Time-Frequency Fading Algorithms Based on Gabor Multipliers

Ama Marina Kreme, Valentin Emiya, Caroline Chaux, Bruno Torrésani

Time-frequency fading algorithms based on Gabor multipliers

A. Marina Krémé, Valentin Emiya, Caroline Chaux, *Senior member, IEEE*, and Bruno Torrésani, *Member, IEEE*

***Abstract*—This paper addresses the problem of suppressing or attenuating time-frequency localized sources from audio signals, using only the knowledge of their time-frequency localization property. This problem, termed *time-frequency fading*, is formulated as a quadratic optimization problem in the signal domain, with data fidelity and penalization terms in the time-frequency plane. This allows the time-frequency coefficients of the estimated signal to be closed to the time-frequency coefficients of the observed signal outside the support of the perturbation while at the same time having a reduced energy inside the support of the perturbation. A closed-form solution is obtained that may be written in terms of Gabor multipliers, i.e. linear operators defined by pointwise multiplication in the time-frequency space. This approach, combined with efficient low-rank approximation algorithms and a suitable heuristic for hyperparameter selection, is able to efficiently process realistic signals. The effectiveness of the proposed approach is demonstrated on several audio signals where perturbations are filtered out while leading to good signal reconstruction quality.**

I. Introduction

FILTERING is one of the fundamental techniques in signal processing. By filtering one often means linear time invariant (LTI) filtering, which has a simple characterization in time and/or frequency domains. We refer to [1] for an in-depth description, in view of audio applications. Sticking to linear filters, time varying (LTV) filtering is also fundamental, as most devices actually feature time variations. LTV filters form a much wider class than LTI filters, and may be represented in several different ways, for example as matrices as described in [1] for digital signals, or pseudo-differential operators in the mathematical literature [2] in the analog case.

Among LTV filters, filters that can be efficiently represented in a joint time-frequency domain have received significant interest in the last two decades. These include filters whose spectral characteristics (frequency response) vary slowly as a function of time, or filters designed to enhance, attenuate or cancel out components which possess specific localization properties in joint time-frequency domain (e.g., chirps with prescribed or parameterized frequency modulation laws, transient oscillatory waves). Several formulations for non-stationary Wiener filtering have been proposed, that rely on time-frequency representations such as the spectrogram [3] or Wigner distribution [4], [5], [6]. These generally rely on pointwise multiplication in the time-frequency domain prior to inversion, very much in the spirit of classical LTI filtering. Filtering in the time-frequency domain has been used for a wide range of applications including decision, detection, or time-frequency segmentation.

Time-frequency filtering exploits the ability of some time-frequency representations to efficiently represent signals of interest. Limited to linear transforms, usual choices are Gabor transform [7] or Short Time Fourier Transform [8], on which we will focus in the present paper, or wavelet transform, constant Q transform or S transform, the choice often depending on the application domain.

We are interested here in the problem of removing or attenuating time-frequency localized components in signals. Unlike Wiener-type filtering, almost no knowledge on the component to be removed or attenuated is available: we only exploit its time-frequency localization properties, i.e., an effective time-frequency support. This problem, named *time-frequency fading* (TFF) in this paper, can be formulated in several ways, such as denoising, source separation, time-frequency segmentation or time-frequency inpainting, each formulation leading to distinct strategies. We are specifically interested here in removing interference sources that are well localized in a specific region Ω in the time-frequency domain Λ. A simple (and often used in practice) approach is to multiply the time-frequency representation inside Ω with a small value and reconstruct. However the time-frequency representation modified in such a way is not consistent —i.e., it is not the time-frequency transform of any signal—, which often results in artifacts after reconstruction [9].

We propose a variational formulation, which can be summarized as follows. Let $\mathcal{T}$ denote a time-frequency transform, mapping any $\boldsymbol{x} \in \mathbb{C}^L$ to $\mathcal{T}\boldsymbol{x} \in l^2(\Lambda)$. Let us denote by $\boldsymbol{x}^o$ the target signal and $\boldsymbol{y}^o$ the signal to be cancelled out. $\boldsymbol{y}^o$ is assumed to be essentially localized in a specific region Ω (that may be the union of connected sub-regions Ω_p far away from each other) in the time-frequency domain Λ. The observed signal $\boldsymbol{z}$ is thus given by $\boldsymbol{z} = \boldsymbol{x}^o + \boldsymbol{y}^o$. The objective is, given Ω, to retrieve $\boldsymbol{x}^o$ from $\boldsymbol{z}$ or equivalently, to remove $\boldsymbol{y}^o$ from $\boldsymbol{z}$. We formulate the problem as the following quadratic optimization problem

