Phase retrieval with Bregman divergences and application to audio signal recovery

Pierre-Hugo Vial, Paul Magron, Thomas Oberlin, Cédric Févotte

Phase retrieval with Bregman divergences and application to audio signal recovery*

Pierre-Hugo Vial†, Paul Magron†, Thomas Oberlin‡, Cédric Févotte†

Abstract

Phase retrieval (PR) aims to recover a signal from the magnitudes of a set of inner products. This problem arises in many audio signal processing applications which operate on a short-time Fourier transform magnitude or power spectrogram, and discard the phase information. Recovering the missing phase from the resulting modified spectrogram is indeed necessary in order to synthesize time-domain signals. PR is commonly addressed by considering a minimization problem involving a quadratic loss function. In this paper, we adopt a different standpoint. Indeed, the quadratic loss does not properly account for some perceptual properties of audio, and alternative discrepancy measures such as beta-divergences have been preferred in many settings. Therefore, we formulate PR as a new minimization problem involving Bregman divergences. We consider a general formulation that actually addresses two problems, since it accounts for the non-symmetry of these divergences in general. To optimize the resulting objective, we derive two algorithms based on accelerated gradient descent and alternating direction method of multiplier. Experiments conducted on audio signal recovery from either exact or modified spectrograms highlight the potential of our proposed methods for audio restoration. In particular, leveraging some of these Bregman divergences induce better performance than the quadratic loss when performing PR from highly degraded spectrograms.

1 Introduction

Data reconstruction from phaseless measurements is a problem that arises in various fields including X-ray crystallography [1], optics [2] and astronomy [3]. This task, hereafter termed phase retrieval (PR), is also ubiquitous in audio signal processing, where much research has focused on the processing of nonnegative time-frequency representations such as short-time Fourier transform (STFT) magnitude or power spectrograms. Processing STFT spectrograms results in discarding or not accounting for the phase information, therefore it is necessary to retrieve the missing phase in order to synthesize time-domain signals. Therefore, PR is of paramount importance for tasks that involve audio signal reconstruction from incomplete time-frequency observations. Consequently, it has attracted some attention for many applications such as speech enhancement [4, 5, 6], source separation [7, 8, 9, 10] or audio restoration (e.g., click removal [11] or time-frequency inpainting [12]).

PR consists in recovering a signal $\mathbf{x}^\star \in \mathbb{C}^L$ from nonnegative measurements $\mathbf{r} \approx |\mathbf{A}\mathbf{x}^\star|^d \in \mathbb{R}_+^K$, where $\mathbf{A} \in \mathbb{C}^{K\times L}$ is the measurement matrix and d is usually equal to 1 or 2, depending whether one considers magnitude or power measurements. This problem is inherently ill-posed as different signals can generate identical measurements. Thus, $\mathbf{x}^\star$ can only be recovered up to several ambiguities which depend on $\mathbf{A}$. In particular, the STFT magnitude of a considered signal cannot uniquely represent this signal without specific constraints or a priori knowledge about part of the samples [13]. For example, estimation is subject to a global phase ambiguity, as the magnitude spectrograms of $\mathbf{x}^\star$ and $c\mathbf{x}^\star$ are identical when $c \in \mathbb{C}$ and $|c| = 1$. PR is commonly formulated as a nonconvex minimization problem involving a quadratic loss function, as follows:

$$\min_{\mathbf{x}\in\mathbb{C}^L} \quad E(\mathbf{x}) := \|\mathbf{r} - |\mathbf{A}\mathbf{x}|^d\|_2^2. \tag{1}$$

Problem (1) may be tackled with conventional optimization algorithms such as gradient descent [14], [15], alternating projections [16, 17], majorization-minimization [18] and alternating direction method of multipliers (ADMM) [19, 20]. An extensive review of those algorithms from a numerical perspective

*This work is supported by the European Research Council (ERC FACTORY-CoG-6681839).

†IRIT, Université de Toulouse, CNRS, Toulouse, France (e-mail: firstname.lastname@irit.fr).

‡ISAE-SUPAERO, Université de Toulouse, France (e-mail: firstname.lastname@isae-supaero.fr).

can be found in [21]. Convex optimization approaches are also considered in [22, 23, 24, 25] by lifting the problem to a higher dimensional space (i.e., solving a constrained quadratic problem involving $\mathbf{x}\mathbf{x}^{\mathsf{H}}$) and relaxing the rank-one constraint. However, they are impracticable for processing audio signals, as they square the dimensionality of the problem [26]. The Griffin-Lim algorithm (GLA) [27], a variant of the Gerchberg-Saxton algorithm (GSA) [16] adapted to STFT measurements, is one of the most popular techniques in the audio literature and is generally considered as a baseline for signal recovery. PR has also been tackled using signal modeling [11, 28, 29] or deep neural networks [30]. However, optimization-based approaches remain efficient, provide theoretical guarantees and may still be used with model-based approaches [31].

Even though a considerable amount of research has been conducted to tackle the PR problem as described in (1), such an approach suffers from one drawback when it comes to audio. Indeed, it is well established that the quadratic loss is not the best-suited metric for evaluating discrepancies in the time-frequency domain. For instance, it does not properly characterize the perceptually-related properties of audio such as its large dynamic range [32].

As such, in this work we propose to replace the quadratic loss function in (1) by alternative divergences which are more appropriate for audio signal processing. We consider general Bregman divergences, a family of loss functions which encompasses the β-divergence [33, 34] and some of its well-known special cases, the general Kullback-Leibler (KL) and Itakura-Saito (IS) divergences. The latter divergences are acknowledged for their superior performance in nonnegative audio spectral decomposition [35, 36, 37]. Besides, these divergences naturally arise from a statistical perspective. For instance, minimizing the KL divergence between an observed spectrogram and a parametric one assumes that the observations follow a Poisson model. Similarly, minimizing the IS divergence implies a multiplicative Gamma noise model [37]. In order to be as general as possible, we consider any power d (we do not restrict to either 1 nor 2) and we account for the fact that these divergences are not symmetric in general, which actually leads to tackling two different problems. To optimize the resulting objective, we derive two algorithms, based on accelerated gradient descent [38] and ADMM [39]. We experimentally assess the potential of our approach for PR on music and speech restoration tasks. Our experimental results show that our proposed methods either compare favorably or outperform traditional methods based on the quadratic loss (in particular with highly degraded spectrograms).

The rest of the paper is organized as follows. Section 2 reviews several baseline algorithms for PR. Section 3 describes the PR problem extended to Bregman divergences and the two proposed algorithms. Section 4 presents the experimental results for audio signal recovery applications. Finally, Section 5 draws some concluding remarks. For the sake of generality, we assume $\mathbf{x}$ to be complex-valued everywhere in Sections 2 and 3. Transposition to the real-valued case is discussed in Section 4 and in the Appendix.

Mathematical notations:

- $\mathbf{A}$ (capital, bold font): matrix.
- $\mathbf{x}$ (lower case, bold font): time-domain signal.
- $\tilde{\mathbf{x}}$ (lower case, bold font, with tilde): time-frequency domain vector, such that $\tilde{\mathbf{x}} = \mathbf{A}\mathbf{x}$.
- z (regular): scalar.
- $|.|$, $\angle(.)$, $(.)^*$: magnitude, complex angle, and complex conjugate, respectively.
- $^{\mathsf{T}}$, $^{\mathsf{H}}$: transpose and Hermitian transpose, respectively.
- $\mathfrak{R}$, $\mathfrak{I}$: real and imaginary part functions.
- $\odot$, $\frac{\cdot}{\cdot}$, $(.)^d$: element-wise matrix or vector multiplication, division, and power, respectively.
- $\cdot^{-1}$: matrix inverse.
- $\mathbf{I}_K$: identity matrix of size K.
- $\mathcal{P}_{\mathcal{S}}$: projection operator on the set $\mathcal{S}$, defined by $\mathcal{P}_{\mathcal{S}}(y) = \underset{z \in \mathcal{S}}{\operatorname{argmin}} \|y - z\|_2$.

2 Related work

In this section, we present three state-of-the-art approaches related to our own contributions: alternating projections (Section 2.1), gradient descent (Section 2.2), and ADMM (Section 2.3). Note that PR being a non-convex optimization problem, the descent methods considered in this paper are influenced by the initialization.

2.1 Alternating projections

In the seminal work [27], the authors address the PR problem (1) with $d = 1$ and with $\mathbf{A}$ being the STFT operator. They propose to alternate projections on $\mathcal{M}$, the set of time-frequency coefficients whose magnitude is equal to the observed measurements, and $\mathcal{C}$, the set of *consistent* coefficients, that is, complex coefficients that correspond to the STFT of time-domain signals [40]. More formally, we have:

$$\mathcal{M} = \{\tilde{\mathbf{x}} \in \mathbb{C}^K \mid |\tilde{\mathbf{x}}| = \mathbf{r}\} \text{ and } \mathcal{C} = \{\tilde{\mathbf{x}} \in \mathbb{C}^K \mid \tilde{\mathbf{x}} = \mathbf{A}\mathbf{A}^\dagger\tilde{\mathbf{x}}\}, \tag{2}$$

where $\tilde{\mathbf{x}}$ is a vector of time-frequency coefficients and $\mathbf{A}^\dagger = (\mathbf{A}^\mathsf{H}\mathbf{A})^{\cdot -1}\mathbf{A}^\mathsf{H}$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}$ (which encodes the inverse STFT). When the window used in the STFT is self-dual (i.e., can be used both for analysis and synthesis), we have $\mathbf{A}^\mathsf{H}\mathbf{A} = \mathbf{I}$ and as such $\mathbf{A}^\dagger = \mathbf{A}^\mathsf{H}$ (see Appendix A for more details about the STFT and duality). We make such an assumption throughout the paper (without loss of generality). The two projections then write:

$$\mathcal{P}_\mathcal{M}(\tilde{\mathbf{x}}) = \mathbf{r} \odot \frac{\tilde{\mathbf{x}}}{|\tilde{\mathbf{x}}|} \text{ and } \mathcal{P}_\mathcal{C}(\tilde{\mathbf{x}}) = \mathbf{A}\mathbf{A}^\mathsf{H}\tilde{\mathbf{x}}. \tag{3}$$

Although $\mathcal{M}$ is not a subspace and is not convex, we still call $\mathcal{P}_\mathcal{M}$ a projection since it maps an element of $\mathbb{C}^K$ to its closest element in $\mathcal{M}$ (in the mean squared error sense), which is unique [41, 42] when $\mathcal{M}$ is defined as in (2). Alternating these projections results in GLA, which is proved to converge to a critical point of the quadratic loss in (1) [27]. Alternatively, this algorithm can also be obtained by majorization-minimization [18].

