance function and is almost everywhere the unique element of $\overline{P}_{C_2}$.

Proposition II.1. *For all $c \in \mathbb{C}^M$ the distance to C_2 can be written as*

$$d_{C_2}(c) = \| |c| - s \| \tag{7}$$

and $d_{C_2}(c) = \|c - P_{C_2}(c)\|$ holds. If $c \notin D$, where $D := \{c \in \mathbb{C}^M \mid \exists i \in \{1,...,M\} : c_i = 0$ and $s_i \neq 0\}$, then $P_{C_2}(c)$ is the unique closest point to c in C_2, namely $\overline{P}_{C_2}(c) = \{P_{C_2}(c)\}$.

The set D consists of the points, where the magnitude of one coefficient is zero while the respective measured one – s_i – is not. This set has been observed to be problematic, for example, in the work [19] it is proven that the phase derivative of the STFT is numerically unstable, i.e., there is a peculiar pole phenomenon at points in this set. Later on, we will prove that for general frames the function $\frac{1}{2}d^2_{C_2}$ is differentiable only on the complement of D. Therefore this set will turn out to play a significant role in the convergence analysis of the iterates.

Proposition II.1 motivates the choice of the objective function f, since for $c \in C_1$ we see that $f(c)$ coincides with the function value of (1) at $T^\dagger c$. One can show that for any local/global minimizer c^* of (5), $x^* = T^\dagger c^*$ is a local/global minimizer of (1) and vice versa.

III. The Algorithms

In 1984, Griffin and Lim presented the algorithm known as *Griffin-Lim algorithm* (GLA). They showed that the iterates of the algorithm converge to a set of critical points of a magnitude-only distance measure and that the objective function values of the iterates are non-increasing. In the following, N denotes the amount of iterations of the algorithms, which can be chosen as $+\infty$ as well. Here, without any risk of confusion, we use the subscript n to mean the n^{th} iteration, like c_n, t_n. On the other hand, the subscript i means c_i is the i^{th} component of the vector $c \in \mathbb{C}^M$.

Algorithm 1 Griffin-Lim algorithm

INITIALIZE $c_0 \in \mathbb{C}^M$

Iterate for $n = 1, \dots, N$

$c_n = P_{C_1}(P_{C_2}(c_{n-1}))$

RETURN $T^\dagger c_N$

In 2013, Perraudin, Balazs and Søndergaard stated that GLA can be seen as the method of alternating projections of the iterates onto C_2 and C_1 and proposed the *Fast Griffin-Lim Algorithm* (FGLA) by adding an inertial/momentum step [11]. The algorithm is based on the idea of the inertial proximal-gradient method, in the spirit of the Heavy Ball method [20] and Fast iterative shrinkage threshold algorithm (FISTA) [10], which also works in the nonconvex setting [21].

This algorithm achieved better results than GLA in numerical experiments, but a convergence guarantee was not addressed.

Algorithm 2 Fast Griffin-Lim algorithm

INITIALIZE $c_0 \in \mathbb{C}^M$, $t_0 \in C_1$ and $\alpha > 0$

Iterate for $n = 1, \dots, N$

$t_n = P_{C_1}(P_{C_2}(c_{n-1}))$

$c_n = t_n + \alpha(t_n - t_{n-1}),$

RETURN $T^\dagger c_N$

In this paper we present a further modification to GLA, by adding a second inertial sequence $(d_n)_{n\in\mathbb{N}}$, which will not be projected. Its purpose is to stabilize the algorithm and avoid getting stuck at the points, in which FGLA stops at, while the distance between the projection of $(c_n)_{n\in\mathbb{N}}$ and the nonprojected $(d_n)_{n\in\mathbb{N}}$ is still large. A similar idea can be found in [21, 22].

Algorithm 3 Accelerated Griffin-Lim algorithm

INITIALIZE $c_0 \in \mathbb{C}^L$, $t_0, d_0 \in C_1$ and $\alpha, \beta, \gamma > 0$

Iterate for $n = 1, \dots, N$

$t_n = (1-\gamma)d_{n-1} + \gamma P_{C_1}(P_{C_2}(c_{n-1}))$

$c_n = t_n + \alpha(t_n - t_{n-1}),$

$d_n = t_n + \beta(t_n - t_{n-1})$

RETURN $T^\dagger c_N$

AGLA is a generalization of FGLA since for $\gamma = 1$ the generated sequences by FGLA and AGLA coincide. In the following chapter, we will state parameter choices of α, β, and γ, for which we prove the convergence of the algorithm.

IV. Convergence of the function values

To guarantee consistent performance of iterative schemes for optimization problems, one has to establish theoretical convergence results. This involves analyzing the convergence behavior of the function values at the generated iterates and addressing both the subsequential and global convergence of the iterates towards a critical point or a local/global minimum of the underlying problem.

In this chapter, we prove the convergence result for AGLA. We will use the following identity, which is a generalization of the parallelogram law, typically used in the proof of convergence of the function values of the iterates for algorithms with inertial sequences. It will be used several times in the analysis. It can be shown by a simple calculation, hence we omit the detail. Precisely, for any two vectors $a, b \in \mathbb{C}^M$ and any two real numbers $\tau, \sigma \in \mathbb{R}$, it holds

$$\|\tau a + \sigma b\|^2 = (\tau+\sigma)\,\tau\,\|a\|^2 + (\tau+\sigma)\,\sigma\,\|b\|^2 - \tau\sigma\,\|a-b\|^2. \quad (8)$$

Using this identity we will state the proof of the function values of the iterates generated by AGLA for certain parameter regimes.

Theorem IV.1. *Let $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ be the sequences generated by AGLA. Suppose that*

$$0 < \gamma < 2 \text{ and } 0 \le 2\beta\,|1-\gamma| < 2-\gamma, \quad (9)$$

and

$$0 \le \alpha < \begin{cases} \left(1-\frac{1}{\gamma}\right)\beta + \frac{1}{\gamma} - \frac{1}{2} & \text{if } 0 < \gamma \le 1, \\ \frac{1}{2\beta(\gamma-1)+\gamma} - \frac{1}{2} & \text{if } 1 < \gamma < 2. \end{cases} \quad (10)$$

Then the following statements are true

(i) *There exist constants $K_1 > K_2 > 0$ such that the following descent property holds for all $n \ge 1$*

$$d^2_{C_2}(c_n) + K_1\Delta^2_{t_n} \le d^2_{C_2}(c_{n-1}) + K_2\Delta^2_{t_{n-1}},$$

where $\Delta_{t_n} := \|t_n - t_{n-1}\|$. *Therefore* Δ_{t_n} *converges to zero and* $\lim_{n\to+\infty} d^2_{C_2}(c_n) \in \mathbb{R}$ *exists.*
(ii) $(c_n)_{n\in\mathbb{N}}$ *is a bounded sequence and every cluster point* c^* *of* $(c_n)_{n\in\mathbb{N}}$ *fulfills*

$$P_{C_1}(P_{C_2}(c^*)) = c^*.$$

Proof. Let $n \geq 1$. By the definition of the algorithm, the sequences $(c_n)_{n\geq 1}$, $(t_n)_{n\in N}$ and $(d_n)_{n\in\mathbb{N}}$ lie in the linear subspace C_1. This observation will be useful throughout the proof and turns our focus on analyzing, how the distances of the iterates to C_2 behave.

We aim at an expression for $\|c_{n-1} - P_{C_2}(c_{n-1})\|^2$ by using the properties of the projections. For this purpose, note that P_{C_1} is the orthogonal projection onto the linear subspace C_1 and therefore the identity

$$\|P_{C_1}(x) - x\|^2 + \|y\|^2 = \|P_{C_1}(x) - x + y\|^2 \quad (11)$$

holds for all $(x, y) \in \mathbb{C}^L \times C_1$. Define $x_n := P_{C_2}(c_{n-1})$ and $y_n = P_{C_1}(x_n)$, then for arbitrary $z \in C_1$ it holds $z - y_n \in C_1$. Hence according to (11) we have

$$\|y_n - x_n\|^2 + \|z - y_n\|^2 = \|z - x_n\|^2. \quad (12)$$

By rewriting the definition of the algorithm, we see that

$$\frac{1}{\gamma}t_n + \frac{\gamma-1}{\gamma}d_{n-1} = P_{C_1}(P_{C_2}(c_{n-1})), \quad (13)$$

and therefore we can write $y_n = \frac{1}{\gamma}t_n + \frac{\gamma-1}{\gamma}d_{n-1}$. Since $z - y_n = \frac{1}{\gamma}(z - t_n) + \frac{\gamma-1}{\gamma}(z - d_{n-1})$ the following expression can be expanded using the identity (8)

$$\|z - y_n\|^2 = \frac{1}{\gamma}\|z - t_n\|^2 + \frac{\gamma-1}{\gamma}\|z - d_{n-1}\|^2 - \frac{\gamma-1}{\gamma^2}\|t_n - d_{n-1}\|^2. \quad (14)$$