$$\boldsymbol{x}_{\boldsymbol{\lambda}} = \underset{\boldsymbol{x}\in\mathbb{C}^L}{\operatorname{argmin}} \quad \|\mathcal{T}\boldsymbol{x} - \mathcal{T}\boldsymbol{z}\|_{\overline{\Omega}}^2 + \sum_{p=1}^{P} \lambda_p \|\mathcal{T}\boldsymbol{x}\|_{\Omega_p}^2 . \qquad (1)$$

where $\|\boldsymbol{x}\|_{\Omega}^2 := \sum_{k\in\Omega} |\boldsymbol{x}[k]|^2$ and $\boldsymbol{\lambda} = (\lambda_p)_p$, $\lambda_p > 0$. This optimization problem amounts to find an estimation $\boldsymbol{x}_{\boldsymbol{\lambda}}$

A. Marina Krémé, Caroline Chaux and Bruno Torrésani are with Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France (Email: ama-marina.kreme@univ-amu.fr, caroline.chaux@univ-amu.fr, bruno.torresani@univ-amu.fr).

A. Marina Krémé and Valentin Emiya are with Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France (Email: ama-marina.kreme@univ-amu.fr, valentin.emiya@lis-lab.fr).

A. Marina Krémé was supported by a PhD grant from Région Sud.

of $\boldsymbol{x}$ which best matches $\boldsymbol{x}$ outside Ω and whose energy within Ω is reduced. The problem (1) leads to a linear system, whose solution involves time-frequency multipliers (i.e. time-frequency filters defined by pointwise multiplication in the time-frequency domain), as shown in [10], where a spectral approach based upon eigenvalue decomposition of Gabor multipliers was proposed. However, this approach turns out to be computationally expensive with large signals. Furthermore, the approach was restricted to the case $P = 1$.

In this work, we build on the results of [10] and extend the approach to the case $P > 1$, still using Gabor multipliers. We propose several strategies to reduce the computational cost and thus enable real audio signal processing. We also propose a complete pipeline that involves the construction of sub-regions Ω_p and the estimation of optimal values for hyperparameters λ_p.

The paper is organized as follow: Section II is devoted to the introduction of Gabor tools while Section III presents the proposed method. The problems of erasing one region or several sub-regions are treated as well as efficient algorithmic issues. Numerical experiments are drawn in Section IV on various audio signals for which the proposed algorithm outperforms baseline approaches. Finally, Section V concludes the paper. More technical aspects are postponed to the appendix.

II. Elements of Gabor analysis

Gabor analysis is a particular instance of time-frequency analysis, that aims at representing signals (functions, sequences or vectors) as linear combinations of translated and modulated copies of a generic signal called window. The coefficients of the expansion provide an alternative representation of the object of interest, called a time-frequency representation. Gabor analysis has enjoyed important developments since its introduction in the seminal paper by D. Gabor [7]. We refer to [11] for a detailed account of mathematical developments, which also addresses extensions to operator theory that will be of interest here, and to [12], [13], [14] for signal processing related developments. Gabor analysis has found many applications in various areas of signal processing such as audio, speech and image processing, communication theory, radar and sonar. We give below a short account fo the main aspects that are important for this paper.

Let us first introduce some notations. Throughout this paper, matrices and vectors are represented using bold characters type (such as $\boldsymbol{X}$ or $\boldsymbol{x}$), and scalars are represented using normal characters; the ith entry of $\boldsymbol{x}$ is $\boldsymbol{x}[i]$, similar notations are used for matrix elements. The inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^H \boldsymbol{x}$, where H denotes Hermitian conjugation, and the standard norm is denoted by $\|\cdot\|_2$.

Given a matrix $\boldsymbol{X}$, we denote by $\|\boldsymbol{X}\|_2 = \sqrt{\mathsf{Tr}(\boldsymbol{X}^H \boldsymbol{X})}$ its Frobenius norm (with Tr denoting the trace) and by $\|\boldsymbol{X}\|$ its operator norm. Given a subset Ω of the matrix index set, the Frobenius norm of the restriction of $\boldsymbol{X}$ to Ω is denoted by $\|\boldsymbol{X}\|_\Omega = \|\boldsymbol{X} 1_\Omega\|_2$, 1_Ω being the indicator function of Ω.