In [42], an accelerated version of GLA, termed Fast GLA (FGLA), is proposed with a Nesterov-like scheme with constant acceleration parameter. FGLA was shown experimentally to reach lower local minima of the problem (1) with $d = 1$, yet without theoretical convergence guarantee. Other improvements of GLA include real-time purposed versions [43, 44] and its extension to multiple signals for source separation [45, 46].

GLA is similar to GSA [16] as they are both alternating projection algorithms. They yet differ in that GSA uses the discrete Fourier transform (DFT) as the measurement operator and accounts for an additional constraint on the support of the time-domain signal to make the solution unique. For GLA, this constraint is not necessary as uniqueness can be obtained thanks to the redundancy of the STFT [47].

2.2 Gradient descent

In [14], Candès et al. address the PR problem (1) with power measurements (i.e., $d = 2$) and a general measurement matrix $\mathbf{A}$ (such as Gaussian random or DFT vectors). They propose to minimize the error E with a gradient method. As the objective function implies complex quantities but is not holomorphic (i.e., not complex-differentiable), the authors express the gradient using the Wirtinger formalism [48] detailed in Appendix B. This leads to:

$$\nabla E(\mathbf{x}) = \mathbf{A}^\mathsf{H}[(\mathbf{A}\mathbf{x}) \odot (|\mathbf{A}\mathbf{x}|^2 - \mathbf{r})]. \tag{4}$$

The gradient algorithm update then writes:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \mu_{t+1}\nabla E(\mathbf{x}_t), \tag{5}$$

where t is the iteration index and μ_t stands for the step size at iteration t. This approach is called the Wirtinger flow algorithm [14].

2.3 ADMM

2.3.1 Minimization problem

In [20], Liang et al. express PR as a constrained problem and introduce auxiliary variables $\mathbf{u}$ and $\boldsymbol{\theta}$ for magnitude and phase, leading to the following problem:

$$\min_{\mathbf{x}\in\mathbb{C}^L} \|\mathbf{r} - \mathbf{u}\|_2^2 \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}, \tag{6}$$

where $\mathbf{u} \in \mathbb{R}_+^K$ and $\boldsymbol{\theta} \in [0; 2\pi[^K$. From (6) one can derive the augmented Lagrangian:

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\lambda}) =& \|\mathbf{r} - \mathbf{u}\|_2^2 + \Re\left(\boldsymbol{\lambda}^{\mathsf{H}}(\mathbf{A}\mathbf{x} - \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}})\right) \\ &+ \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}\|_2^2,\end{aligned} \tag{7}$$

where $\boldsymbol{\lambda}$ is the vector of Lagrange multipliers corresponding to the constraint $\mathbf{A}\mathbf{x} = \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}$ and ρ is the penalty parameter. From this expression, the authors derive the following ADMM update rules:

$$\begin{aligned}\{\mathbf{u}_{t+1}, \boldsymbol{\theta}_{t+1}\} &= \operatorname*{argmin}_{\mathbf{u} \geq 0, \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}_t, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\lambda}_t), \\ \mathbf{x}_{t+1} &= \operatorname*{argmin}_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{u}_{t+1}, \boldsymbol{\theta}_{t+1}, \boldsymbol{\lambda}_t), \\ \boldsymbol{\lambda}_{t+1} &= \boldsymbol{\lambda}_t + \mathbf{A}\mathbf{x}_{t+1} - \mathbf{u}_{t+1} \odot e^{\mathrm{i}\boldsymbol{\theta}_{t+1}}.\end{aligned} \tag{8}$$

The first two updates are available in closed form as detailed in the original paper [20] and form a special case of our proposed algorithm presented in Section 3.3.

2.3.2 Feasibility problem

In [19], Wen et al. address PR with DFT measurements as a feasibility problem. Instead of (1), they consider the following formulation:

$$\text{find} \quad \mathbf{x} \in \mathbb{C}^L \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{S}_{\mathcal{F}} \cap \mathcal{S}_0, \tag{9}$$

where $\mathcal{S}_{\mathcal{F}}$ is the set of signals whose DFT magnitude is $\mathbf{r}$ and $\mathcal{S}_0$ is the set of signals respecting an additional constraint (in optics, a typical constraint is that the signal is real-valued and nonnegative). They derive the following ADMM updates:

$$\begin{aligned}\mathbf{x}_{t+1} &= \mathcal{P}_{\mathcal{S}_0}(\mathbf{y}_t - \mathbf{w}_t), \\ \mathbf{y}_{t+1} &= \mathcal{P}_{\mathcal{S}_{\mathcal{F}}}(\mathbf{x}_{t+1} + \mathbf{w}_t), \\ \mathbf{w}_{t+1} &= \mathbf{w}_t + \eta(\mathbf{x}_{t+1} - \mathbf{y}_{t+1}).\end{aligned} \tag{10}$$

The authors also note that when $\eta = 1$, this algorithm is equivalent to the hybrid input-output algorithm, which is well-known in optics [17].

In a similar fashion, Masuyama et al. [49] use ADMM to tackle PR with STFT measurements (like in GLA) as a feasibility problem:

$$\text{find} \quad \tilde{\mathbf{x}} \in \mathbb{C}^K \quad \text{s.t.} \quad \tilde{\mathbf{x}} \in \mathcal{M} \cap \mathcal{C}. \tag{11}$$

They derive the following updates:

$$\begin{aligned}\tilde{\mathbf{x}}_{t+1} &= \mathcal{P}_{\mathcal{M}}(\tilde{\mathbf{y}}_t - \tilde{\mathbf{w}}_t), \\ \tilde{\mathbf{y}}_{t+1} &= \mathcal{P}_{\mathcal{C}}(\tilde{\mathbf{x}}_{t+1} + \tilde{\mathbf{w}}_t), \\ \tilde{\mathbf{w}}_{t+1} &= \tilde{\mathbf{w}}_t + \tilde{\mathbf{x}}_{t+1} - \tilde{\mathbf{y}}_{t+1}.\end{aligned} \tag{12}$$

This algorithm will be referred to as GLADMM. One can note than when $\mathbf{w}$ and $\tilde{\mathbf{w}}$ are equal to 0, the algorithms defined by (10) and (12) are respectively equivalent to GSA and GLA.

3 Proposed methods

In this section, we first propose a generalization of problem (1) to the family of Bregman divergences (Section 3.1). Then, relying on some of the related works presented in Section 2, we derive two algorithms based on accelerated gradient descent (Section 3.2) and ADMM (Section 3.3).

3.1 Phase retrieval with general Bregman divergence

We propose to generalize the problem (1) by substituting the quadratic loss by a general Bregman divergence. A Bregman divergence $\mathcal{D}_\psi$ is defined from a generating function ψ as follows:

$$\mathcal{D}_\psi(\mathbf{y} \,|\, \mathbf{z}) = \sum_k \left[\psi(y_k) - \psi(z_k) - \psi'(z_k)(y_k - z_k)\right] \tag{13}$$

Divergence	$\mathcal{D}_\psi(y\|z)$	$\psi(z)$	$\psi'(z)$	$\psi''(z)$
Quadratic loss	$\frac{1}{2}(y-z)^2$	$\frac{1}{2}z^2$	z	1
Kullback-Leibler	$y(\log y - \log z) - (y-z)$	$z\log z$	$1+\log z$	z^{-1}
Itakura-Saito	$\frac{y}{z} - \log\frac{y}{z} - 1$	$-\log z$	$-z^{-1}$	z^{-2}
β-divergence ($\beta \neq \{0,1\}$)	$\frac{y^\beta}{\beta-1} - \frac{\beta y z^{\beta-1}}{\beta-1} + z^\beta$	$\frac{z^\beta}{\beta(\beta-1)} - \frac{z}{\beta-1} + \frac{1}{\beta}$	$\frac{z^{\beta-1}-1}{\beta-1}$	$z^{\beta-2}$

Table 1: Typical Bregman divergences generating functions with their first and second derivatives. The KL and IS divergences are limit cases of the β-divergence for $\beta = 1$ and $\beta = 0$, respectively. The quadratic loss is obtained for $\beta = 2$.

where ψ is a strictly-convex scalar function, continuously-differentiable on a closed convex definition domain with derivative ψ', see, e.g., [50]. We here further assume that ψ is twice-differentiable with second derivative ψ''. $\mathcal{D}_\psi$ is always convex with respect to its first argument, but not necessarily with respect to its second one [51].

The motivation for using Bregman divergences is two-fold. First, they encompass several divergences that are well suited for audio spectrograms such as KL or IS, as illustrated in Table 1. Second, writing those divergences under the form (13) will ease the derivations, as will be seen hereafter.

As $\mathcal{D}_\psi$ is not necessarily symmetric, we will tackle the two following formulations of the problem:

$$\min_{\mathbf{x}\in\mathbb{C}^L} \overrightarrow{J}(\mathbf{x}) := \mathcal{D}_\psi(\mathbf{r} \,|\, |\mathbf{A}\mathbf{x}|^d), \tag{14}$$

$$\min_{\mathbf{x}\in\mathbb{C}^L} \overleftarrow{J}(\mathbf{x}) := \mathcal{D}_\psi(|\mathbf{A}\mathbf{x}|^d \,|\, \mathbf{r}). \tag{15}$$

We will refer to problems (14) and (15) as "right PR" and "left PR" respectively.

3.2 Gradient descent and acceleration

Similarly to [14], we first propose a Wirtinger gradient descent algorithm to minimize the objective functions defined in (14) and (15). The gradients of a general Bregman divergence with respect to its first and second arguments are given by

$$\nabla_{\mathbf{z}}\mathcal{D}_\psi(\mathbf{y} \,|\, \mathbf{z}) = \psi''(\mathbf{z}) \odot (\mathbf{z} - \mathbf{y}), \tag{16}$$

$$\nabla_{\mathbf{y}}\mathcal{D}_\psi(\mathbf{y} \,|\, \mathbf{z}) = \psi'(\mathbf{y}) - \psi'(\mathbf{z}). \tag{17}$$

Using the chain rule [52], we obtain:

$$\nabla\overrightarrow{J}(\mathbf{x}) = (\nabla|\mathbf{A}\mathbf{x}|^d)^{\mathsf{H}}[\psi''(|\mathbf{A}\mathbf{x}|^d) \odot (|\mathbf{A}\mathbf{x}|^d - \mathbf{r})], \tag{18}$$

$$\nabla\overleftarrow{J}(\mathbf{x}) = (\nabla|\mathbf{A}\mathbf{x}|^d)^{\mathsf{H}}[\psi'(|\mathbf{A}\mathbf{x}|^d) - \psi'(\mathbf{r})], \tag{19}$$

where the derivative ψ' and second-derivative ψ'' are applied entrywise and $\nabla|\mathbf{A}\mathbf{x}|^d$ denotes the Jacobian of the multivariate function $\mathbf{x} \to |\mathbf{A}\mathbf{x}|^d$ (the Jacobian being the extension of the gradient for multivariate functions, we may use the same notation ∇).[1] Using differentiation rules for element-wise matrix operations [52], we have:

$$\nabla|\mathbf{A}\mathbf{x}|^d = \frac{d}{2}\,\mathrm{diag}(|\mathbf{A}\mathbf{x}|^{d-2} \odot (\mathbf{A}\mathbf{x}))\mathbf{A}. \tag{20}$$

Expressions of ψ, ψ' and ψ'' for some typical Bregman divergences are given in Table 1.