Combining (12) and (14) leads to

$$\|y_n - x_n\|^2 - \frac{\gamma-1}{\gamma^2}\|t_n - d_{n-1}\|^2 = \|z - x_n\|^2 - \frac{1}{\gamma}\|z - t_n\|^2 - \frac{\gamma-1}{\gamma}\|z - d_{n-1}\|^2. \quad (15)$$

The left hand side is independent of the choice of z, hence we can substitute z by c_n and c_{n-1} and equate both right hand sides of (15) to get

$$\|c_n - x_n\|^2 - \frac{1}{\gamma}\|c_n - t_n\|^2 - \frac{\gamma-1}{\gamma}\|c_n - d_{n-1}\|^2 = d^2_{C_2}(c_{n-1}) - \frac{1}{\gamma}\|c_{n-1} - t_n\|^2 - \frac{\gamma-1}{\gamma}\|c_{n-1} - d_{n-1}\|^2, \quad (16)$$

where we used the fact that $d_{C_2}(c_{n-1}) = \|c_{n-1} - x_n\|$. By the definition of the generated sequences, we can see that the following identities hold

$$\|c_n - t_n\|^2 = \alpha^2\Delta^2_{t_n}, \quad (17)$$
$$\|c_n - d_n\|^2 = (\alpha - \beta)^2\Delta^2_{t_n}, \quad (18)$$
$$\|t_n - d_n\|^2 = \beta^2\Delta^2_{t_n}. \quad (19)$$

Furthermore, we can use the property that the distance of c_n to C_2 is not larger than the distance of c_n to an arbitrary point in C_2

$$d_{C_2}(c_n) = \min_{x\in C_2}\|c_n - x\| \leq \|c_n - x_n\|. \quad (20)$$

Applying (17), (18) and (20) into (16) yields

$$d^2_{C_2}(c_n) - \frac{\alpha^2}{\gamma}\Delta^2_{t_n} - \frac{\gamma-1}{\gamma}\|c_n - d_{n-1}\|^2 \leq d^2_{C_2}(c_{n-1}) - \frac{1}{\gamma}\|c_{n-1} - t_n\|^2 - \frac{\gamma-1}{\gamma}(\alpha-\beta)^2\Delta^2_{t_{n-1}}. \quad (21)$$

In order to handle the differences $c_{n-1} - t_n$ and $c_n - d_{n-1}$, we will first represent these as linear combinations of differences of consecutive terms of the sequence $(t_n)_{n\in\mathbb{N}}$. Therefore we need to distinguish between the two cases $\gamma \in (0, 1]$ and $\gamma > 1$.

Case 1: Assume that $\gamma \in (0, 1]$. By the definition of the sequences, we see that

$$c_n - d_{n-1} = (1+\alpha)(t_n - t_{n-1}) - \beta(t_{n-1} - t_{n-2})$$
$$t_n - c_{n-1} = (t_n - t_{n-1}) - \alpha(t_{n-1} - t_{n-2})$$

and by applying the identity (8) and leaving out a positive term, we get the following estimations

$$\|c_n - d_{n-1}\|^2 \geq (1+\alpha-\beta)((1+\alpha)\Delta^2_{t_n} - \beta\Delta^2_{t_{n-1}}),$$
$$\|t_n - c_{n-1}\|^2 \geq (1-\alpha)(\Delta^2_{t_n} - \alpha\Delta^2_{t_{n-1}}).$$

Applying these estimates into (21) leads to

$$d^2_{C_2}(c_n) + K_1\Delta^2_{t_n} \leq d^2_{C_2}(c_{n-1}) + K_2\Delta^2_{t_{n-1}}, \quad (22)$$

with $K_1 := \frac{1-\gamma}{\gamma}(1+2\alpha+\alpha^2-\beta-\alpha\beta) + \frac{1}{\gamma}(1-\alpha-\alpha^2)$ and $K_2 := \frac{1-\gamma}{\gamma}(\beta-\alpha\beta+\alpha^2) + \frac{1}{\gamma}(\alpha-\alpha^2)$. Lemma A.1 ensures that (9) and (10) imply $K_1 > K_2 > 0$.

Case 2: Assume that $\gamma > 1$. Similarly, we see that

$$c_n - d_{n-1} = (1+\alpha)(t_n - t_{n-1}) + \beta(t_{n-2} - t_{n-1})$$
$$t_n - c_{n-1} = (t_n - t_{n-1}) - \alpha(t_{n-1} - t_{n-2})$$

and by applying the identity (8), we get

$$\|c_n - d_{n-1}\|^2 \leq (1+\alpha+\beta)((1+\alpha)\Delta^2_{t_n} + \beta\Delta^2_{t_{n-1}}),$$
$$\|t_n - c_{n-1}\|^2 \geq (1-\alpha)(\Delta^2_{t_n} - \alpha\Delta^2_{t_{n-1}}).$$

Applying these estimates into (21) leads to

$$d^2_{C_2}(c_n) + K_1\Delta^2_{t_n} \leq d^2_{C_2}(c_{n-1}) + K_2\Delta^2_{t_{n-1}}, \quad (23)$$

with $K_1 := \frac{1-\gamma}{\gamma}(1+2\alpha+\alpha^2+\beta+\alpha\beta)+\frac{1}{\gamma}(1-\alpha-\alpha^2)$ and $K_2 := \frac{1-\gamma}{\gamma}(\alpha^2-\beta-3\alpha\beta)+\frac{1}{\gamma}(\alpha-\alpha^2)$. Lemma A.1 asserts that (9) and (10) imply $K_1 > K_2 > 0$.

In both scenarios, we obtain the desired descent inequality. Moreover, we can deduce from (22) and (23) that the sequence

$$\left(d^2_{C_2}(c_n) + K_2\Delta^2_{t_n}\right)_{n\in\mathbb{N}} \tag{24}$$

is positive and non-increasing as n increases, therefore the sequence (24) converges as $n \to +\infty$. In order to use telescoping sum arguments, we can rewrite (22) and (23) to get

$$0 \leq (K_1-K_2)\Delta^2_{t_n} \leq d^2_{C_2}(c_{n-1}) - d^2_{C_2}(c_n) + K_2(\Delta^2_{t_{n-1}} - \Delta^2_{t_n}).$$

If we sum up this inequality we deduce that

$$(K_1-K_2)\sum_{j=2}^{N}\Delta^2_{t_j} \leq d^2_{C_2}(c_1) - d^2_{C_2}(c_N) + K_2(\Delta^2_{t_1} - \Delta^2_{t_N}).$$

Since for all $n \geq 2$ it holds $d^2_{C_2}(c_0) + K_2\Delta^2_{t_0} \geq 0$, by taking the limit $n \to +\infty$ we see that

$$\sum_{j=2}^{+\infty}\Delta^2_{t_j} \leq d^2_{C_2}(c_1) + K_2\Delta^2_{t_1} < +\infty$$

holds, and therefore $\Delta_{t_n} \to 0$ as $n \to +\infty$. Moreover, we already showed that (24) is converges and as a consequence $d_{C_2}(c_n)$ also converges as $n \to +\infty$. Hence the sequence $(c_n)_{n\in\mathbb{N}}$ must be bounded, since its distance to the bounded set C_2 converges. By (17) and (18), we see that $(\|c_n - t_n\|)_{n\in\mathbb{N}}$ and $(\|c_n - d_n\|)_{n\in\mathbb{N}}$ converge to zero as well. Using this observation and (13) we conclude

$$P_{C_1}(P_{C_2}(c_n)) - c_n \to 0 \text{ as } n \to +\infty. \tag{25}$$

Furthermore, since c_n is a bounded sequence, there exist cluster points. For each cluster point c^*

$$P_{C_1}(P_{C_2}(c^*)) = c^*$$

has to hold by (25). □

In numerical experiments we will see that whenever β or γ are not chosen to fulfill (9), then AGLA does not converge. This suggests that it might not be possible to extend the condition (9). On the other hand, (9) and (10) together are sufficient conditions to guarantee convergence.