A. Gabor analysis

We limit ourselves to finite dimensional situations and describe Gabor analysis on $\mathbb{C}^L$, where L is a positive integer. Let a and b be two divisors of L, we set $N = L/a$ and $M = L/b$ and introduce the time-frequency lattice $\Lambda^\circ = b\mathbb{Z}_M \times a\mathbb{Z}_N$ and the dimensionless lattice $\Lambda = \mathbb{Z}_M \times \mathbb{Z}_N$ (which we will use as index set for Gabor analysis). Let $\boldsymbol{g} \in \mathbb{R}^L$ be a nonzero vector, hereafter called the *analysis window*. The family of Gabor atoms associated with $(\boldsymbol{g}, \Lambda^\circ)$ is the family of vectors $\boldsymbol{g}_{mn} \in \mathbb{C}^L$ obtained by translations and modulations of $\boldsymbol{g}$ on the lattice Λ°. Namely, the Gabor atoms $\boldsymbol{g}_{mn}$ are defined by

$$\boldsymbol{g}_{mn}[l] = \boldsymbol{g}[l - na] e^{2i\pi mbl/M}, \quad l \in \mathbb{Z}_L$$

A family of Gabor atoms as above is a *Gabor frame* when there exists two constants $0 < A \leq B < \infty$ verifying, for any $\boldsymbol{x} \in \mathbb{C}^L$.

$$A\|\boldsymbol{x}\|_2^2 \leq \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |\langle \boldsymbol{x}, \boldsymbol{g}_{mn} \rangle|^2 \leq B\|\boldsymbol{x}\|_2^2 \tag{2}$$

The constants A and B are respectively the lower and upper frame bounds. When $A = B$, the Gabor frame is *tight*, and is called a Parseval frame when $A = B = 1$. The redundancy of the Gabor frame is the parameter The redundancy of the representation is the number $\mathsf{red} = L/ab = MN/L$, it has to be larger than 1.

The Discrete Gabor transform (DGT) $\mathcal{V}$ maps any $\boldsymbol{x} \in \mathbb{C}^L$ into a matrix $\mathcal{V}\boldsymbol{x} \in \mathbb{C}^{M \times N}$ of *Gabor coefficients*, defined by

$$\mathcal{V}\boldsymbol{x}[m, n] = \langle \boldsymbol{x}, \boldsymbol{g}_{mn} \rangle = \sum_{l=0}^{L-1} \boldsymbol{x}[l] g[l - na] e^{-2i\pi mbl/M}.$$

The adjoint operator $\mathcal{V}^H : \mathbb{C}^{M \times N} \longrightarrow \mathbb{C}^L$, also called synthesis operator is given as follows. For all $\boldsymbol{c} \in \mathbb{C}^{M \times N}$,

$$\mathcal{V}^H \boldsymbol{c} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \boldsymbol{c}[m, n] \boldsymbol{g}_{mn} \ .$$

The frame operator is defined by $\boldsymbol{S} = \mathcal{V}^H \mathcal{V}$. It is bounded, self-adjoint and semi positive definite by construction, and actually positive definite as a consequence of (2). Therefore $\boldsymbol{S}$ is invertible, which permits to reconstruct any $\boldsymbol{x} \in \mathbb{C}^L$ from its Gabor coefficients.

Of particular interest are the above mentioned Parseval frames, for which $\boldsymbol{S} = \mathbf{I}$, i.e., $\mathcal{V}^H$ is a left inverse of $\mathcal{V}$. In such situations, the Gabor expansion takes a very simple form: for all $\boldsymbol{x} \in \mathbb{C}^L$

$$\boldsymbol{x} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle \boldsymbol{x}, \boldsymbol{g}_{mn} \rangle \boldsymbol{g}_{mn} \ .$$

For the sake of simplicity, we will limit ourselves to this case throughout this paper. Parseval Gabor frames can easily be generated from any Gabor frame. Indeed, it can be shown that given a Gabor frame associated with $(\boldsymbol{g}, \Lambda^\circ)$, with frame operator $\boldsymbol{S}$, and defining a new window $\boldsymbol{\gamma} = \boldsymbol{S}^{-1/2} \boldsymbol{g}$, the Gabor frame associated with $(\boldsymbol{\gamma}, \Lambda^\circ)$ is a Parseval Gabor frame. In this paper we will focus on Parseval frames constructed in this way, starting from a Gaussian window and from a Hann window.