We rewrite the gradients (18) and (19) in the following compact form:

$$\nabla J(\mathbf{x}) = (\nabla|\mathbf{A}\mathbf{x}|^d)^{\mathsf{H}}\,\mathbf{z}, \tag{21}$$

[1] Note that the gradient is not defined at $\mathbf{x} = 0$ in some cases ($d = 1$ and/or $\beta \leq 1$). In practice, we compute the gradient using $|\mathbf{A}\mathbf{x} + \epsilon|$ with $\epsilon \ll 1$ instead of $|\mathbf{A}\mathbf{x}|$, which alleviates potential numerical issues. A more rigorous treatment that should involve subgradients where the objective is not differentiable is left for future work.

where J can be either $\overrightarrow{J}$ or $\overleftarrow{J}$ and

$$\text{for "right" PR, } \mathbf{z} = \psi''(|\mathbf{A}\mathbf{x}|^d) \odot (|\mathbf{A}\mathbf{x}|^d - \mathbf{r}), \tag{22}$$

$$\text{for "left" PR, } \mathbf{z} = \psi'(|\mathbf{A}\mathbf{x}|^d) - \psi'(\mathbf{r}). \tag{23}$$

As such and together with (20), we obtain:

$$\nabla J(\mathbf{x}) = \frac{d}{2}\,\mathbf{A}^{\mathsf{H}}\left[|\mathbf{A}\mathbf{x}|^{d-2} \odot (\mathbf{A}\mathbf{x}) \odot \mathbf{z}\right]. \tag{24}$$

Using a constant step-size μ, our generic gradient algorithm writes:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \mu \nabla J(\mathbf{x}_t). \tag{25}$$

Similarly as in FGLA [42], we furthermore use a Nesterov-like acceleration scheme [38] resulting in the following updates:

$$\begin{aligned} \mathbf{y}_{t+1} &= \mathbf{x}_t - \mu \nabla J(\mathbf{x}_t), \\ \mathbf{x}_{t+1} &= \mathbf{y}_{t+1} + \gamma(\mathbf{y}_{t+1} - \mathbf{y}_t), \end{aligned} \tag{26}$$

where γ is the acceleration parameter.
Remark: When considering a quadratic loss (i.e., $\psi(z) = \frac{1}{2}z^2$), problems (1), (14) and (15) become equivalent. In particular, when $d = 1$, both gradients (18)-(19) write:

$$\nabla J(\mathbf{x}) = \mathbf{x} - \mathbf{A}^{\mathsf{H}}\left(\mathbf{r} \odot \frac{\mathbf{A}\mathbf{x}}{|\mathbf{A}\mathbf{x}|}\right). \tag{27}$$

Generic gradient descent with step size equal to 1 thus yields:

$$\mathbf{x}_{t+1} = \mathbf{A}^{\mathsf{H}}\left(\mathbf{r} \odot \frac{\mathbf{A}\mathbf{x}_t}{|\mathbf{A}\mathbf{x}_t|}\right), \tag{28}$$

which is nothing but the GLA update given by alternating the projections in (3). This shows that GLA can be seen as a gradient descent applied to the PR problem (1).

3.3 ADMM algorithm

In a similar fashion as in [20], we propose to reformulate PR with Bregman divergences as a constrained problem. We detail hereafter the left PR problem, and a similar derivation can be conducted for its right counterpart. The problem rewrites:

$$\min_{\mathbf{x} \in \mathbb{C}^L} \mathcal{D}_\psi(\mathbf{r} \,|\, \mathbf{u}) \text{ subject to } (\mathbf{A}\mathbf{x})^d = \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}, \tag{29}$$

from which we obtain the augmented Lagrangian:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\lambda}) =& \mathcal{D}_\psi(\mathbf{r} \,|\, \mathbf{u}) + \Re\left(\boldsymbol{\lambda}^{\mathsf{H}}((\mathbf{A}\mathbf{x})^d - \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}})\right) \\ &+ \frac{\rho}{2}\left\|(\mathbf{A}\mathbf{x})^d - \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}\right\|_2^2, \end{aligned}$$

where ρ is the penalty parameter. The first step of our ADMM algorithm consists in updating the values of $\mathbf{u}$ and $\boldsymbol{\theta}$ given $\mathbf{x}_t$ and $\boldsymbol{\lambda}_t$:

$$\{\mathbf{u}_{t+1}, \boldsymbol{\theta}_{t+1}\} = \operatorname*{argmin}_{\mathbf{u} \geq 0, \boldsymbol{\theta}} \mathcal{L}(\mathbf{x}_t, \mathbf{u}, \boldsymbol{\theta}, \boldsymbol{\lambda}_t). \tag{30}$$

This problem can be equivalently formulated as:

$$\{\mathbf{u}_{t+1}, \boldsymbol{\theta}_{t+1}\} = \operatorname*{argmin}_{\mathbf{u} \geq 0, \boldsymbol{\theta}} \mathcal{D}_\psi(\mathbf{r} \,|\, \mathbf{u}) + \frac{\rho}{2}\|\mathbf{h}_t - \mathbf{u} \odot e^{\mathrm{i}\boldsymbol{\theta}}\|_2^2, \tag{31}$$

with:

$$\mathbf{h}_t = (\mathbf{A}\mathbf{x}_t)^d + \frac{\boldsymbol{\lambda}_t}{\rho}. \tag{32}$$

With $\mathbf{u}$ fixed, the second term in (31) is minimized when the phase of $\mathbf{h}_t$ is equal to $\boldsymbol{\theta}$. Thus, $\boldsymbol{\theta}$ is updated as follows:

$$\boldsymbol{\theta}_{t+1} = \angle \mathbf{h}_t. \tag{33}$$

The problem in $\mathbf{u}$ can then be formulated as:

$$\mathbf{u}_{t+1} = \underset{\mathbf{u}\geq 0}{\operatorname{argmin}} \quad \mathcal{D}_\psi(\mathbf{r} \,|\, \mathbf{u}) + \frac{\rho}{2}\||\mathbf{h}_t| - \mathbf{u}\|_2^2. \tag{34}$$

As shown in Appendix C, the minimization problem involved in (34) remains unchanged when the positivity constraint on $\mathbf{u}$ is disregarded. The $\mathbf{u}$ update can therefore be written

$$\mathbf{u}_{t+1} = \operatorname{prox}_{\rho^{-1}\mathcal{D}_\psi(\mathbf{r}\,|\,\cdot)}(|\mathbf{h}_t|), \tag{35}$$

where prox_f denotes the proximal operator of a convex function f. The expressions of prox_f for some of the divergences considered in our experiments are given in Appendix C. The second step of our ADMM algorithm consists in updating the value of $\mathbf{x}$:

$$\mathbf{x}_{t+1} = \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{x}, \mathbf{u}_{t+1}, \boldsymbol{\theta}_{t+1}, \boldsymbol{\lambda}_t). \tag{36}$$

This problem can be rewritten as:

$$\mathbf{x}_{t+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\| (\mathbf{A}\mathbf{x})^d - \mathbf{u}_{t+1} \odot e^{\mathrm{i}\boldsymbol{\theta}_{t+1}} + \frac{\boldsymbol{\lambda}_t}{\rho} \right\|_2^2, \tag{37}$$

and has the following solution:

$$\mathbf{x}_{t+1} = \mathbf{A}^{\mathsf{H}}\big(\mathbf{u}_{t+1} \odot e^{\mathrm{i}\boldsymbol{\theta}_{t+1}} - \frac{\boldsymbol{\lambda}_t}{\rho}\big)^{1/d}. \tag{38}$$

The final step of our ADMM algorithm consists in updating the Lagrange multipliers $\boldsymbol{\lambda}$, as follows:

$$\boldsymbol{\lambda}_{t+1} = \boldsymbol{\lambda}_t + \rho(\mathbf{A}\mathbf{x}_{t+1} - \mathbf{u}_{t+1} \odot e^{\mathrm{i}\boldsymbol{\theta}_{t+1}}). \tag{39}$$

The whole ADMM procedure then consists in iteratively applying the updates given by (35), (38) and (39).

The derivation of the updates for the left PR problem is similar, and the resulting algorithm is unchanged, except for the update of $\mathbf{u}$ in (35), which becomes:

$$\mathbf{u}_{t+1} = \operatorname{prox}_{\rho^{-1}\mathcal{D}_\psi(\cdot\,|\,\mathbf{r})}(|\mathbf{h}_t|). \tag{40}$$

3.4 Implementation

We have presented gradient descent and ADMM algorithms for phase retrieval in the general case. We now address some specificities of audio signal recovery from a phaseless spectrogram, i.e., when $\mathbf{A}$ is the STFT matrix and $\mathbf{x}$ is real-valued. The STFT matrix $\mathbf{A}$ and its inverse are large structured matrices that allow for efficient implementations of matrix-vector products of the forms $\mathbf{A}\mathbf{x}$ and $\mathbf{A}^{\mathsf{H}}\mathbf{y}$. In that setting, it is more customary to handle time-frequency matrices of size $M \times N$, where M is the number of frequency channels and N the number of time frames, rather than vectors of size $K = MN$. As such, we provide in Algorithms 1 and 2 the pseudo-code for practical implementation of our accelerated gradient and ADMM algorithms, respectively, in the time-frequency audio recovery setting.

For generality, we assumed $\mathbf{x} \in \mathbb{C}^L$ in the previous sections. However, audio signals are real-valued and this deserves some comments. As shown in Appendix D, the estimates $\mathbf{x}_t$ remain real-valued under the following conditions. In a nutshell, a signal is real-valued if and only if its STFT $\mathbf{X} \in \mathbb{C}^{M\times N}$ is frequency-Hermitian, that is:

$$[\mathbf{X}]_{m,n} = [\mathbf{X}]^*_{M-m,n}. \tag{41}$$

When $\mathbf{R}$ is the spectrogram of a real-valued signal and when Algorithms 1 and 2 are initialized with a frequency-Hermitian matrix $\mathbf{X}$, all the time-frequency matrices involved in the updates remain frequency-Hermitian (because operations only involve sum and element-wise product with frequency-Hermitian matrices). This in turn ensures that the variable $\mathbf{x}$ remains real-valued. As such, the STFT and inverse STFT (iSTFT) operations in Algorithms 1 and 2 need only return/process the first $\lfloor \frac{M}{2} \rfloor + 1$ frequency channels (usually termed "positive frequencies"), as customary with real-valued signals.