V. Convergence of the iterates

The goal of this section is to analyze the conditions under which the convergence of the entire sequence of iterates can be guaranteed. To achieve this, we will associate $c \in \mathbb{C}^M$ with $y \in \mathbb{R}^{M\times 2}$, where $y_i = (\mathrm{Re}(c_i), \mathrm{Im}(c_i))$ for all $i \in \{1,\dots,M\}$, in order to apply subdifferential calculus and the *Kurdyka–Łojasiewicz-property* (KL-property) V.4, which are defined for functions from $\mathbb{R}^{M\times 2}$ to $\overline{\mathbb{R}} := \mathbb{R}\cup\{\pm\infty\}$. The set $C_1 \subseteq \mathbb{R}^{M\times 2}$ remains a linear subspace and the set C_2 can be written as

$$C_2 = \{y \in \mathbb{R}^{M\times 2} : \|y_i\| = s_i, \forall i = 1,...,M\}. \tag{26}$$

The domain of an extended real-valued function f is defined as $\mathrm{dom}\, f = \{y : f(y) < +\infty\}$. For our objective function (5) it holds $\mathrm{dom}\, f = C_1$. A function is called proper if it never attains the value $-\infty$ and its domain is a nonempty set. Since our objective function is nonconvex and not everywhere differentiable, we introduce the notion of the limiting subdifferential following [23, Definition 8.3].

Definition V.1. For the proper function $f : \mathbb{R}^{M\times 2} \to \overline{\mathbb{R}}$ and point $\bar{y} \in \mathrm{dom}\, f$, the *regular subgradient* $\hat{\partial} f(\bar{y})$ is defined as the set of vectors $\bar{v} \in \mathbb{R}^{M\times 2}$ which fulfill

$$\liminf_{y\to\bar{y}} \frac{f(y) - f(\bar{y}) - \langle \bar{v}, y-\bar{y}\rangle}{\|y-\bar{y}\|} \geq 0,$$

and the *limiting subdifferential* $\partial f(\bar{y})$ is defined as the set of vectors $\bar{v} \in \mathbb{R}^{M\times 2}$ such that there exist sequences $(y_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^{M\times 2}$ and $v_n \in \hat{\partial} f(y_n)$ such that $y_n \to \bar{y}$, $f(y_n) \to f(\bar{y})$ and $v_n \to \bar{v}$ as $n \to +\infty$.

The subdifferential is a set-valued operator and its domain is defined as $\mathrm{dom}\,\partial F := \{u \in \mathbb{R}^{M\times 2} : \partial F(u) \neq \emptyset\}$. If f is continuously differentiable on a open set, then ∂f is single-valued and thus coincide with ∇f on this set [23, Corrolary 9.19].

The following proposition states the formula for the subgradient of our objective function, whose proof is postponed to the appendix.

Proposition V.2. *For the function $f : \mathbb{R}^{M\times 2} \to \overline{\mathbb{R}}$, $y \mapsto \delta_{C_1}(y) + \frac{1}{2}d^2_{C_2}(y)$ the limiting subdifferential $\partial f : \mathbb{R}^{M\times 2} \rightrightarrows \mathbb{R}^{M\times 2}$ is given by*

$$\partial f(y) = C_1^{\perp} + y - P_{C_2}(y), \text{ for } y \in C_1 \setminus D,$$

where

$$D = \{y \in \mathbb{R}^{M\times 2} \mid \exists i \in \{1,...,M\} : y_i = 0 \text{ and } s_i \neq 0\}.$$

Furthermore for $y \in C_1 \setminus D$ it holds

$$d_{\partial f(y)}(0) \leq \|y - P_{C_1}(P_{C_2}(y))\|.$$

Here for simplicity, we keep the notation f for the objective function and write $f(y)$ when it maps from $\mathbb{R}^{M\times 2}$ to $\overline{\mathbb{R}}$, and $f(c)$ when $f\colon \mathbb{C}^M \to \overline{\mathbb{R}}$.

Before we introduce the KL-property, we have to define the class of concave and continuous functions.

Definition V.3. Let $\eta \in (0, +\infty]$. We denote by Φ_η all function $\varphi : [0, \eta) \to [0, +\infty)$ which satisfy the following conditions

(i) $\varphi(0) = 0$
(ii) φ is C^1 on $(0, \eta)$ and continuous at 0
(iii) for all $s \in (0, \eta)$: $\varphi'(s) > 0$

We state the definition of the KL-property, which can be used to prove convergence iterates of first-order and second-order methods, with nonconvex objective functions [24]. Intuitively speaking, functions that satisfy this property are not too flat at their respective local minimizers and critical points. For $b > a$, we will write $[\![a < F < b]\!]$ to denote the level set $\{u \in \mathbb{R}^{M\times 2} : a < F(u) < b\}$ of a function $F : \mathbb{R}^{M\times 2} \to \overline{\mathbb{R}}$.

Definition V.4. Let $F : \mathbb{R}^{M\times 2} \to \overline{\mathbb{R}}$ be proper and lower semicontinuous. ∂F denotes the subdifferential of F. The function F is said to have the *Kurdyka–Łojasiewicz-property* at $\overline{u} \in \operatorname{dom} \partial F$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\overline{u}$ and a function $\varphi \in \Phi_\eta$, such that for all

$$u \in U \cap [\![F(\overline{u}) < F < F(\overline{u}) + \eta]\!],$$

the following holds

$$\varphi'(F(u) - F(\overline{u}))d_{\partial F(u)}(0) \geq 1.$$

The function φ is called the desingularizing function.

The following result, taken from [25, Lemma 6], provides a uniformized KL property on a neighborhood and will be crucial in our convergence analysis.

Lemma V.5. *Let Ω be a compact set and F: $\mathbb{R}^{M\times 2} \to \overline{\mathbb{R}}$ be a proper and lower semicontinuous function. Assume that F is constant on Ω and satisfies the KL property at each point of Ω. Then there exist $\varepsilon > 0, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for every $\overline{u} \in \Omega$ and every element u in the intersection*

$$\left\{u \in \mathbb{R}^{M\times 2} \colon d_\Omega(u) < \varepsilon\right\} \cap [\![F(\overline{u}) < F < F(\overline{u}) + \eta]\!]$$

the following holds:

$$\varphi'(F(u) - F(\overline{u}))\, d_{\partial F(u)}(0) \geq 1.$$

The KL-property holds for a broad class of functions, especially for the indicator functions and distance functions of semi-algebraic sets as stated in [26]. Since C_1 as a linear subspace is algebraic and C_2 is algebraic as we see in (26), we know that our objective function f in (5) has the KL-property.

For the proof of the convergence of the iterates, we introduce the regularized version of f, namely $F_K : \mathbb{R}^{M\times 2} \times \mathbb{R}^{M\times 2}$ defined as

$$F_K(y, z) = \delta_{C_1}(y) + \frac{1}{2}d^2_{C_2}(y) + \frac{K}{2\alpha^2}\|y - z\|^2,$$

with $K > 0$. Then by Proposition 10.5 and Corollary 10.9 of [23] the formula of the subdifferential $\partial F_K : \mathbb{R}^{M\times 2} \times \mathbb{R}^{M\times 2} \rightrightarrows \mathbb{R}^{M\times 2} \times \mathbb{R}^{M\times 2}$ is given as

$$\partial F_K(y, z) = \left\{\partial f(y) + \frac{K}{\alpha^2}(y - z)\right\} \times \left\{\frac{K}{\alpha^2}(z - y)\right\}. \tag{27}$$

Furthermore, by [27] we know that F_K inherits the KL-property from f. Simple calculations show that F is non-increasing, and the distance of the subgradient to 0 can be estimated from above.

Proposition V.6. *Let $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ be the sequence generated by AGLA and assume that* (9) *and* (10) *hold. Then the following statements are true:*

(i) *There exist a constant $\kappa_1 > 0$ such that for all $n \in \mathbb{N}$ it holds*

$$F_{K_2}(c_n, t_n) + \kappa_1\Delta^2_{t_n} \leq F_{K_2}(c_{n-1}, t_{n-1}), \tag{28}$$

where K_2 is defined as in Theorem IV.1.

(ii) *If $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, where D is defined as in Proposition V.2, then there exist an integer $m \in \mathbb{N}$ and a constant $\kappa_2 > 0$ such that for all $n \geq m$*

$$d_{\partial F_{K_2}(c_n,t_n)}(0) \leq \kappa_2(\Delta_{t_{n+1}} + \Delta_{t_n}) \tag{29}$$

holds.

The proof of this proposition can be found in the appendix. Now we can state the proof that the generated iterates of AGLA converge by using the KL-property and the previous observations.

Theorem V.7. *Let $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ be the sequence generated by AGLA. Assume that* (9) *and* (10) *hold and that $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, where D is defined as in Proposition V.2. Then $(c_n)_{n\in\mathbb{N}}$ converges and the limit c^* is a critical point of f, if $c^* \notin D$.*

Proof. For simplicity we will denote $F_n = F_{K_2}(c_n, t_n)$ and $\lim_{n\to\infty} F_n = F^*$, which exists by Theorem IV.1. For any cluster point c^* of the sequence $(c_n)_{n\in\mathbb{N}}$, we see that $F(c^*) = F^*$ by Theorem IV.1. If there exists an integer $n_0 \in \mathbb{N}$ such that $F_{n_0} = F^*$, then the inequality (28) implies that $\Delta_{t_n} = 0$ for all $n \geq n_0$. Hence by the definition of the algorithm and (17)-(19), the sequences $(t_n)_{n\geq n_0}$, $(d_n)_{n\geq n_0}$ and $(c_n)_{n\geq n_0}$ are constant and the statement is proven.