B. Gabor multipliers

Gabor multipliers are linear operators that perform time-varying signal filtering by pointwise multiplication in the Gabor domain.

Definition 2.1 (Gabor multiplier): Let $\boldsymbol{g}$ and Λ° be as above, and let $\mathbf{m} \in \mathbb{C}^{M\times N}$. The Gabor multiplier associated to $(\boldsymbol{g}, \Lambda^\circ)$ and $\mathbf{m}$, is the linear operator $\mathfrak{M}_\mathbf{m} = \mathcal{V}^H \mathbf{m} \mathcal{V}$, acting on $\boldsymbol{x} \in \mathbb{C}^L$ as

$$\mathfrak{M}_\mathbf{m} \boldsymbol{x} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{m}[m,n] \langle \boldsymbol{x}, \boldsymbol{g}_{mn} \rangle \boldsymbol{g}_{mn} \ .$$

$\mathbf{m}$ is called the mask (or time-frequency transfer function) of the Gabor multiplier[1].

Gabor multipliers form a particular class of time varying filters [15], [5], which has been shown to be relevant in various domains including time varying system identification [16] or modulation theory [17] or sound morphing. Approximation of linear systems by Gabor multipliers have been studied in various places (see e.g. [18], [19], [20] and references therein). A main property is that they are good at approximating so called underspread linear systems, i.e. linear systems that do not involve large time shifts and frequency shifts.

The class of Gabor multipliers considered in this paper enjoy useful properties, which we address below. We first start from generic properties.

Properties 2.1:

(i) If $\mathbf{m}$ is real-valued then $\mathfrak{M}_\mathbf{m}$ is self-adjoint. It is then diagonalizable, with real eigenvalues, and there exists an orthonormal basis of $\mathbb{C}^L$ formed by $\mathfrak{M}_\mathbf{m}$ eigenvectors.

(ii) The Gabor multiplier generated by $\mathbf{m} \equiv 1$ is a multiple of the identity operator if and only if $(\boldsymbol{g}, \Lambda^\circ)$ generates a tight Gabor frame.

(iii) For arbitrary $\mathbf{m} \in \mathbb{C}^{M\times N}$, then $\mathfrak{M}_\mathbf{m}$ defines a bounded operator with operator norm $\|\mathfrak{M}_\mathbf{m}\| \leq C\|\mathbf{m}\|_\infty$, for some constant C. In particular, if $\boldsymbol{g}$ and Λ° generate a Parseval frame, then $\|\mathfrak{M}_\mathbf{m}\| \leq \|\mathbf{m}\|_\infty$.

We will mainly focus on the so-called time-frequency localization operators, which are Gabor multipliers associated with a mask $\mathbf{m} = 1_\Omega$ equal to the indicator function of a region Ω of the time-frequency lattice Λ. The mask $\mathbf{m}$ being binary, and since we limit the present investigations to Parseval frames, the corresponding eigenvalues range between 0 and 1.

Time-frequency localization properties of corresponding eigenvectors (which can be seen as generalized prolate spheroidal functions) have been studied in [21], [15], see [22] and references therein for recent accounts. A main result is that when the area of the region Ω is large enough, eigenvectors with large eigenvalues (i.e. close to 1) tend to be concentrated within Ω, while eigenvectors with smaller eigenvalues tend to localize themselves outside Ω. Therefore, the rank (or effective rank, i.e. dimension of the subspace generated by eigenvectors with significant eigenvalues) grows with the area of Ω (we refer to [22] for more precise statements). We illustrate these properties in the next subsection.

[1]By abuse of notation we have used the same symbol to represent the mask and the operator of pointwise multiplication by the mask.

C. Numerical illustrations

We illustrate in this section the behavior of Gabor transform and Gabor multipliers in situations of interest here. We rely on dedicated toolboxes, namely the LTFAT Octave/Matlab toolbox [23], [24], and the Python version ltfatpy [25].