More rigorously, we may also re-derive our gradient and ADMM algorithms for $\mathbf{x} \in \mathbb{R}^L$, using real-valued differentiation instead of Wirtinger gradients (and involving the real and imaginary parts of $\mathbf{A}$ in the objective function). This is addressed in Appendix D which shows that we indeed obtain the same algorithms.

Algorithm 1: Accelerated gradient descent for PR with the Bregman divergence.

Inputs: Measurements $\mathbf{R} \in \mathbb{R}_+^{M \times N}$, initial phase $\boldsymbol{\phi}_0 \in [0, 2\pi[^{M \times N}_+$, step size μ and acceleration parameter γ.
Initialization:
$\mathbf{X} = \mathbf{R} \odot e^{\mathrm{i}\boldsymbol{\phi}_0}$
$\mathbf{x} = \mathrm{iSTFT}(\mathbf{X})$
$\mathbf{y}_{\mathrm{old}} = 0$
while *stopping criteria not reached* **do**
  $\mathbf{X} = \mathrm{STFT}(\mathbf{x})$
  if *PR left* **then**
    $\mathbf{Z} = \psi'(|\mathbf{X}|^d) - \psi'(\mathbf{R})$
  else if *PR right* **then**
    $\mathbf{Z} = \psi''(|\mathbf{X}|^d) \odot (|\mathbf{X}|^d - \mathbf{R})$
  $\mathbf{g} = \mathrm{iSTFT}(\frac{d}{2}\mathbf{X} \odot |\mathbf{X}|^{d-2} \odot \mathbf{Z})$
  $\mathbf{y} = \mathbf{x} - \mu\mathbf{g}$
  $\mathbf{x} = \mathbf{y} + \gamma(\mathbf{y} - \mathbf{y}_{\mathrm{old}})$
  $\mathbf{y}_{\mathrm{old}} = \mathbf{y}$
end
Output: $\mathbf{x}$

4 Experiments

In this section, we conduct experiments on PR tasks. We first assess the potential of the proposed algorithms for recovering signals from exact (i.e., non-modified) spectrograms. Then, we consider a PR task from modified spectrograms, as often encountered in audio applications. In the spirit or reproducible research, we will release the code related to those experiments along with the final version of the paper. Besides, audio examples of reconstructed signals are available online.[2]

4.1 Experimental setup

4.1.1 Data

As acoustic material, we use two corpora in our experiments. The first one, referred to as "speech", is composed of 10 utterances taken randomly from the TIMIT database [53]. The second one, referred to as "music", comprises 10 snippets from the Free Music Archive dataset [54]. All audio excerpts are single-channel, sampled at $22,050$ Hz and cropped to be 2 seconds-long. The STFT is computed with a 1024 samples-long (46 ms) self-dual sine bell window [55] (leading to an effective number of 513 frequency bins) and 50% percent overlap. We used the `librosa` Python package [56].

4.1.2 Methods

PR is conducted using the algorithms presented in Section 3 under different settings as described next.

Proposed gradient descent algorithm We experimented the accelerated gradient algorithm described in Alg. 1 in the following settings:

- KL ($\beta = 1$) for the "right" and "left" problems with $d \in \{1, 2\}$,
- $\beta = 0.5$ for the "right" and "left" problems and with $d \in \{1, 2\}$,
- IS ($\beta = 0$) for the "right" problem with $d = 2$,
- quadratic loss ($\beta = 2$) with $d \in \{1, 2\}$ (in that case the "right" and "left" problems are equivalent).

The "right" problems with KL, $d = 1$ on the one hand, and IS, $d = 2$ on the other hand, correspond to standard designs in NMF [37, 57]. The trade-off value $\beta = 0.5$ with either $d = 1$ or 2 has also been advocated in various papers, e.g., [58].

[2] `https://magronp.github.io/demos/jstsp21.html`

Algorithm 2: ADMM for PR with the Bregman divergence.

Inputs: Measurements $\mathbf{R} \in \mathbb{R}_+^{M \times N}$, initial phase $\boldsymbol{\phi}_0 \in [0, 2\pi[_+^{M \times N}$ and augmentation parameter ρ.
Initialization:
$\mathbf{X} = \mathbf{R} \odot e^{\mathrm{i}\boldsymbol{\phi}_0}$
$\mathbf{x} = \mathrm{iSTFT}(\mathbf{X})$
$\boldsymbol{\Lambda} = 0$
while *stopping criteria not reached* **do**
  $\mathbf{X} = \mathrm{STFT}(\mathbf{x})$
  $\mathbf{H} = \mathbf{X}^d + \frac{1}{\rho}\boldsymbol{\Lambda}$
  $\boldsymbol{\Theta} = \angle\mathbf{H}$
  if *PR left* **then**
    $\mathbf{U} = \mathrm{prox}_{\rho^{-1}\mathcal{D}_\psi(\cdot \,|\, \mathbf{r})}(|\mathbf{H}|)$
  else if *PR right* **then**
    $\mathbf{U} = \mathrm{prox}_{\rho^{-1}\mathcal{D}_\psi(\mathbf{r} \,|\, \cdot)}(|\mathbf{H}|)$
  $\mathbf{Z} = \mathbf{U} \odot e^{\mathrm{i}\boldsymbol{\Theta}}$
  $\mathbf{x} = \mathrm{iSTFT}((\mathbf{Z} - \frac{1}{\rho}\boldsymbol{\Lambda})^{1/d})$
  $\boldsymbol{\Lambda} = \boldsymbol{\Lambda} + \rho(\mathrm{STFT}(\mathbf{x}) - \mathbf{Z})$
end
Output: $\mathbf{x}$

The algorithms are used with constant step-size μ and acceleration parameter $\gamma = 0.99$ (like in [42]). The step-size is empirically set to the largest negative power of 10 enabling convergence for each loss and value of d in the setting of the experiments reported in Section 4.2. A summary of the parameter configurations and choice of loss functions is given in Table 2.

Proposed ADMM algorithm Applicability of ADMM is more limited than with gradient descent because it requires the expression of the proximal operators (35) and (40). We here consider the quadratic loss and "left" KL and IS problems. We set $d = 1$ and $\rho = 1$, which corresponds to the setting used by Liang et al. [20] for the quadratic loss (which thus falls as a special case of our setting).

Other baselines and parameters The previous algorithms are compared with the following other baselines: GLA, FGLA and GLADMM, presented in Section 2 and which use $d = 1$. All the algorithms (baseline and contributed) are run for 2500 iterations and initialized with the same uniform random phase.

4.1.3 Evaluation metrics

The reconstruction quality is evaluated in the time-frequency domain with the standard spectral convergence (SC) metric, defined as:

$$\mathrm{SC}(\mathbf{r}, \mathbf{x}) = \frac{\|\mathbf{r}^{1/d} - |\mathbf{A}\mathbf{x}|\|_2}{\|\mathbf{r}\|_2}. \tag{42}$$

PR performance is also assessed in the temporal domain using the signal-to-noise ratio (SNR):

$$\mathrm{SNR}(\mathbf{x}^\star, \mathbf{x}) = 20 \log_{10} \frac{\|\mathbf{x}^\star\|_2}{\|\mathbf{x}^\star - \hat{\mathbf{x}}\|_2}, \tag{43}$$

where $\hat{\mathbf{x}}$ is a scaled and time-shifted version of the estimate $\mathbf{x}$ returned by any algorithm. It is chosen to maximally correlate with the original signal $\mathbf{x}^\star$: as a result, the SNR is invariant to a global delay and scale factor. To better assess the performance of the algorithms, we display the SNR improvement over the initial estimate (that is, using a random phase).

Additionally, for the "speech" corpus, we also consider the short-term objective intelligibility (STOI) measure [59]. This score is obtained by first decomposing the clean and processed speech signals through a DFT-like filterbank, and then computing the correlation between the resulting representations. It has been shown to correlate well with subjective intelligibility measurements of speech, whether in clean or noisy conditions. It is computed with the `pystoi` library [60].

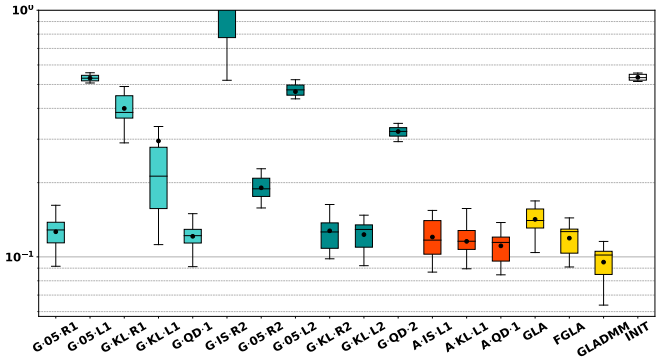
Algorithm	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent
Loss	β-div. ($\beta = 0.5$)	β-div. ($\beta = 0.5$)	Kullback-Leibler	Kullback-Leibler	Quadratic	Itakura-Saito
Direction	right	left	right	left	N/A	right
d	1	1	1	1	1	2
Hyperparameters	$\mu = 10^{-1}$	$\mu = 10^{-6}$	$\mu = 10^{-4}$	$\mu = 10^{-1}$	$\mu = 1$	$\mu = 10^{-7}$
Associated code	G·05·R1	G·05·L1	G·KL·R1	G·KL·L1	G·QD·1	G·IS·R2
Algorithm	Gradient descent	Gradient descent	Gradient descent	Gradient descent	Gradient descent	ADMM
Loss	β-div. ($\beta = 0.5$)	β-div. ($\beta = 0.5$)	Kullback-Leibler	Kullback-Leibler	Quadratic	Itakura-Saito
Direction	right	left	right	left	N/A	left
d	2	2	2	2	2	1
Hyperparameters	$\mu = 10^{-3}$	$\mu = 10^{-5}$	$\mu = 10^{-1}$	$\mu = 10^{-1}$	$\mu = 10^{-5}$	$\rho = 10^{-1}$
Associated code	G·05·R2	G·05·L2	G·KL·R2	G·KL·L2	G·QD·2	A·IS·L1
Algorithm	ADMM	ADMM	Griffin-Lim	Fast Griffin-Lim	GLADMM	Initialisation
Loss	Kullback-Leibler	Quadratic	(Quadratic)	(Quadratic)	(Indicator function)	N/A
Direction	left	N/A	N/A	N/A	N/A	N/A
d	1	1	1	1	1	N/A
Hyperparameter	$\rho = 10^{-1}$	$\rho = 10^{-1}$	N/A	N/A	N/A	N/A
Associated code	A·KL·L1	A·QD·1	GLA	FGLA	GLADMM	INIT

Table 2: Summary of setups considered in the experiments with their parameters (loss function, exponent d, type of algorithm and hyperparameter). Each setup is described by a code that follows this format: *algorithm·loss·direction-d*.