Now assume that $F_n > F^*$ for all $n \in \mathbb{N}$. If $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, then we can choose $m \in \mathbb{N}$ such that for all $n \geq m$ the vector $c_n \notin D$.

From Theorem IV.1, we know that the sequence $((c_n, t_n))_{n\in\mathbb{N}}$ is bounded, and therefore there exist cluster points. Let us denote by Ω the set of all cluster points of the sequence $((c_n, t_n))_{n\in\mathbb{N}}$. We can see that Ω is closed and also bounded. Moreover, the value of F over Ω always equals F^*. Since F has the KL-property, according to Lemma V.5, there exist $\varepsilon, \eta > 0$ and a function $\varphi \in \Phi_\eta$ such that all element (y, z) in the intersection

$$\big\{(y,z) \in \mathbb{R}^{M\times 2} \times \mathbb{R}^{M\times 2} \colon d_\Omega((y,z)) < \varepsilon\big\} \cap [\![F^* < F_{K_2} < F^* + \eta]\!] \quad (30)$$

it holds

$$\varphi'\left(F_{K_2}\left((y,z)\right) - F^*\right) d_{\partial F_{K_2}((y,z))}(0) \geq 1. \quad (31)$$

In the following we will use the KL-property to prove that the sequence $(t_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence and therefore is convergent. Since F_n converges to F^* and $d_{\partial F_{K_2}(c_n,t_n)}(0) \to 0$ as $n \to +\infty$ due to Proposition V.6, there exists $n_1 \geq m$ such that (c_n, t_n) belongs to the intersection (30) for all $n \geq n_1$. This means by (31) and (29), the inequality

$$\kappa_2\varphi'(F_n - F^*)(\Delta_{t_{n+1}} + \Delta_{t_n}) \geq \varphi'(F_n - F^*) d_{\partial F_n}(0) \geq 1, \quad (32)$$

holds for all $n \geq n_1$. Since φ is concave and differentiable, we know that

$$\varphi(z) - \varphi(y) \geq \varphi'(z)(z - y).$$

By choosing $z = F_n - F^*$ and $y = F_{n+1} - F^*$ then plugging (28) and (32), we obtain after some rearranging

$$\varphi(F_n - F^*) - \varphi(F_{n+1} - F^*) \geq \frac{\kappa_1}{\kappa_2} \frac{\Delta_{t_{n+1}}^2}{\Delta_{t_{n+1}} + \Delta_{t_n}}$$

holds for $n \geq n_1$. By applying Lemma A.2 in the appendix we can see that this implies that

$$\frac{9\kappa_2}{4\kappa_1}\left(\varphi(F_n - F^*) - \varphi(F_{n+1} - F^*)\right) \geq 2\Delta_{t_{n+1}} - \Delta_{t_n}$$

holds. Summing up this inequality for $j = n_1, \ldots, \bar{n}$ leads to

$$\sum_{j=n_1}^{\bar{n}} \Delta_{t_j+1} \leq \frac{9\kappa_2}{4\kappa_1}\left(\varphi(F_{n_1} - F^*) - \varphi(F_{\bar{n}+1} - F^*)\right) + \Delta_{t_{n_1}} - \Delta_{t_{\bar{n}+1}}.$$

Since φ is nonnegative, by passing $\bar{n} \to \infty$, we deduce

$$\sum_{j=n_1}^{+\infty} \Delta_{t_j} \leq \frac{9\kappa_2}{4\kappa_1}\varphi(F_{n_1} - F^*) + \Delta_{t_{n_1}} < +\infty.$$

This implies that $(t_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence and therefore converges. Since $(c_n)_{n\in\mathbb{N}}$ and $(d_n)_{n\in\mathbb{N}}$ are linear combinations of $(t_n)_{n\in\mathbb{N}}$, we conclude that these series converge as well and the limit of $(c_n)_{n\in\mathbb{N}}$ has to be c^* according to Theorem IV.1. Furthermore $0 \in \partial f(c^*)$ by Proposition V.2 if $c^* \notin D$. □

Since AGLA with $\gamma = 1$ reduces to FGLA, we can state the following convergence properties of FGLA to complement the result in [11]. It is a direct consequence of Theorem IV.1 and V.7.

Corollary V.8. *Let $(c_n)_{n\in\mathbb{N}}$ be the sequences generated by FGLA. If $\alpha \in (0, 0.5)$, then*

(i) *There exist constants $K_1 > K_2 > 0$ such that the following descent property holds for all $n \geq 1$*

$$d_{C_2}^2(c_n) + K_1\Delta_{t_n}^2 \leq d_{C_2}^2(c_{n-1}) + K_2\Delta_{t_{n-1}}^2,$$

where $\Delta_{t_n} := \|t_n - t_{n-1}\|$. Therefore Δ_{t_n} converges to zero and $\lim_{n\to+\infty} d_{C_2}^2(c_n) \in \mathbb{R}$ exists.

(ii) *Furthermore $(c_n)_{n\in\mathbb{N}}$ is a bounded sequence and every convergent subsequence converges to a point c^*, which fulfills*

$$P_{C_1}(P_{C_2}(c^*)) = c^*$$

and is therefore a critical point if $c^* \notin D$, where D is defined as in Proposition V.2.

(iii) *If $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, then $(c_n)_{n\in\mathbb{N}}$ is a convergent sequence.*

To summarize our results, we have established conditions and parameter regimes for AGLA and FGLA that guarantee that the generated sequences $(c_n)_{n\in\mathbb{N}}$ assert a decreasing property and that every cluster point $c^* \notin D$ is a local minimizer of our objective function. To guarantee convergence, we needed the condition $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, since outside of D the projection $\overline{P}_{C_2}$ is unique and $\frac{1}{2}d_{C_2}^2$ is continuously differentiable. It is important to mention that the inequalities used in the statements above do not cover all possible choices of parameters for which we have observed convergence in the simulations. A similar situation occurs even if $\frac{1}{2}d_{C_2}^2$ is assumed to have Lipschitz-continuous gradient, see [21, 22]. Notice that Lipschitz-continuity does not hold in our model, according to Proposition V.2.

Most importantly, our proofs show that for AGLA and FGLA a good performance and minimizing properties for the phase retrieval problem can be guaranteed, when the parameters are well chosen.

In Figure 1 we plotted for fixed $\beta \in [0, 5]$ and $\gamma \in [0.1, 1.5]$ the upper bound for α s.t. α, β, γ satisfy the convergence guarantees (9) and (10) of Theorem IV.1. The white areas symbolize the combinations of β and

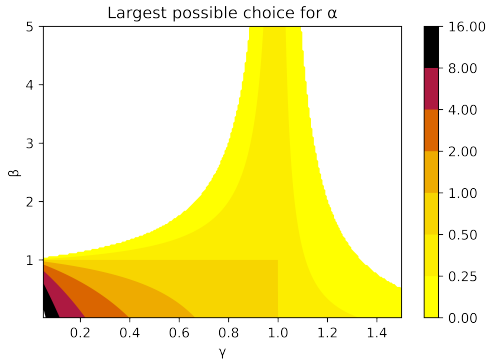

Fig. 1: Largest possible choice for α in AGLA for given β and γ based on the conditions (9) and (10) that guarantee convergence

γ, for which (9) is not fulfilled, i.e. where we expect AGLA not to converge.

Before we take a look at the numerical simulations we will state a corollary, which motivates for which iterates we are going to plot the function values.

Corollary V.9. *Let $(c_n)_{n\in\mathbb{N}}$ be the sequence generated by AGLA. Assume that* (9) *and* (10) *hold and that $c_n \in D$ for at most finitely many $n \in \mathbb{N}$, where D is defined as in Proposition V.2. Then for $y_n = P_{C_1}(P_{C_2}(c_n))$ the following properties hold:*

(i) *For $n \geq 0$*

$$d^2_{C_2}(y_n) \leq d^2_{C_2}(c_n) - \|y_n - c_n\|^2 .$$

(ii) *The sequence $(y_n)_{n\in\mathbb{N}}$ converges as $n \to +\infty$.*

Proof. Let $n \geq 0$. By (11) we know that

$$\|y_n - P_{C_2}(c_n)\|^2 = \|c_n - P_{C_2}(c_n)\|^2 - \|y_n - c_n\|^2 .$$

By the definition of the distance function we see that $d^2_{C_2}(y_n) \leq \|y_n - P_{C_2}(c_n)\|^2$ and $d^2_{C_2}(c_n) = \|c_n - P_{C_2}(c_n)\|^2$ hold and by this (i) is proven. Furthermore by the results of the Theorems IV.1 and V.7 we can deduce by (13) that $(y_n)_{n\in\mathbb{N}}$ converges as well. □

VI. Numerical Experiments

In this section we will present the results of our numerical experiments and test the performance of the algorithm AGLA for the STFT-spectrogram inversion. As signals we used the © EBU Audio Test sequences, which provide 70 audio files for testing. To reduce the computational time for each signal, we trimmed down every signal to their first two seconds. In this paper we only present the results of some selected signals of this test set. A reproducible research addendum is available at http://ltfat.org/notes/059/, where we provide the code, from which one can test different configurations of windows and parameters and the results of our parameter testing. It is important to mention, that the presented algorithm and convergence proofs work for a broad range of transformations. In this publication we restrict ourselves to the application in acoustics using a STFT for T.