We first highlight the behavior of eigenvectors and eigenvalues of Gabor multipliers on a mask generated from a mixture of car engine sound and bird song with length 16384 (see Section IV for the full experimental setting). The mask was constructed by an appropriate thresholding of the Gabor transform of the mixture, followed by simple morphological operations that ensure sufficient connectivity of sub-regions (see Section IV-A below for details). The mask is displayed in Fig 1 (left panel), the eigenvalues of the corresponding Gabor multiplier are displayed on the right panel, sorted in decreasing order. Two windows have been considered, respectively generated from a Gaussian and Hann window (we recall that initial windows are modified so as to generate Parseval frames). As expected, eigenvalues range between 0 and 1. A plateau of values equal to 1 can be observed at the beginning, followed by a rapid decay. As an example, less than 1800 eigenvalues (out of 16384) are enough to reach a precision of 10^{-6} for the Gaussian-type window (2000 for Hann), which opens the door to significant dimension reduction. We also illustrate in Fig. 2 the time-frequency localization of eigenvectors related to large and smaller eigenvalues. It can be seen on the top row that eigenvectors associated with the top eigenvalue $\sigma[1]$ (equal to 1 with an excellent precision, namely 3.4×10^{-10} for the Gaussian type window, and 1.5×10^{-6} for the Hann type window) tend to localize in the same regions as one of the connected sub-regions of the mask (here sub-regions 3 and 5). Eigenvectors associated with lower eigenvalues are displayed on the bottom row, (we chose $\sigma[1540] \approx 2.254 \times 10^{-4}$ and $\sigma[1585] \approx 6.399 \times 10^{-5}$) tend to be localized within time-frequency "rings" around the connected components of the mask, the smaller the eigenvalue the larger the ring radius. In both rows, the frequency localization of the Hann type window seems poorer, as could be expected.

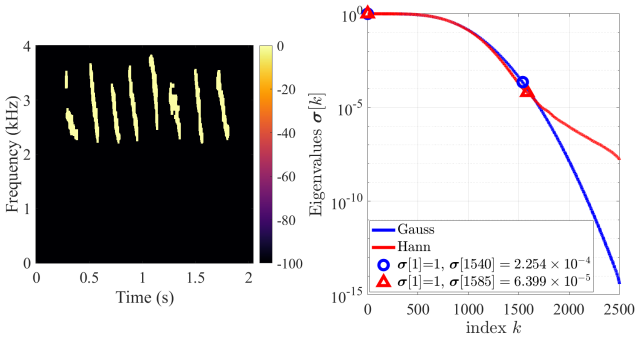

Fig. 1. Binary Mask generated with a Gauss window estimated from a bird song signal (left), and Gabor multiplier's eigenvalues $\sigma[k]$ in decreasing order, displayed in logarithmic scale (right). Eigenvectors associated with eigenvalues marked by dots on the right panel are displayed in Fig. 2

The relationship between the support size of the mask and the behavior of eigenvalues of the corresponding Gabor multiplier has been studied theoretically by several authors,

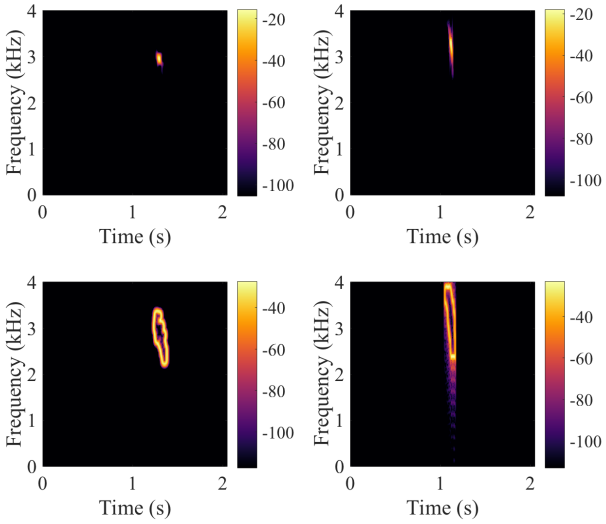

Fig. 2. Column 1 (top to bottom): spectrogram of the first and 1540 th eigenvectors of the Gabor multiplier generated with a Gauss window. Column 2 (top to bottom): spectrogram of the first and 1585 th eigenvectors of the Gabor multiplier generated with a Hann window. The mask associated with each of these multipliers was generated using the Gaussian type window, as shown in Fig 1.

in continuous time setting (see e.g. [22] and references therein). Assuming one may extrapolate to the discrete, finite-dimensional case, one would expect the number of eigenvalues larger than some threshold to vary (a least approximately) linearly with the support size. We checked this property numerically in situations such as the ones we consider in this paper.