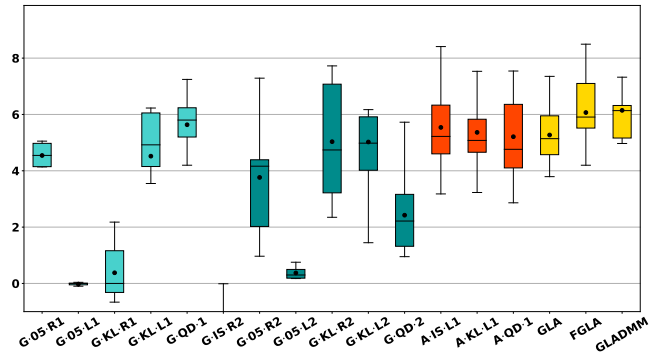
To summarize, SC is directly related to the PR quadratic loss problem (1), formulated in the time-frequency domain. On the other hand, the time-domain SNR and perceptual STOI are more related to the applicative needs. Note that for SC the lower the better, whereas for both SNR and STOI the higher the better.

4.2 PR from exact spectrograms

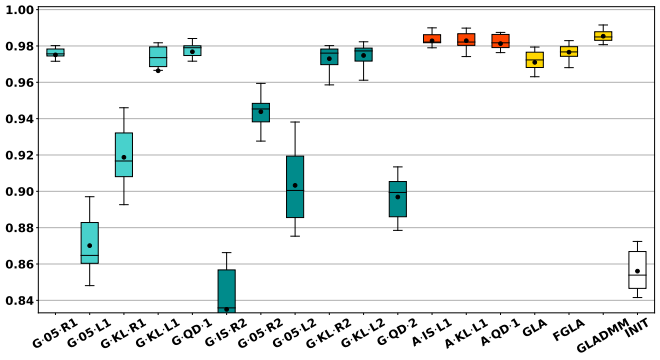
First, we consider a PR task conducted on exact spectrograms. In this setting, measurements are directly obtained from the ground truth signals $\mathbf{x}^\star$, such that $\mathbf{r} = |\mathbf{A}\mathbf{x}^\star|^d$. These measurements $\mathbf{r}$ are then fed as inputs to the algorithms described in 4.1.2.



(a) Spectral convergence

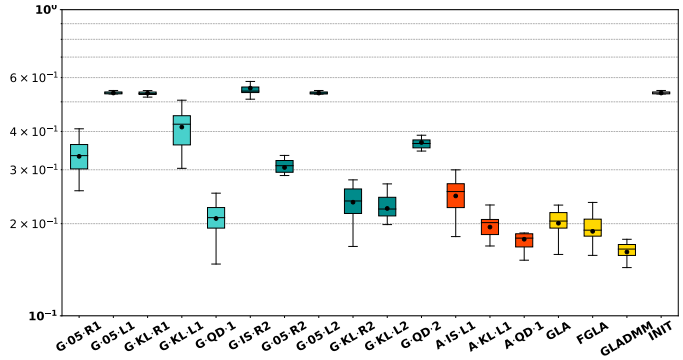


(b) Improvement in SNR

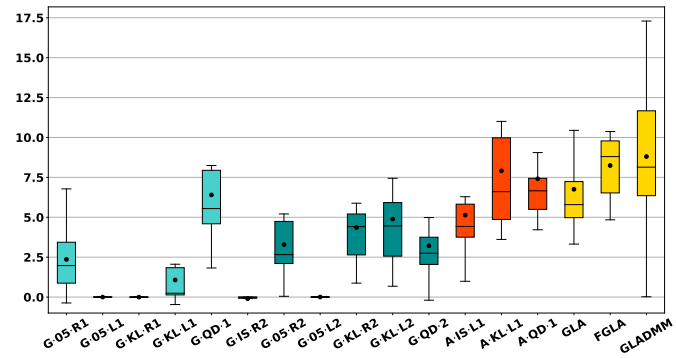


(c) STOI

Figure 1: Performance of PR from exact spectrograms for the "speech" corpus. Turquoise, orange and yellow respectively denote gradient descent algorithms, ADMM algorithms and GLA-like algorithms. For SC, a lower value corresponds to a better spectrogram approximation. For SNR improvement and STOI, a higher value corresponds to a better signal approximation. The boxes indicate the two middle quartiles among the ten excerpts, the middle bar is for the median, the dot for the mean, and the whiskers denote the extremal values.



(a) Spectral convergence



(b) Improvement in SNR

Figure 2: Performance of PR from exact spectrograms for the "music" corpus.

The results on the "speech" and "music" corpora are presented in Figures 1 and 2 respectively, from which overall similar conclusions can be drawn.

The best performances in terms of SC are achieved by GLADMM and algorithms optimizing the quadratic loss with $d = 1$. Note however that the advantage of quadratic loss-based algorithms against competing methods is less significant in terms of SNR and STOI. As recalled above, SC is directly related to the PR problem with quadratic loss (1) and consequently favors algorithms that directly tackle this problem.

A performance similar to that of quadratic loss-based algorithms is reached by some of the proposed alternative methods, such as the ADMM algorithms A·IS·L1 and A·KL·L1 and the gradient descent algorithms GD·05·R1, GD·KL·R2 and GD·KL·L2, in terms of SC, SNR and STOI (note that for the latter, the best performing methods exhibit a lower variance than the others). This outlines the potential of using alternative divergences to the quadratic loss.

Besides, we observe that the performance of these methods depend on a variety of factors. For instance, the difference between the performance reached by GD·KL·L2 and GD·KL·L1, or between GD·QD·1 and GD·QD·2 (for all metrics and both corpora) outlines the impact of d on the reconstruction quality. Likewise, considering a "left" rather than a "right" PR problem may yield very different results (see for instance the two corresponding gradient algorithms with $\beta = 0.5$ and $d = 1$).

Finally, for a given problem, the impact of the optimization strategy (i.e., ADMM vs. gradient descent) depends on the nature of the signals. For the "speech" corpus, ADMM algorithms (for KL and the quadratic loss) perform overall similarly to their gradient algorithms respective counterparts. However, for the "music" corpus, A·KL·L1 outperforms GD·KL·L1 in terms of both SC and SNR.

To summarize, when retrieving a signal from an exact spectrogram, GLADMM and quadratic-minimizing algorithms (with $d = 1$) seem to perform best. Some alternative methods yield competitive results, but require to carefully select the problem (power d, loss β, "right" or "left" formulation) and optimization strategy (ADMM vs. gradient descent), as well as considering the nature of the signals (speech or music). Note than when the data $\mathbf{r}$ is an exact spectrogram (i.e., $\mathbf{r} = |\mathbf{A}\mathbf{x}^\star|^d$), the loss functions (14) and (15) share the same minimum value 0 and global solution $\mathbf{x}^\star$ (up to ambiguities) for all ψ. This may explain why the somehow easier-to-optimize quadratic loss performs well. However this result is to be contrasted when using degraded spectrograms, as shown next.

4.3 PR from modified spectrograms

We now consider a PR task from modified spectrograms. In audio restoration applications such as source separation [61], audio inpainting [62] or time-stretching [63], the spectrogram that results from diverse operations does not necessarily correspond to the magnitude of the STFT of a signal. We propose to simulate this situation by modifying the spectrograms as in [49]. We add Gaussian white noise at various input SNRs to each excerpt in the "speech" corpus. We then apply an oracle Wiener filter [64] to the mixture in the STFT domain. This yields the magnitude spectrogram estimate $\mathbf{r}$ to which we apply the considered PR algorithms.

The results in terms of output SNR and STOI are presented in Figures 3 and 4, respectively. Note that we do not report the SC, since it is mostly impacted by the spectrogram deformation procedure, not by the subsequent PR task.

At high input SNR (0 to 10 dB), we observe a similar trend than in the previous experiment: GLADMM and quadratic loss-based algorithms (with $d = 1$) enable better reconstruction in terms of output SNR overall than other categories of algorithms. This confirms that such algorithms are appropriate for addressing the PR problem when the spectrograms are either exact or slightly degraded.

However, at lower input SNRs, we observe a different trend. For instance, at -20 dB input SNR, the best performing algorithms in terms of median SNR and STOI are GD·KL·L2 and GD·05·L2, respectively. Conversely, GLADMM which overall perform best at high input SNRs is outperformed by other GL-based or ADMM algorithms for an input SNR of -20 dB.

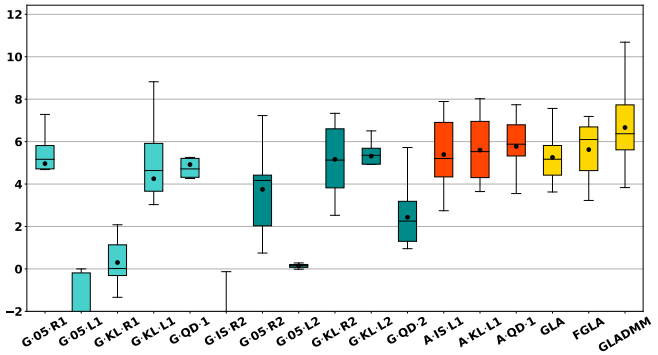
Besides, we observe that some algorithms based on alternative losses exhibit more robustness to the input noise. For instance, at all input SNRs, GD·KL·R2 compares favorably with ADMM algorithms in terms of output SNR. On the other hand, the performance of several algorithms, such as GD·KL·L1 and GD·IS·R2, significantly drops when more noise is added. Note that we experimentally observed some convergence problems at low input SNRs for those algorithms (which we recall were tuned using exact spectrogram data): the fine tuning of the gradient step becomes challenging as the level of input noise increases, and this needs to be more carefully investigated in future work.

Overall, the usefulness of PR with general Bregman divergences is revealed when the spectrograms are highly corrupted, as quadratic loss-based algorithms are outperformed by alternative loss-based algorithms in such a scenario. This might be explained by the ability of such divergences to better model and account for the nature of this destructive noise.

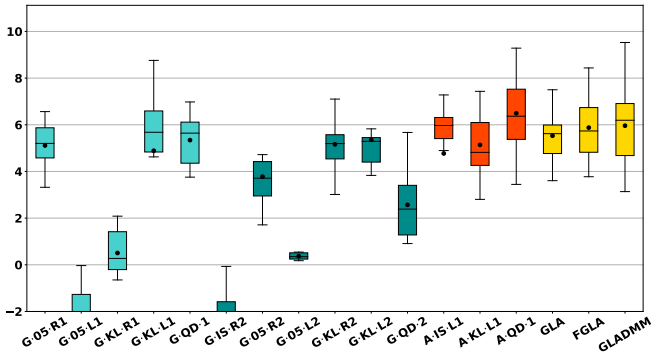
5 Conclusion

We have addressed the problem of PR when the quadratic loss is replaced by general Bregman divergences, a family of discrepancy measures with special cases that are well-suited for audio applications. We derived a gradient algorithm and an ADMM scheme for solving this problem and implemented them in the context of audio signal recovery. We evaluated the performance of these algorithms for PR from exact and modified spectrograms. We experimentally observed that when performing PR from exact or slightly degraded spectrograms, traditional algorithms based on the quadratic loss perform best. However, in the presence of high level of degradation, these are outperformed by algorithms based on alternative losses. This highlights the potential of PR with the Bregman divergence for audio signal recovery from highly corrupted spectrograms. However it is difficult to recommend a specific alternative divergence at this stage. The choice is dependent on the nature of the degradation and possibly on the nature of the data itself (e.g., speech vs music). Gradient algorithms are very convenient because they can be applied to any setting, however finding efficient step sizes in every setting was challenging and this issue deserves more attention. In that respect, our ADMM algorithms appeared more stable with respect to the level of noise and to the nature of the data but their applicability is more limited as they depend on the availability of specific proximal operators for each setting.