Best 5 combinations with convergence guarantee			
α	β	γ	Average SSNR
0.09	1.10	0.20	9.45696
0.39	1.90	0.90	9.44955
0.48	1.15	0.95	9.44320
0.54	0.55	0.90	9.44319
0.51	0.60	0.95	9.44246

TABLE I: The best parameter combinations satisfying the convergence guarantee with respect to average SSNR after 100 iterations over the test set

Best 5 combinations over all			
α	β	γ	Average SSNR
0.99	1.00	1.30	11.47489
1.00	1.10	1.30	11.45270
1.00	1.05	1.30	11.43239
0.99	0.99	1.30	11.42138
1.05	1.30	1.25	11.39242

TABLE II: The best parameter combinations with respect to average SSNR after 100 iterations over the test set

The simulations were performed with hop size of 32 and 256 FFT bins and a Gaussian window using the LTFAT toolbox [28]. This choice of hop size and bins ensures that it is possible to reconstruct the signals uniquely from their measurements [17]. For more numerical experiments with a different window function, we refer to [29]. As a quality measure we used the *Spectrogram Signal to Noise Ratio*, which is defined as

$$\mathrm{SSNR}(x) = -10\log_{10}\left(\frac{\||Tx| - s\|}{\|s\|}\right).$$

It measures the similarity of the spectrogram of the signal to the desired spectrogram, but it does not necessarily relate to perceived audio quality. Maximizing the SSNR for a signal x is equivalent to minimizing our objective function (5), since $d_{C_2} = \||\cdot| - s\|$ for given s and the negative logarithm is monotonely decreasing.

Having three parameters in AGLA rather than one in FGLA makes the algorithm on the one hand more flexible and adaptable, but on the other hand makes it more difficult to assert, which parameter combination is the best. Therefore we first present a summary of our numerical parameter tests for the parameters of α, β and γ. We took ten signals of the test sequence set

and let the algorithms run for 100 iterations, initialized with the projected measurements vector, namely $c_0 = t_0 = d_0 = P_{C_1}(s)$, where s denotes the measurements of the STFT of a signal. The signals were chosen to cover a wide range of acoustical signals, including single instruments, human speech and full orchestras. We tested the performance of parameter combinations, satisfying the sufficient conditions to guarantee convergence, and also the best possible combinations, disregarding the convergence guarantee. It is important to note that these results will vary for different choices of window functions and different redundancies of the STFT. A table including the final SSNR for all tested combinations can be found in the research addendum.

We computed for $\beta \in [0.1, 2]$ and $\gamma \in [0.2, 1.6]$ the largest possible α up to two decimal points, which satisfies the conditions (9) and (10) and tested the parameter combinations for ten different signals. The results are displayed in Table I.

It is interesting to note that all parameter combinations satisfying (9) and (10) resulted in a very similar SSNR after 100 iterations. Overall the combination with $\alpha = 0.09$, $\beta = 1.1$ and $\gamma = 0.2$ performed the best after 100 iterations, followed by combinations, where $\alpha \approx 0.5$.

We did the same test considering further 532 combinations with $\alpha \in [0.95, 1.15]$, $\beta \in [0.95, 1.5]$ and $\gamma \in [0.95, 1.3]$. Even though these combinations are not covered by the convergence guarantee, they show the best possible convergence of our proposed algorithm AGLA. In these tests, we noticed that if β and γ do not fulfill (9), then the algorithm does not converge. This observation is included in the research addendum. It seems optimal to choose $\gamma = 1.3$, $\alpha \approx 1$ and $\beta \geq \alpha$.

We compared AGLA to two other algorithms, which were observed to perform well in the STFT phase retrieval in [30]. The first algorithm is the *Relaxed Averaged Alternating Projections* (RAAR) proposed by R. Luke for the phase retrieval problem in Diffraction Imaging [31].

Algorithm 4 Relaxed Averaged Alternating Reflections

INITIALIZE $c_0 \in \mathbb{C}^n$ and $\lambda \in (0, 1]$

Iterate for $n = 1, \ldots, N$

$c_{n+1} = \frac{\lambda}{2}\left(c_n + R_{C_1}(R_{C_2}(c_n))\right) + (1-\lambda)P_{C_2}(c_n)$

RETURN $T^\dagger c_N$

According to [32] the best performance of RAAR for speech signals can be expected for $\lambda = 0.9$.

The second algorithm is the *Difference Map* (DM) proposed by V. Elser for phase retrieval in [33].

Algorithm 5 Difference Map

INITIALIZE $c_0 \in \mathbb{C}^n$ and $\rho \in \mathbb{R} \setminus \{0\}$

Iterate for $n = 1, \ldots, N$

$t_n = P_{C_2}(c_n) + \frac{1}{\rho}(P_{C_2}(c_n) - c_n)$

$s_n = P_{C_1}(c_n) + \frac{1}{\rho}(P_{C_1}(c_n) - c_n)$

$c_{n+1} = c_n + \rho(P_{C_1}(t_n) - P_{C_2}(s_n))$

RETURN $T^\dagger c_N$

In general choosing ρ close to 1 yields the best performance of DM and in [30] it was observed that the choice $\rho = 0.8$ performs best for speech signals. To the best of our knowledge the convergence for DM is unproven.

Based on the observation in Corollary V.9, we plotted for FGLA and AGLA the SSNR of $y_n = P_{C_1}(P_{C_2}(c_n))$ and for RAAR and DM the SSNR of the iterates c_n respectively.

For the comparison between the algorithms, we evaluated AGLA, FGLA, DM, and RAAR by initializing with the projected measurements vector, namely, taking $c_0 = t_0 = d_0 = P_{C_1}(s)$. For the first simulations we used the 60 signals which were not included in the parameter search to remove any bias. For the purpose of a better overview, we did not include the algorithm GLA in this test, since the observations in [11] clearly showed that FGLA outperforms GLA in nearly every setting. For DM and RAAR we chose $\rho = 0.8$ and $\lambda = 0.9$ respectively, for FGLA $\alpha = 0.99$ and for AGLA the best combination from Table II. In Figure 2 we depicted the SSNR reached by the algorithms after 100 iterations over the tests set. Furthermore we included in Figure 3 four graphs for the SSNR reached by the iterates of the four algorithms over 1000 iterations. These selected graphs are not necessarily representative of the overall average SSNR reached by the algorithms, but are included to show certain reoccurring behaviours, we want to highlight.

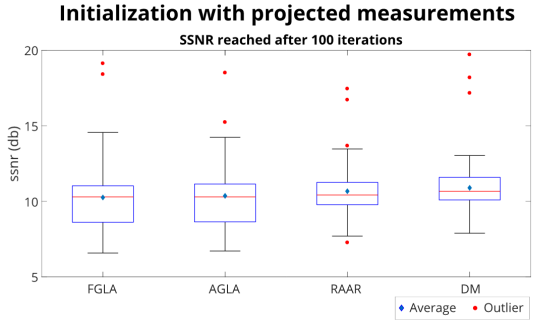

Fig. 2: Box plots depicting the SSNR that FGLA, AGLA, RAAR, and DM reached over the test set after 100 iterations with their respective best parameter choice and initialization with the projected measurements vector

For the next experiment, we initialized the algorithms with the reconstructed signal from the *Phase Gradient Heap Integration* (PGHI) presented in [34] and compared them in Figure 5 and Figure 4 for the same signals.

PGHI is a noniterative method for the phase retrieval problem, based on the phase-magnitude relations of a continuous STFT. It is efficient, but when used for the discrete setting it introduced inaccuracies depending on the parameters of the discrete STFT, and it was observed to give a good starting point for iterative algorithms [7].

For the last experiment, we ran GLA with FGLA and AGLA with their respective best parameter choices covered by the convergence guarantee, proven in this paper. The results are displayed in Figure 7.