More precisely, we computed a rank estimate for Gabor multipliers with rectangular masks of constant shape and increasing areas. Namely, we computed the number of eigenvalues above a fixed threshold (set here to 10^{-13}), as a function of the area of the support of the mask. We use a signal length $L = 16384$, a Gaussian window with length 64, a hop size $a = 16$ and $b = 256$ bins. Results are displayed in Fig. 3. Within the range of interest the rank estimate appears to vary linearly as a function of the mask area. Fig. 3 also includes another estimate called `Rand-EVD` and based upon random projections, which we use in Section III-C2 below. This estimate exhibits the same linear behavior.

III. Time-frequency fading (TFF)

A. Problem statement

As mentionned in Section I, we assume that we observe a signal $\boldsymbol{z}$ of length L, of the form

$$\boldsymbol{z} = \boldsymbol{x}^o + \boldsymbol{y}^o \ , \tag{3}$$

i.e., the sum of a signal of interest $\boldsymbol{x}^o$ and a perturbation $\boldsymbol{y}^o$. We assume further that the perturbation is strongly concentrated within a known region $\Omega \subset \Lambda$ in the time-frequency (TF) domain Λ. We denote $\overline{\Omega} = \Lambda \setminus \Omega$ the complementary region of Ω.

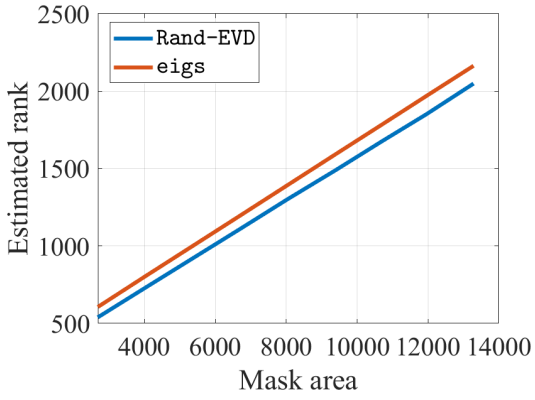

Fig. 3. Rank estimation as a function of mask area. The window is a *Gaussian* window.

Our objective is to filter out the perturbation $\boldsymbol{y}^o$ within Ω and reduce the artifacts generated by filtering. We term this problem *time-frequency fading* (TFF).

As can be intuited from Figure 1, region Ω may be structured into P sub-regions Ω_p where $1 \leq p \leq P$ by grouping time-frequency points that are close, given a closeness measure to be properly defined (see Section III-C1).

A standard practice for such a TFF problem is to perform a pointwise multiplication in the time-frequency domain by some transfer function, that penalizes the component within Ω, i.e. using a Gabor multiplier. However the choice of the transfer function is often arbitrary, and standard choices (such as binary masks) often generate artifacts.

We choose here to formulate the problem as an optimization problem, which we write as

$$\boldsymbol{x}_{\boldsymbol{\lambda}} = \underset{\boldsymbol{x}\in\mathbb{R}^L}{\operatorname{argmin}}\, f_{\boldsymbol{\lambda}}(\boldsymbol{x}) := \|\mathcal{V}\boldsymbol{x} - \mathcal{V}\boldsymbol{z}\|^2_{\overline{\Omega}} + \sum_{p=1}^{P} \lambda_p \|\mathcal{V}\boldsymbol{x}\|^2_{\Omega_p} \tag{4}$$

where the objective function $f_{\boldsymbol{\lambda}}$ depends on regularization parameters $\lambda_p > 0$ for all sub-regions $p = 1, \dots P$, which we collect in a set $\boldsymbol{\lambda} = \{\lambda_1, \dots \lambda_P\}$.

The first term of the objective function in (4) is a data fidelity term that matches the DGT of the estimated signal to that of the observation outside Ω. The second term controls its energy in each sub-region Ω_p, and the regularization parameters controls the trade-off among all terms.

The objective function in (4) is a quadratic form, its optimization leads to a linear system which we study below.