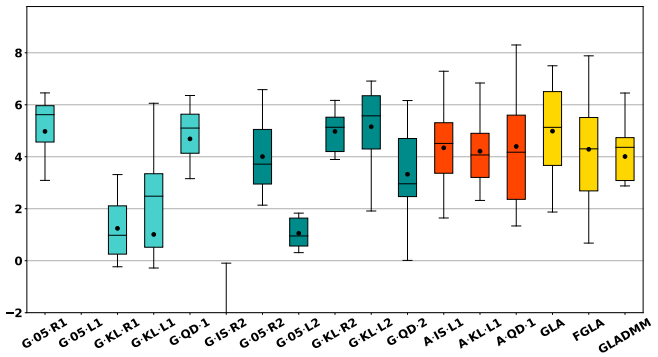
In future work, we intend to further improve the proposed gradient descent algorithms and to explore other optimisation strategies such as majorization-minimization. We also intend to tackle PR with non-quadratic measures of fit in frameworks where some additional phase information is available, such as in speech enhancement and source separation applications.



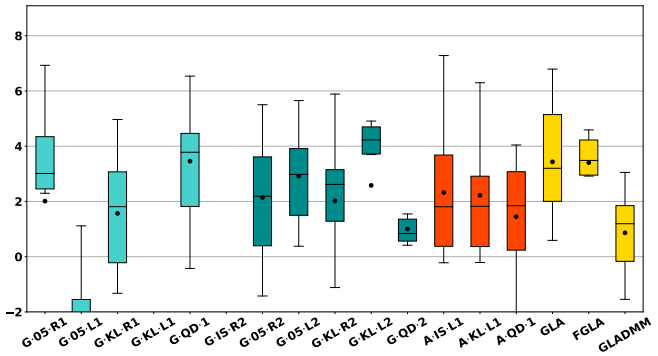
(a) +10dB



(b) 0dB

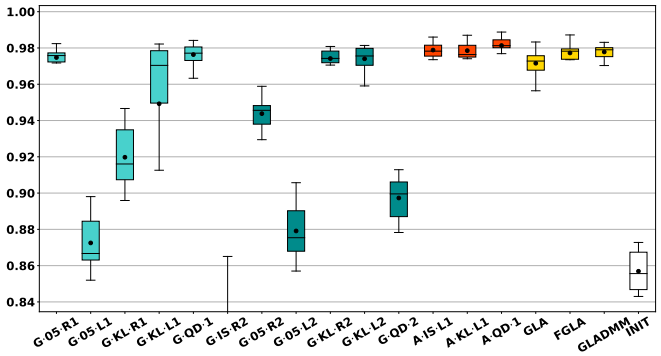


(c) -10dB

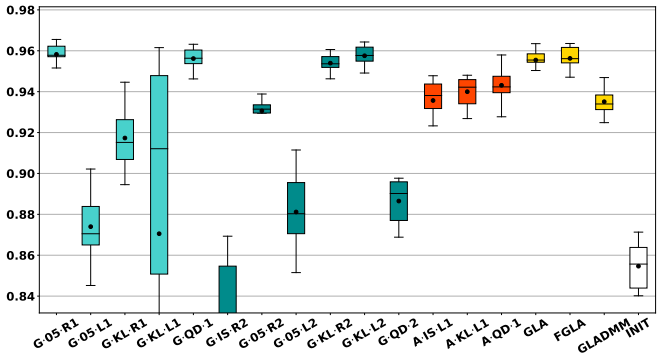


(d) -20dB

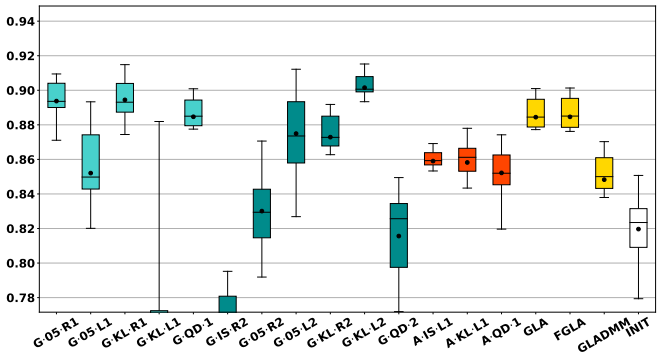
Figure 3: Performance of PR from modified speech spectrograms (measured by the improvement of output SNR) at various input SNRs.



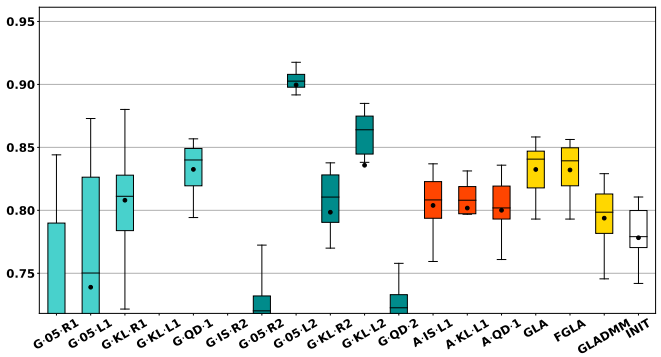
(a) +10dB



(b) 0dB



(c) -10dB



(d) -20dB

Figure 4: STOI for PR from modified speech spectrograms at various input SNRs.

A Short-time Fourier transform

Given a signal $\mathbf{x} \in \mathbb{C}^L$ and an *analysis window* $\mathbf{w} \in \mathbb{R}^T$ such that $T < L$, the discrete short-time Fourier transform (STFT) is a linear application $\mathcal{A}_w$ defined by

$$[\mathcal{A}_w \mathbf{x}](m,n) := \sum_{t=0}^{T-1} x(t+nH) w(t) e^{-\mathrm{i}2\pi \frac{m}{M} t}, \tag{44}$$

where

- $n = 0, \ldots, N-1$ indexes time frames,
- $m = 0, \ldots, M-1$ indexes frequency bins,
- H is a "hop" size.

The STFT essentially chops the signal $\mathbf{x}$ into windowed segments of size T and applies a DFT of size M to each segment. H controls the overlap between segments. H and M are user-defined. Generally, $H \leq T$. When T is even, $H = T/2$ corresponds to a 50% overlap between segments, which is a common choice. Generally, $M \geq T$ (more frequencies than samples). A common choice is $M = T$, which corresponds to using a standard "square" DFT. The value of N is determined by the length of the signal L, the length of the window T and the hop-size H. Common practice consists in zero-padding the signal x with $T - H$ zeroes at the beginning and as many zeroes as needed at the end so that $L = (T-H) + NH = T + (N-1)H$. This is in particular needed to have perfect reconstruction at the borders when defining an inverse-operator. We here assume that the signal $\mathbf{x}$ (of length L) has undergone such zero-padding at its borders.

Given a time-frequency matrix $\mathbf{C} \in \mathbb{C}^{M \times N}$ and a *synthesis window* $\mathbf{v} \in \mathbb{R}^T$, an inverse-STFT can be defined through the linear application $\mathcal{S}_v$ defined by

$$[\mathcal{S}_v \mathbf{C}](\ell) := \frac{1}{M} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} c(m,n) v(\ell - nH) e^{\mathrm{i}2\pi \frac{m}{M}(\ell - nH)}, \tag{45}$$

where $\ell = 0, \ldots, L-1$. We use the convention that $w(t) = v(t) = 0$ whenever $t \notin [0, T-1]$. The inverse-STFT essentially applies an inverse DFT to each time-frame of $\mathbf{C}$ and overlap-adds the resulting temporal signals. The windows $w(t)$ and $v(t)$ are said to be *dual* whenever they satisfy the following condition:

$$\forall \ell\,, \quad \sum_{n=0}^{N-1} w(\ell - nH) v(\ell - nH) = 1. \tag{46}$$

In this case (and when $M \geq T$), perfect reconstruction is achieved, i.e.,

$$\mathcal{S}_v \mathcal{A}_w \mathbf{x} = \mathbf{x}. \tag{47}$$

The STFT can alternatively be written as a Gabor frame. Indeed, (44) can be written as the output of inner products between $\mathbf{x}$ and Gabor atoms $\mathbf{g}_{mn} \in \mathbb{C}^L$ defined as pure windowed complex exponential, such that

$$g_{mn}(\ell) = w(\ell - nH) e^{\mathrm{i}2\pi \frac{m}{M}(\ell - nH)}. \tag{48}$$

Ignoring the time-frequency ordering and collecting the Gabor atoms into the columns of an $L \times MN$ matrix $\mathbf{G}_w$, the STFT can equivalently be obtained by $\mathbf{G}_w^{\mathsf{H}} \mathbf{x}$ (and as such $\mathbf{G}_w^{\mathsf{H}}$ is equal to the matrix $\mathbf{A}$ used in the main body of the paper). Under general conditions [65], the matrix $\mathbf{G}_w$ defines a frame in the sense that there exists positive constants C_1 and C_2 such that

$$C_1 \|\mathbf{x}\|_2^2 \leq \|\mathbf{G}_w^{\mathsf{H}} \mathbf{x}\|_2^2 \leq C_2 \|\mathbf{x}\|_2^2. \tag{49}$$

Similarly, the synthesis operator $\mathcal{S}_v$ can be expressed as

$$\mathcal{S}_v \mathbf{C} = \mathbf{G}_v \mathbf{c} \tag{50}$$

where $\mathbf{c}$ is a vectorized version of $\mathbf{C}$. As such, the windows $w(t)$ and $v(t)$ are dual if and only if $\mathbf{G}_v \mathbf{G}_w^{\mathsf{H}} \mathbf{x} = \mathbf{x}$. When the same window can be used for analysis and synthesis with perfect reconstruction (an example being the "sinebell" window [65]), then it can be shown that $C_1 = C_2 = 1$ and $\mathbf{G}_w^{\mathsf{H}}$ defines a so-called *Parseval frame*. This last assumption holds everywhere in the main body of the paper (i.e., $\mathbf{A}^{\mathsf{H}} \mathbf{A} = \mathbf{I}$).