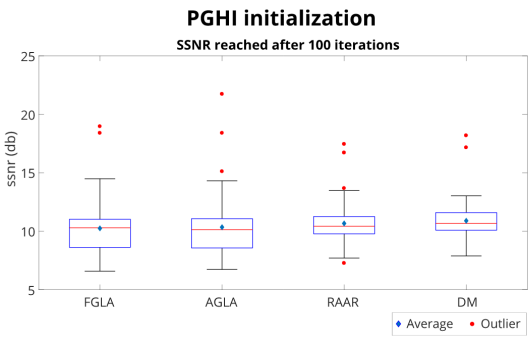

Fig. 4: Box plots depicting the SSNR that FGLA, AGLA, RAAR, and DM reached over the test set after 100 iterations with their respective best parameter choice and PGHI initialization

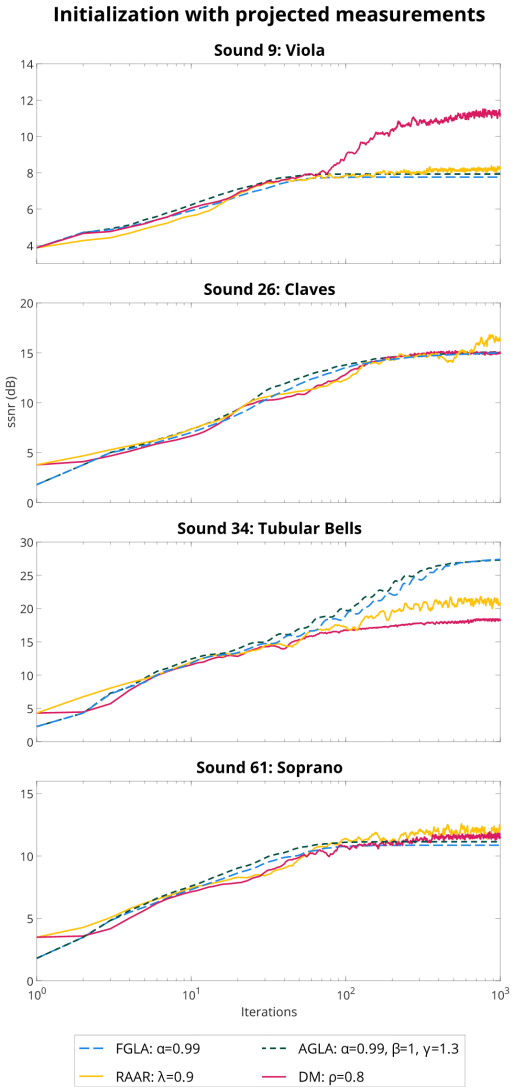

Fig. 3: Comparison of FGLA, AGLA, RAAR, and DM with their respective best parameter choice and initialization with the projected measurements vector

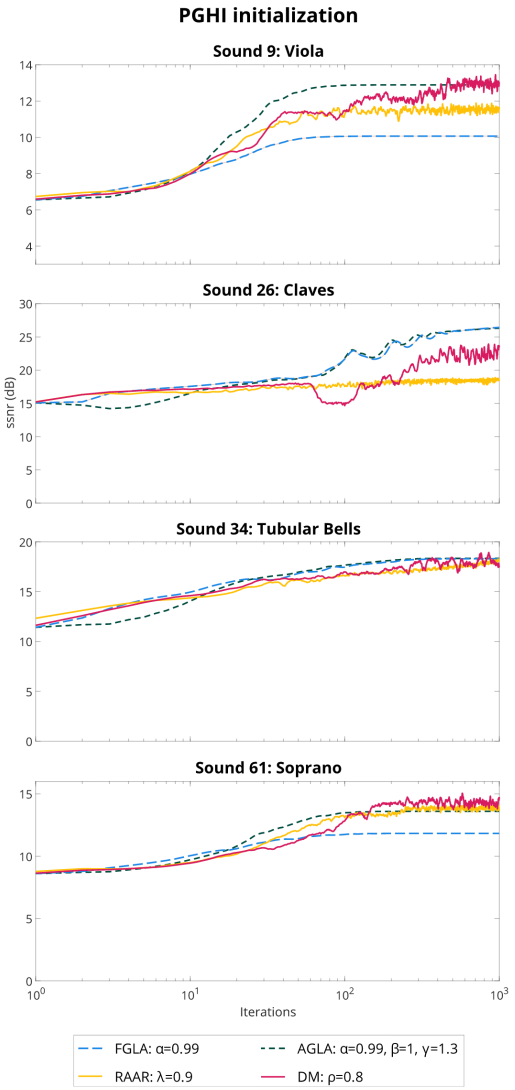

Fig. 5: Comparison of FGLA, AGLA, RAAR, and DM with their respective best parameter choice and PGHI initialization

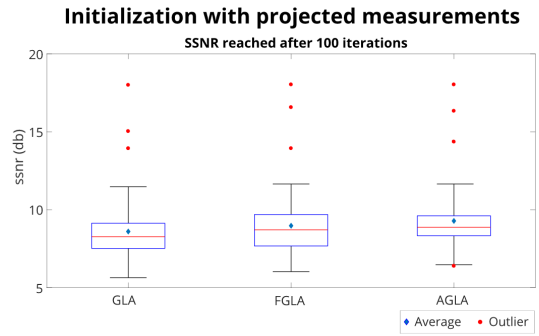

Fig. 6: Box plots depicting the SSNR that GLA, FGLA and AGLA reached over the test set after 100 iterations with their respective best parameter choice satisfying the convergence guarantee

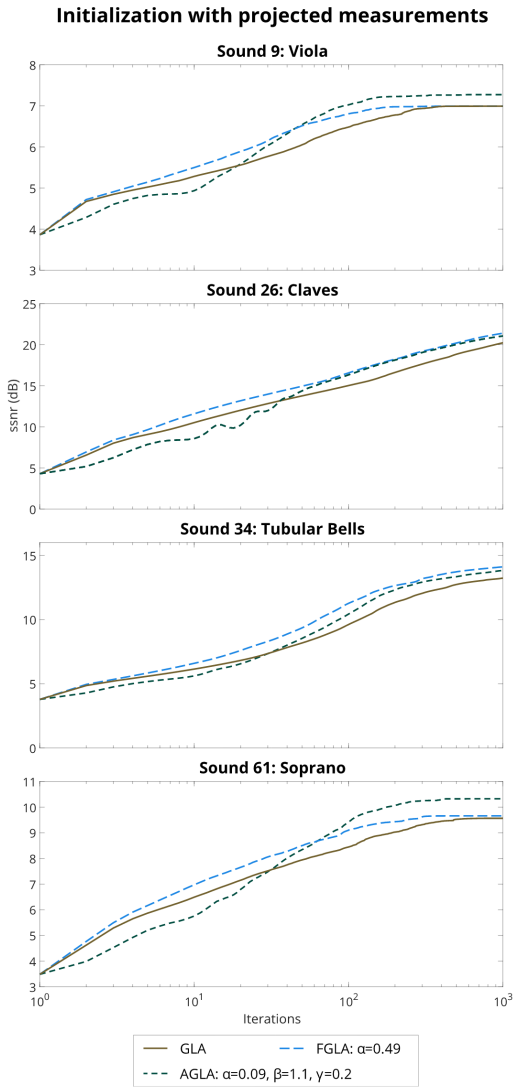

Fig. 7: Comparison of GLA, FGLA, and AGLA, with their respective parameter choice satisfying the convergence guarantee

We can observe that DM performed on average best, but it is also important to note that DM computes twice as many different projections per iterations compared to the other algorithms and the projections are the computationally most expensive segments of the algorithms. This was noticeable in the computational time per iteration for each signal, where the iterates of DM took at least 50% longer to compute compared to AGLA. We see that AGLA performed on average better than FGLA and is able to reach comparable SSNR to DM and RAAR. Over the test set we observed that, DM and RAAR tend to start oscillating after 100 iterations, whereas FGLA and AGLA stayed stable, an important trait in algorithms. With PGHI-initialization we observe that there was a slight increase in the intensity of the oscillations of DM and RAAR, but for FGLA and AGLA nearly none. These observations suggest that AGLA is an algorithm suitable for hybrid schemes, where one initializes with either a non-iterative scheme or the end point of another, maybe faster, scheme, after a fixed number of iterations as it has been studied for FGLA in [30]. Choosing the parameters of FGLA and AGLA to satisfy the convergence guarantee, can lead to slightly accelerated and more consistent behaviour compared to GLA. We notice that for the chosen combination of AGLA, we can experience slightly lower SSNR in the beginning, which then catches up or even surpasses the other algorithms. This behaviour can be attributed to the relatively large value of $\beta = 1.1$ and the small value of $\gamma = 0.2$. As a result, the nonprojected sequence exhibits a significant inertial/momentum step, while the projected iterates are less influential. Consequently, we experience a slower performance in the beginning.

VII. Conclusion

In this paper we presented the *Accelerated Griffin-Lim algorithm* and proved convergence results for it and its predecessor, the *Fast Griffin-Lim algorithm*. If the parameters are chosen to fulfill the necessary conditions to guarantee convergence and the minimizing properties, both of them outperform the *Griffin-Lim algorithm*, making it now possible for them to replace this classical method as the standard and reliable phase retrieval algorithm for acoustic. We showed that one can expect good convergence behaviour of inertial based methods theoretically and practically in the phase retrieval setting. The numerical results indicate that there are parameter combinations outside of the convergence regimes for which the algorithm asserts fast behaviour, further sparking interest in studying these methods. The simulations indicate that the *Accelerated Griffin-Lim algorithm* has the possibility to perform similarly error-wise to *Relaxed Averaged Alternating Reflections* and *Difference Map*, the two other powerful retrieval methods. The

proposed method achieves better convergence behaviour in the sense of having fewer to none oscillations and performing better for good initialization. Given the good results of the *Accelerated Griffin-Lim algorithm* under the initialization with *Phase Gradient Heap Integration*, the exploration of hybrid schemes presents an intriguing question for future research. Furthermore, the performance of the proposed method under different audio quality measures warrants further investigation.

REFERENCES

[1] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Stft phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE J Sel Top*, vol. 10, no. 4, pp. 770–781, 2016.

[2] R. A. Bedoui, Z. Mnasri, and F. Benzarti, "Phase retrieval: Application to audio signal reconstruction," in *2022 19th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2022, pp. 21–30.

[3] O. Yurduseven, D. R. Smith, and T. Fromenteze, "Phase retrieval in frequency-diverse imaging," in *IEEE Int. Symp. Antennas Propag. & USNC/URSI National Radio Sci. Meet.*, 2018, pp. 1797–1798.

[4] Y. Shechtman, Y. Eldar, O. Cohen, H. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," *IEEE Signal Process. Mag.*, vol. 32, 02 2014.

[5] P. Bowlan and R. Trebino, "Phase retrieval and the measurement of the complete spatiotemporal electric field of ultrashort pulses," in *CLEO/QELS: Laser Sci. to Photonic Appl.*, 2010, pp. 1–2.

[6] M. Johansson, H.-S. Lui, A. Fhager, and M. Persson, "Electromagnetic source modeling using phase retrieval methods," in *XXXth URSI Gen. Assem. Sci. Symp.*, 2011, pp. 1–2.

[7] A. Marafioti, N. Holighaus, and P. Majdak, "Time-frequency phase retrieval for audio—the effect of transform parameters," *IEEE Trans. Signal Process.*, vol. 69, pp. 3585–3596, 2021.

[8] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, pp. 236 – 243, 1984.

[9] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–250, 1971.

[10] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[11] N. Perraudin, P. Balazs, and P. Søndergaard, "A fast Griffin–Lim algorithm," *IEEE Workshop Appl. Signal Process. Audio Acoust.*, pp. 1–4, 2013.

[12] M. Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, 1st ed. Springer Publishing Company, Incorporated, 2015.

[13] K. Yatabe, "Consistent ICA: Determined bss meets spectrogram consistency," *IEEE Signal Processing Letters*, vol. 27, pp. 870–874, 2020.

[14] A. Marafioti, P. Majdak, N. Holighaus, and N. Perraudin, "Gacela - a generative adversarial context encoder for long audio inpainting," *IEEE Journal of Selected Topics in Signal Processing*, in press.

[15] N. Saleem, J. Gao, M. Irfan, E. Verdu, and J. P. Fuente, "E2e-v2sresnet: Deep residual convolutional neural networks for end-to-end video driven speech synthesis," *Image and Vision Computing*, vol. 119, p. 104389, 2022.

[16] O. Christensen, *An Introduction to Frames and Riesz Bases*. Birkhäuser Cham, 2016.

[17] R. Balan, P. Casazza, and D. Edidin, "On signal reconstruction from absolute value of frame coefficients," *Proceedings of SPIE*, 2005.

[18] P. Grohs, S. Koppensteiner, and M. Rathmair, "Phase retrieval: Uniqueness and stability," *SIAM Review*, 2020.

[19] P. Balazs, D. Bayer, F. Jaillet, and P. Søndergaard, "The pole behaviour of the phase derivative of the short-time fourier transform," *Applied and Computational Harmonic Analysis*, vol. 40, 2015.

[20] B. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[21] R. I. Boţ, E. R. Csetnek, and S. C. Laszlo, "An inertial forward–backward algorithm for the minimization of the sum of two nonconvex functions," *EURO J. Comput. Optim.*, vol. 4, p. 3–25, 2016.

[22] S. C. Laszlo, "Forward-backward algorithms with different inertial terms for structured non-convex minimization problems," *J. Optim. Theory Appl.*, 2023.

[23] R. Rockafellar and R. Wets, *Variational analysis*, ser. Grundlehren Math. Wiss. Springer, 2011, vol. 317.

[24] H. Attouch and J. Bolte, "On the convergence of the proximal algorithm for nonsmooth functions involving analytic features," *Math. Program.*, vol. 116, p. 5–16, 22009.

[25] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, 08 2013.

[26] H. Attouch, J. Bolte, and B. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods," *Math. Program.*, vol. 137, pp. 91–129, 2013.

[27] G. Li and T. K. Pong, "Calculus of the exponent of kurdyka-łojasiewicz inequality and its applications to linear convergence of first-order methods," *Found. Comput. Math.*, vol. 18, no. 5, pp. 1199–1232, 2018.

[28] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The large time-frequency analysis toolbox 2.0," in *Sound, Music, and Motion*, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds. Cham: Springer International Publishing, 2014, pp. 419–442.

[29] R. Nenov, D.-K. Nguyen, and P. Balazs, "Faster than Fast: Accelerating the Griffin-Lim algorithm," *IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP)*, pp. 1–5, 2023.

[30] T. Peer, S. Welker, and T. Gerkmann, "Beyond Griffin-Lim: Improved iterative phase retrieval for speech," in *Int. Workshop Acoustic Signal Enhancement (IWAENC)*, 2022, pp. 1–5.

[31] D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Probl.*, vol. 21, 2004.

[32] T. Kobayashi, T. Tanaka, K. Yatabe, and Y. Oikawa, "Acoustic application of phase reconstruction algorithms in optics," *IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP)*, pp. 6212–6216, 2022.

[33] V. Elser, "Phase retrieval by iterated projections," *J. Opt. Soc. Am. A*, vol. 20, pp. 40–55, 2003.

[34] Z. Průša, P. Balazs, and P. Søndergaard, "A noniterative method for reconstruction of phase from stft magnitude," *IEEE/ACM IEEE Trans. Audio Speech Language Process.*, vol. 25, pp. 1154–1164, 05 2017.

[35] G. Garrigos, "Square distance functions are polyak-Łojasiewicz and vice-versa," *arXiv: 2301.10332*, 2023.

APPENDIX

Proof of Proposition II.1. In order to prove (7), we are going to look at a general property of the distances of complex numbers to elements on the one dimensional sphere. For given $a \in \mathbb{C}$ and $r > 0$, one can deduce by rewriting in polar coordinates that

$$\begin{aligned}\min_{|b|=r} |a-b|^2 &= \min_{\theta\in[0,2\pi)} \left|a - re^{i\theta}\right|^2 \\ &= \min_{\theta\in[0,2\pi)} |a|^2 + r^2 - 2|a|\, r\cos(\angle a - \theta).\end{aligned} \tag{33}$$

Since this minimum is attained at $\theta = \angle a$, we get

$$\begin{aligned}\min_{|b|=r} |a-b|^2 &= \left|a - re^{i\angle a}\right|^2 \\ &= ||a| - r|^2 \left|e^{i\angle a}\right|^2 = ||a| - r|^2.\end{aligned}$$

If $r = 0$, then $b = 0$ is the unique minimizer of (33). Therefore we can see that

$$\begin{aligned}d^2_{C_2}(c) &= \min_{v\in C_2} \|c - v\|^2 \\ &= \sum_{i=1}^{L} \min_{v_i\in\mathbb{C}, |v_i|=s_i} |c_i - v_i|^2 \\ &= \sum_{i=1}^{L} ||c_i| - s_i|^2 = \||c| - s\|^2,\end{aligned} \tag{34}$$

which proves the first statement. For the second statement, we use the definition (6) of P_{C_2} and (7)

$$\begin{aligned}\|c - P_{C_2}(c)\|^2 &= \sum_{i=1}^{L} |c_i - (P_{C_2}(c))_i|^2 \\ &= \sum_{c_i\neq 0} \left|c_i - c_i\frac{s_i}{|c_i|}\right|^2 + \sum_{c_i=0} |s_i|^2 \\ &= \sum_{c_i\neq 0} |c_i|^2 \left|1 - \frac{s_i}{|c_i|}\right|^2 + \sum_{c_i=0} |s_i|^2 \\ &= \sum_{c_i\neq 0} ||c_i| - s_i|^2 + \sum_{c_i=0} |s_i|^2 \\ &= \sum_{i=1}^{L} ||c_i| - s_i|^2 = d^2_{C_2}(c).\end{aligned}$$