B Wirtinger formalism

A function f, which can be either complex- or real-valued, of a complex variable $x = x_r + \mathrm{i}x_i$ can be seen as a function of (x_r, x_i). The Wirtinger derivatives [66, 67, 48] are then defined as:

$$\begin{aligned}\frac{\partial f}{\partial x}(x) &:= \frac{1}{2}\left(\frac{\partial f}{\partial x_r}(x_r, x_i) - \mathrm{i}\frac{\partial f}{\partial x_i}(x_r, x_i)\right), \\ \frac{\partial f}{\partial x^*}(x) &:= \frac{1}{2}\left(\frac{\partial f}{\partial x_r}(x_r, x_i) + \mathrm{i}\frac{\partial f}{\partial x_i}(x_r, x_i)\right).\end{aligned} \tag{51}$$

In practice, computing the derivative of f with respect to x (resp. x^*) can be done using usual differentiation by treating x (resp. x^*) as a real variable with x^* (resp. x) treated as a constant [48, 68]:

$$\frac{\partial f}{\partial x} = \left.\frac{\partial f(x, x^*)}{\partial x}\right|_{x^*=\text{const.}}, \tag{52}$$

$$\frac{\partial f}{\partial x^*} = \left.\frac{\partial f(x, x^*)}{\partial x^*}\right|_{x=\text{const.}}. \tag{53}$$

Besides, if f is real-valued, the following property is verified:

$$\left(\frac{\partial f}{\partial x}\right)^* = \frac{\partial f}{\partial x^*}. \tag{54}$$

In a multivariate setting, the gradient of f is then defined as:

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_K}\right]^{\mathsf{H}}. \tag{55}$$

When f is additionnally real-valued, the following property holds from (51) and (54):

$$\nabla_{\mathbb{R}} f := \left[\frac{\partial f}{\partial x_{r,1}}, ..., \frac{\partial f}{\partial x_{r,K}}\right]^{\mathsf{H}} = 2\Re(\nabla f). \tag{56}$$

As an illustrative example, we derive the expression of the gradient in the Wirtinger Flow algorithm [14]. The loss is:

$$E(\mathbf{x}) = \frac{1}{2}\||\mathbf{A}\mathbf{x}|^2 - \mathbf{r}\|_2^2. \tag{57}$$

Applying the chain rule yields:

$$\nabla E(\mathbf{x}) = \big(\nabla(|\mathbf{A}\mathbf{x}|^2 - \mathbf{r})\big)^{\mathsf{H}}(|\mathbf{A}\mathbf{x}|^2 - \mathbf{r}). \tag{58}$$

Treating $\mathbf{x}^*$ as a constant like in (52), the first term is given by:

$$\begin{aligned}\nabla(|\mathbf{A}\mathbf{x}|^2 - \mathbf{r}) &= \nabla(|\mathbf{A}\mathbf{x}|^2) && (59)\\ &= \nabla((\mathbf{A}\mathbf{x})^* \odot (\mathbf{A}\mathbf{x})) && (60)\\ &= \mathrm{diag}(\mathbf{A}\mathbf{x})^*\nabla(\mathbf{A}\mathbf{x}) + \mathrm{diag}(\mathbf{A}\mathbf{x})\nabla\big((\mathbf{A}\mathbf{x})^*\big) && (61)\\ &= \mathrm{diag}(\mathbf{A}\mathbf{x})^*\nabla(\mathbf{A}\mathbf{x}) + 0 && (62)\\ &= \mathrm{diag}(\mathbf{A}\mathbf{x})^*\mathbf{A}. && (63)\end{aligned}$$

We finally obtain:

$$\begin{aligned}\nabla E(\mathbf{x}) &= \mathbf{A}^{\mathsf{H}}\mathrm{diag}(\mathbf{A}\mathbf{x})(|\mathbf{A}\mathbf{x}|^2 - \mathbf{r}) && (64)\\ &= \mathbf{A}^{\mathsf{H}}[(\mathbf{A}\mathbf{x}) \odot (|\mathbf{A}\mathbf{x}|^2 - \mathbf{r})]. && (65)\end{aligned}$$

C Proximal operators

C.1 Definition

The proximal operator of a convex function $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ is the operator mapping a vector $\mathbf{y} \in \mathbb{R}^K$ to the set of solutions of the following penalized optimization problem [69]:

$$\mathrm{prox}_{\rho^{-1}f}(\mathbf{y}) := \underset{\mathbf{x}\in\mathbb{R}^K}{\mathrm{argmin}} \quad f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{y}\|_2^2. \tag{66}$$

Divergence	Expression	Proximal operator
Quadratic	$\frac{1}{2\rho}\|\cdot-\mathbf{r}\|_2^2$	$\dfrac{\rho\mathbf{y}+\mathbf{r}}{\rho+1}$
KL right	$\rho^{-1}\mathcal{D}_{KL}(\mathbf{r}\,\|\,\cdot)$	$\frac{1}{2\rho}(\mathbf{y}-1\pm\sqrt{\Delta})$ with $\Delta := 4\rho\mathbf{r}+(1-\mathbf{y})^2$
KL left	$\rho^{-1}\mathcal{D}_{KL}(\cdot\,\|\,\mathbf{r})$	$\rho^{-1}\mathcal{W}(\rho\mathbf{r}\odot e^{\rho\mathbf{y}})$
IS left	$\rho^{-1}\mathcal{D}_{IS}(\cdot\,\|\,\mathbf{r})$	$\frac{1}{2\rho}(-\mathbf{r}^{-1}+\rho\mathbf{y}\pm\sqrt{\Delta'})$ with $\Delta' := 4\rho+(\mathbf{r}^{-1}-\rho\mathbf{y})^2$

Table 3: Proximal operators of some standard (convex) Bregman divergences. $\mathcal{W}$ is the Lambert W function (i.e., the inverse relation of $z \mapsto ze^z$) applied entry-wise.

C.2 Proximal operator of usual Bregman divergences

A closed-form expression of the proximal operator can be obtained for some of the usual Bregman divergences, such as the quadratic distance and the KL right and left divergences [69, 70].

These are summarized in Table 3.

To the best of our knowledge, the proximal operator of the IS divergence has not been derived in closed-form in the literature. Therefore, for the sake of completeness, we derive it hereafter. Let us consider ψ such that $\psi(z) = -\log z$. We consider the problem (66) with $f(\mathbf{x}) = D_\psi(\mathbf{x}\,|\,\mathbf{r})$.

Note that such a function is defined only for vectors with nonnegative entries. However, we can extend its definition domain to $\mathbb{R}^K$ by considering that $\mathcal{D}_\psi(\mathbf{x}\,|\,\mathbf{r}) = +\infty$ if $\mathbf{x} \notin \mathbb{R}_+^K$ [70]. We then search for $\mathbf{x}$ such that $\nabla Q(\mathbf{x}) = \mathbf{0}$, where $Q(\mathbf{x}) = D_\psi(\mathbf{x}\,|\,\mathbf{r}) + \frac{\rho}{2}\|\mathbf{x}-\mathbf{y}\|_2^2$. We have:

$$\nabla Q(\mathbf{x}) = \psi'(\mathbf{x}) - \psi'(\mathbf{r}) + \rho(\mathbf{x}-\mathbf{y}) \tag{67}$$

$$= \mathbf{r}^{-1} - \mathbf{x}^{-1} + \rho(\mathbf{x}-\mathbf{y}). \tag{68}$$

Therefore,

$$\nabla Q(\mathbf{x}) = \mathbf{0} \iff \mathbf{x}\odot\mathbf{r}^{-1} - \mathbf{1} + \rho\mathbf{x}\odot(\mathbf{x}-\mathbf{y}) = \mathbf{0} \tag{69}$$

$$\iff \rho\mathbf{x}^2 + (\mathbf{r}^{-1} - \rho\mathbf{y})\odot\mathbf{x} - \mathbf{1} = \mathbf{0}. \tag{70}$$

Finally:

$$\operatorname{prox}_{\rho^{-1}\mathcal{D}_\psi(\cdot\,|\,\mathbf{r})}(\mathbf{y}) = \frac{1}{2\rho}(-\mathbf{r}^{-1} + \rho\mathbf{y} \pm \sqrt{\Delta'}), \tag{71}$$

where $\Delta' := 4\rho + (\mathbf{r}^{-1} - \rho\mathbf{y})^2$.

C.3 Nonnegativity constraint in problem (35)

Here we prove that the nonnegativity constraint on $\mathbf{u}$ in problem (35) can be ignored. Let us first rewrite this problem into scalar form, as this problem is separable entrywise:

$$\operatorname*{argmin}_{u_k \geq 0} d_\psi(r_k\,|\,u_k) + \frac{\rho}{2}\|\,|h_k| - u_k\|^2. \tag{72}$$

where d_ψ denotes the Bregman divergence applied to scalars (instead of vectors). We will remove the index k in what follows for clarity. We aim to prove that:

$$\text{If } u < 0, \quad d_\psi(r\,|\,0) + \frac{\rho}{2}|h|^2 \;\leq\; d_\psi(r\,|\,u) + \frac{\rho}{2}||h| - u|^2, \tag{73}$$

If this inequality holds, then the minimizer of the function defined in (72) necessarily belongs to $\mathbb{R}_+$. Consequently, the nonnegativity constraint can be dismissed. Equation (73) rewrites:

$$\psi(r) - \psi(0) - \psi'(0)r + \frac{\rho}{2}|h|^2 \leq \psi(r) - \psi(u) - \psi'(u)(r-u) + \frac{\rho}{2}||h| - u|^2, \tag{74}$$

which is equivalent to:

$$\psi(0) - \psi(u) + r\psi'(0) - \psi'(u)(0-u) - r\psi'(u) + \frac{\rho}{2}[-2u|h| + u^2] \geq 0, \tag{75}$$

which finally rewrites:

$$\underbrace{d_\psi(0 \mid u)}_{\text{term 1}} + \underbrace{r(\psi'(0) - \psi'(u))}_{\text{term 2}} + \underbrace{\frac{\rho}{2}[-2u|h| + u^2]}_{\text{term 3}} \geq 0. \tag{76}$$

The latter inequality holds for the following reasons:

- Term 1 is nonnegative by nonnegativity of Bregman divergences.
- Term 2 is nonnegative by convexity of ψ and nonnegativity of r: ψ is convex, therefore ψ' is monotonically non-decreasing. As $u < 0$, $\psi'(u) \leq \psi'(0)$ and $r(\psi'(0) - \psi'(u)) \geq 0$.
- Term 3 is nonnegative because u is negative.

Therefore, (73) holds, which demonstrates that the nonnegativity constraint in (35) can be dismissed. Finally, using a similar proof, we can show that the same holds for the "left" PR problem.

D Algorithms derivations for real-valued signals

We here discuss the adaptation of our proposed gradient and ADMM algorithms to the specific case when the input signal is real-valued $\mathbf{x} \in \mathbb{R}^L$.