Moreover, we see from (33) that $\theta = \angle\alpha$ is the unique minimizer when $a \neq 0$, thus $P_{C_2}(c)$ is the unique minimizer of (34) for $c \notin D$ by the calculations above. □

Proof of Proposition V.2. In [35], an explicit formula for the limiting subgradient of the power of a distance function is derived. By this result, we know that for $y \in \mathbb{R}^{M\times 2}$

$$\partial\left(\frac{1}{2}d^2_{C_2}\right)(y) = y - \overline{P}_{C_2}(y).$$

By Proposition II.1 we can deduce that for $y \notin D$

$$\partial\left(\frac{1}{2}d^2_{C_2}\right)(y) = \{y - P_{C_2}(y)\}$$

holds, since then $\overline{P}_{C_2}(y) = \{P_{C_2}(y)\}$ and that $\frac{1}{2}d^2_{C_2}(y) = \frac{1}{2}\|y - P_{C_2}(y)\|^2$. If we look at the definition of P_{C_2}, we notice that each component is continuously differentiable around a neighbourhood of any $y \notin D$, since $x \mapsto \frac{x}{|x|}$ is continously differentiable around a neighbourhood of any $x \neq 0$. Since $\|\cdot\|^2$ is continuously differentiable everywhere, we deduce that $\frac{1}{2}d^2_{C_2}$ is continuously differentiable around a neighbourhood of any $y \notin D$. Therefore $\partial\frac{1}{2}d^2_{C_2}(y) = \nabla\frac{1}{2}d^2_{C_2}(y)$ for $y \notin D$ by [23, Corrolary 9.19].

For δ_{C_1} the limiting subgradient is given by the orthogonal space $\partial\delta_{C_1}(y) = C_1^\perp$ for $y \in C_1$ by [23, Exercise 8.14]. Furthermore, using the sum rule forumla [23, Exercise 8.8], we can deduce for the objective function $f(y) = \delta_{C_1}(y) + \frac{1}{2}d^2_{C_2}(y)$ that for $y \in C_1 \setminus D$

$$\partial f(y) = C_1^\perp + y - P_{C_2}(y).$$

By the definition of the distance of the sum of sets to a point, we have that for $y \in C_1 \setminus D$

$$\begin{aligned}d_{\partial f(y)}(0) &\leq \min_{z\in C_1^\perp, u\in\overline{P}_{C_2}(y)} \|z + y - u\| \\ &\leq \min_{z\in C_1^\perp} \|z + y - P_{C_2}(y)\|\end{aligned} \tag{35}$$

By taking $z = P_{C_2}(y) - P_{C_1}(P_{C_2}(y)) \in C_1^\perp$, we get the conclusion from (35). □

Proof of Proposition V.6. Choose $m \in \mathbb{N}$ such that for all $n \geq m$ the vector $c_n \notin D$ and let $n \geq m$. By the decreasing property of Theorem IV.1 and by (17) we know that

$$F_{K_2}(c_n, t_n) + \kappa_1\|c_n - t_n\|^2 \leq F_{K_2}(c_{n-1}, t_{n-1}),$$

where $\kappa_1 = \frac{K_1 - K_2}{2\alpha^2}$ and thus proving the first statement, since $K_1 > K_2$. Using the triangle inequality we can see that by (27) and a similar argument as in Proposition V.2

$$\begin{aligned}d_{\partial F_n}(0) \leq \|c_n - P_{C_1}(P_{C_2}(c_n)) + K_2(c_n - t_n)\| \\ + K_2\|c_n - t_n\|,\end{aligned} \tag{36}$$

since $c_n \in C_1 \setminus D$. By (13) we know that

$$c_n - P_{C_1}(P_{C_2}(c_n)) = \frac{1}{\gamma}(c_n - t_{n+1}) + \frac{\gamma - 1}{\gamma}(c_n - d_n)$$

Furthermore, by the definition of the algorithm we see that $c_n - d_n = \frac{\alpha}{\alpha - \beta}(c_n - t_n)$. Combining this observations in (36) we see that

$$d_{\partial F_n}(0) \leq \frac{1}{\gamma}\|c_n - t_{n+1}\| + \mu\|c_n - t_n\|$$

with $\mu = \left|\frac{(\gamma-1)\alpha}{\gamma(\alpha-\beta)} + K_2\right| + K_2 > 0$. Furthermore using the triangle inequality we see that

$$d_{\partial F_n}(0) \leq \frac{1}{\gamma}\left\|t_n - t_{n+1}\right\| + \left(\frac{1}{\gamma} + \mu\right)\left\|c_n - t_n\right\|.$$

Since $c_n - t_n = \alpha(t_n - t_{n-1})$ we conclude that

$$d_{\partial F_n}(0) \leq \kappa_2(\Delta_{t_{n+1}} + \Delta_{t_n})$$

with $\kappa_2 = \max\{\frac{1}{\gamma}, (\frac{1}{\gamma} + \mu)\alpha\}$. □

Lemma A.1. *Suppose that $\gamma > 0$ and*

$$0 \leq 2\beta|1-\gamma| < 2 - \gamma \tag{37}$$

(i) For $0 < \gamma \leq 1$, if

$$0 < \alpha < \left(1 - \frac{1}{\gamma}\right)\beta + \frac{1}{\gamma} - \frac{1}{2} \tag{38}$$

then

$$K_1 > K_2 > 0,$$

where $K_1 := \frac{1-\gamma}{\gamma}(1+2\alpha+\alpha^2-\beta-\alpha\beta)+\frac{1}{\gamma}(1-\alpha-\alpha^2)$ and $K_2 := \frac{1-\gamma}{\gamma}(\alpha^2+\beta-\alpha\beta)+\frac{1}{\gamma}(\alpha-\alpha^2)$.
(ii) For $1 < \gamma < 2$, if

$$0 < \alpha < \frac{1}{2\beta(\gamma-1)+\gamma} - \frac{1}{2} \tag{39}$$

then

$$K_1 > K_2 > 0,$$

where $K_1 := \frac{1-\gamma}{\gamma}(1+2\alpha+\alpha^2+\beta+\alpha\beta)+\frac{1}{\gamma}(1-\alpha-\alpha^2)$ and $K_2 := \frac{1-\gamma}{\gamma}(\alpha^2-\beta-3\alpha\beta)+\frac{1}{\gamma}(\alpha-\alpha^2)$.

Proof. One can check that (37) guarantees the the feasibility of α in both (38) and (39). Moreover, we can see that (38) implies

$$2\gamma\alpha < 2 - \gamma + 2(\gamma-1)\beta,$$

from which we are going to show that it further yields $K_1 - K_2 > 0$. It remains to show that $K_2 > 0$. After some calculations, we see that it is equivalent to

$$\gamma\alpha^2 - (1 + (\gamma-1)\beta)\alpha + (\gamma-1)\beta < 0. \tag{40}$$

Observe that for every $0 < \gamma < 1$ it holds

$$\begin{aligned}\Delta_1 &= (1+(\gamma-1)\beta)^2 - 4\gamma(\gamma-1)\beta \\ &> (1+(\gamma-1)\beta)^2 > 0,\end{aligned}$$

Hence the inequality (40) holds for $\alpha > 0$ if and only if

$$0 < \alpha < \frac{1 + (\gamma-1)\beta + \sqrt{\Delta_1}}{2\gamma}.$$

Furthermore, if α fulfills (38), then for $0 < \gamma < 1$

$$\begin{aligned}2\gamma\alpha &< 2 - \gamma + 2(\gamma-1)\beta \\ &< 2 + 2(\gamma-1)\beta \\ &< 1 + (\gamma-1)\beta + |1 + (\gamma-1)\beta| \\ &< 1 + (\gamma-1)\beta + \sqrt{\Delta_1}.\end{aligned}$$

Therefore (38) implies $K_1 > K_2 > 0$ for this case. The result for $1 < \gamma < 2$ can be deduced similarly. □

Lemma A.2. *Let A, B and C be positive real numbers. Then $A \geq \frac{C^2}{B+C}$ implies*

$$\frac{9}{4}A \geq 2C - B.$$

Proof. Rewriting the given assumption leads to

$$\sqrt{(B+C)A} \geq C.$$

We can apply the weighted arithmetic-geometric mean inequality $\frac{2}{3}\alpha + \frac{3}{2}\beta \geq 2\sqrt{\alpha\beta}$ and get

$$\frac{2}{3}(B+C) + \frac{3}{2}A \geq 2C.$$

By rewriting this inequality and multiplying by $\frac{3}{2}$ we get the conclusion. □