In this setting, the gradient algorithm can be easily deduced from its complex-valued counterpart. Indeed, since $\mathbf{x}$ is real-valued, the gradient of J simply reduces to $\nabla_{\mathbb{R}} J(\mathbf{x})$, as defined in Appendix B. According to the property (56), this gradient is given by:

$$\nabla_{\mathbb{R}} J(\mathbf{x}) = 2\mathfrak{R}(\nabla J(\mathbf{x})). \tag{77}$$

where $\nabla J(\mathbf{x})$ is computed using the Wirtinger derivatives.

Consequently, the gradient update rule is similar to the complex-valued case, up to a constant factor of 2 and with the difference that we only need to retain the real part after applying $\mathbf{A}^\mathsf{H}$ (in practice, the inverse STFT).

Regarding the ADMM algorithm, we need to address the following sub-problem, in lieu of (37):

$$\min_{\mathbf{x} \in \mathbb{R}^L} ||(\mathbf{A}\mathbf{x})^d - \mathbf{y}||_2^2. \tag{78}$$

where we note $\mathbf{y} = \mathbf{u}_{t+1} \odot e^{\mathrm{i}\boldsymbol{\theta}_{t+1}} - \frac{\boldsymbol{\lambda}_t}{\rho}$. Since we only use ADMM algorithms with $d = 1$ in our experiments, we focus hereafter on this setting. By using again (56), we compute the gradient of the loss in (78) and set it at 0:

$$2\mathfrak{R}(\mathbf{A}^\mathsf{H}\mathbf{A}\mathbf{x} - \mathbf{A}^\mathsf{H}\mathbf{y}) = 0. \tag{79}$$

This yields the following solution:

$$\mathbf{x} = (\mathfrak{R}(\mathbf{A}^\mathsf{H}\mathbf{A}))^{-1}\mathfrak{R}(\mathbf{A}^\mathsf{H}\mathbf{y}). \tag{80}$$

When using the STFT with a self-dual window we have $\mathbf{A}^\mathsf{H}\mathbf{A} = \mathbf{I}$ and the update becomes

$$\mathbf{x} = \mathfrak{R}(\mathbf{A}^\mathsf{H}\mathbf{y}). \tag{81}$$

It is the same update as in the complex-valued case (38) up to retaining the real part after applying the inverse STFT $\mathbf{A}^\mathsf{H}$.

References

[1] R. W. Harrison, "Phase problem in crystallography," *Journal of the Optical Society of America A*, vol. 10, no. 5, pp. 1046–1055, 1993.

[2] A. Walther, "The question of phase retrieval in optics," *Optica Acta: International Journal of Optics*, vol. 10, no. 1, pp. 41–49, 1963.

[3] C. Fienup and J. Dainty, "Phase retrieval and image reconstruction for astronomy," *Image recovery: theory and application*, vol. 231, pp. 231–275, 1987.

[4] M. Krawczyk and T. Gerkmann, "STFT phase improvement for single channel speech enhancement," in *Proc. of the International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2012.

[5] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Phase processing for single-channel speech enhancement: History and recent advances," *IEEE Signal Processing Magazine*, vol. 32, pp. 55–66, March 2015.

[6] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Communication*, vol. 81, pp. 1 – 29, July 2016.

[7] P. Magron, R. Badeau, and B. David, "Model-based STFT phase recovery for audio source separation," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, pp. 1095–1105, June 2018.

[8] Z.-Q. Wang, J. Le Roux, D. Wang, and J. R. Hershey, "End-to-end speech separation with unfolded iterative phase reconstruction," in *Proc. of Interspeech*, September 2018.

[9] G. Wichern and J. Le Roux, "Phase reconstruction with learned time-frequency representations for single-channel speech separation," in *Proc. of the International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

[10] Z. Wang, K. Tan, and D. Wang, "Deep learning based phase reconstruction for speaker separation: A trigonometric perspective," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.

[11] P. Magron, R. Badeau, and B. David, "Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration," in *Proc. of the European Signal Processing Conference (EUSIPCO)*, August 2015.

[12] A.-M. Krémé, V. Emiya, and C. Chaux, "Phase reconstruction for time-frequency inpainting," in *Proc. of the International conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, July 2018.

[13] S. Nawab, T. Quatieri, and J. Lim, "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, pp. 986–998, August 1983.

[14] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, pp. 1985–2007, April 2015.

[15] R. Decorsière, P. L. Søndergaard, E. N. MacDonald, and T. Dau, "Inversion of auditory spectrograms, traditional spectrograms, and other envelope representations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 46–56, January 2015.

[16] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237–246, 1972.

[17] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, pp. 2758–2769, August 1982.

[18] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: Phase retrieval via majorization-minimization," *IEEE Transactions on Signal Processing*, vol. 64, pp. 5174–5186, October 2016.

[19] Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inverse Problems*, vol. 28, p. 18, October 2012.

[20] J. Liang, P. Stoica, Y. Jing, and J. Li, "Phase retrieval via the alternating direction method of multipliers," *IEEE Signal Processing Letters*, vol. 25, pp. 5–9, January 2018.

[21] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," *arXiv preprint arXiv:2004.05788*, 2020.

[22] E. J. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, pp. 1241–1274, August 2013.

[23] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 199–225, 2013.

[24] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[25] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Recovering signals from the short-time Fourier transform magnitude," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[26] D. L. Sun and I. J. O. Smith, "Estimating a Signal from a Magnitude Spectrogram via Convex Optimization," in *Audio Engineering Society Convention 133*, October 2012.

[27] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, pp. 236–243, April 1984.

[28] Z. Průša, P. Balazs, and P. L. Søndergaard, "A noniterative method for reconstruction of phase from STFT magnitude," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, pp. 1154–1164, May 2017.

[29] Z. Průša and P. Rajmic, "Toward high-quality real-time signal reconstruction from STFT magnitude," *IEEE Signal Processing Letters*, vol. 24, pp. 892–896, June 2017.

[30] S. Takamichi, Y. Saito, N. Takamune, D. Kitamura, and H. Saruwatari, "Phase reconstruction from amplitude spectrograms based on von-Mises-distribution deep neural network," in *Proc. of the International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 2018.

[31] Y. Masuyama, K. Yatabe, Y. Koizumi, Y. Oikawa, and N. Harada, "Deep Griffin–Lim iteration," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019.

[32] R. Gray, A. Buzo, A. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, pp. 367–376, August 1980.

[33] A. Cichocki and S. Amari, "Families of Alpha- Beta- and Gamma- divergences: Flexible and robust measures of similarities," *Entropy*, vol. 12, pp. 1532–1568, June 2010.

[34] R. Hennequin, B. David, and R. Badeau, "Beta-divergence as a subclass of Bregman divergence," *IEEE Signal Processing Letters*, vol. 18, pp. 83–86, February 2011.

[35] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 1066–1074, March 2007.

[36] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural computation*, vol. 21, pp. 793–830, March 2009.

[37] P. Smaragdis, C. Févotte, G. J. Mysore, N. Mohammadiha, and M. Hoffman, "Static and dynamic source separation using nonnegative factorizations: A unified view," *IEEE Signal Processing Magazine*, vol. 31, pp. 66–75, May 2014.

[38] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate $o(1/k^2)$," *Dokl. Akad. Nauk SSSR*, vol. 269, pp. 543–547, 1983.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[40] J. Le Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in *Proc. of the ISCA Workshop on Statistical and Perceptual Audition (SAPA)*, September 2008.

[41] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization," *Journal of the Optical Society of America A*, vol. 19, pp. 1334–1345, July 2002.

[42] N. Perraudin, P. Balazs, and P. L. Søndergaard, "A fast Griffin-Lim algorithm," in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2013.

[43] G. T. Beauregard, X. Zhu, and L. Wyse, "An efficient algorithm for real-time spectrogram inversion," in *Proc. of the International Conference on Digital Audio Effects (DAFx)*, September 2005.

[44] X. Zhu, G. T. Beauregard, and L. Wyse, "Real-time iterative spectrum inversion with look-ahead," in *Proc. of the IEEE International Conference on Multimedia and Expo (ICME)*, July 2006.

[45] D. Gunawan and D. Sen, "Iterative phase estimation for the synthesis of separated sources from single-channel mixtures," *IEEE Signal Processing Letters*, vol. 17, pp. 421–424, May 2010.

[46] P. Magron and T. Virtanen, "Online spectrogram inversion for low-latency audio source separation," *IEEE Signal Processing Letters*, vol. 27, pp. 306–310, 2020.

[47] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "STFT phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 770–781, June 2016.

[48] P. Bouboulis, "Wirtinger's calculus in general Hilbert spaces," *arXiv preprint arXiv:1005.5170*, 2010.

[49] Y. Masuyama, K. Yatabe, and Y. Oikawa, "Griffin–Lim like phase recovery via alternating direction method of multipliers," *IEEE Signal Processing Letters*, vol. 26, pp. 184–188, January 2019.

[50] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *Journal of Machine Learning Research*, vol. 6, pp. 1705 –1749, December 2005.

[51] H. H. Bauschke and J. M. Borwein, "Joint and separate convexity of the Bregman distance," in *Studies in Computational Mathematics*, vol. 8, pp. 23–36, Elsevier, 2001.

[52] J. R. Magnus and H. Neudecker, "Matrix differential calculus with applications to simple, Hadamard, and Kronecker products," *Journal of Mathematical Psychology*, vol. 29, pp. 474–492, December 1985.

[53] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguistic data consortium*, 1993.

[54] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: A dataset for music analysis," in *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, October 2017.

[55] J. O. Smith III, *Spectral audio signal processing*. W3K publishing, 2011.

[56] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proc. of the Python in Science Conference*, vol. 8, pp. 18–25, 2015.

[57] C. Févotte, E. Vincent, and A. Ozerov, "Single-channel audio source separation with NMF: divergences, constraints and algorithms," in *Audio Source Separation* (S. Makino, ed.), pp. 1–24, Springer, 2018.

[58] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, pp. 528 – 537, March 2010.

[59] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *Proc. of the IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP)*, March 2010.

[60] M. Pariente, "Pystoi," 2018.

[61] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. Wiley, 2018.

[62] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 922–932, March 2012.

[63] J. Driedger and M. Müller, "A review of time-scale modification of music signals," *Applied science*, vol. 6, February 2016.

[64] A. Liutkus and R. Badeau, "Generalized Wiener filtering with fractional power spectrograms," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015.

[65] K. Gröchenig, *Foundations of time-frequency analysis*. Springer Science & Business Media, 2013.

[66] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.

[67] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel hilbert spaces and the complex kernel LMS," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 964–978, 2010.

[68] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," in *ECE275A - Lecture Supplement - Fall 2005*, 2005.

[69] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212, Springer, 2011.

[70] M. El Gheche, G. Chierchia, and J.-C. Pesquet, "Proximity operators of discrete information divergences," *IEEE Transactions on Information Theory*, vol. 64, pp. 1092–1104, February 2017.