Remark 3.1: In the case $P = 1$, the problem can also be formulated in terms of the perturbation $\boldsymbol{y}$ in the observation model (3). Indeed, one may also write

$$g_\lambda(\boldsymbol{y}) = \frac{1}{\lambda} f_\lambda(\boldsymbol{z} - \boldsymbol{y}) = \|\mathcal{V}\boldsymbol{y} - \mathcal{V}\boldsymbol{x}\|^2_{\overline{\Omega}} + \frac{1}{\lambda}\|\mathcal{V}\boldsymbol{y}\|^2_{\overline{\Omega}}$$

so that optimizing f_λ with respect to $\boldsymbol{x}$ is equivalent to optimizing g_λ with respect to $\boldsymbol{y}$. Despite this apparent symmetry, the problems of reconstructing $\boldsymbol{x}$ and reconstructing $\boldsymbol{y}$ are not equivalent. We showed in [10] that the optimal values of λ are not the same for the two problems, mainly due to the imbalance of sizes of Ω and $\overline{\Omega}$.

B. Proposed solutions

1) Closed-form solution: Because $f_{\boldsymbol{\lambda}}$ is quadratic, solutions of the optimization problem (4) are given by corresponding critical points. The gradient of $f_{\boldsymbol{\lambda}}$ can be expressed in terms of Gabor multipliers

$$\nabla f_{\boldsymbol{\lambda}}(\boldsymbol{x}) = \mathfrak{M}_{\overline{\Omega}}\boldsymbol{x} - \mathfrak{M}_{\overline{\Omega}}\boldsymbol{z} + \sum_{p=1}^{P} \lambda_p \mathfrak{M}_{\Omega_p}\boldsymbol{x} \ , \tag{5}$$

where $\mathfrak{M}_{\mathcal{D}}$ denotes the Gabor multiplier associated with the indicator function of the domain $\mathcal{D}$.

Since we have limited our interest to the case of Parseval Gabor frames, we then have $\mathfrak{M}_{\Omega} + \mathfrak{M}_{\overline{\Omega}} = \mathbf{I}$. Therefore, the normal equations can be written as

$$\left(\mathbf{I} - \sum_{p=1}^{P}(1-\lambda_p)\mathfrak{M}_{\Omega_p}\right)\boldsymbol{x} = \sum_{p=1}^{P}\lambda_p\mathfrak{M}_{\Omega_p}\boldsymbol{z} \ . \tag{6}$$

The operator $\mathbf{I}-\sum_{p=1}^{P}(1-\lambda_p)\mathfrak{M}_{\Omega_p}$ is actually a Gabor multiplier with piecewise constant mask $\mathbf{m} = 1-\sum_{p=1}^{P}(1-\lambda_p)1_{\Omega_p}$, where 1_{Ω_p} is the indicator function of the sub-region Ω_p. If $\lambda_p > 0$ for all p, then $\min_{(m,n)}(\mathbf{m}[m,n]) > 0$ and the mask is therefore semi-normalized, which implies invertibility (see Proposition 3.7 in [26]).

From now on we then assume $\lambda_p > 0$ for all p. The solution then reads

$$\begin{aligned}\boldsymbol{x}_{\boldsymbol{\lambda}} &= \left(\mathbf{I} - \sum_{p=1}^{P}(1-\lambda_p)\mathfrak{M}_{\Omega_p}\right)^{-1}\left(\mathbf{I} - \sum_{p=1}^{P}\mathfrak{M}_{\Omega_p}\right)\boldsymbol{z} \\ &= \boldsymbol{z} - \left(\mathbf{I} - \sum_{p=1}^{P}(1-\lambda_p)\mathfrak{M}_{\Omega_p}\right)^{-1}\sum_{p=1}^{P}\lambda_p\mathfrak{M}_{\Omega_p}\boldsymbol{z}\end{aligned} \tag{7}$$

This closed-form solution requires solving a linear system for each choice of $\boldsymbol{\lambda}$, which may be computationally demanding in high dimensions and/or when the region Ω is large. Indeed, owing to the discussion at the end of subsection II-A, the rank of the multipliers increases with the area of the support of the mask. We propose below efficient algorithmic solutions that provide good approximations of (7).

2) Reduced rank approximation for $P = 1$: Let us first examine the case where a single region Ω is considered (for simplicity we then drop the subregion index). Since $P = 1$, a unique regularization parameter $\lambda \in \mathbb{R}_+^*$ is required. The solution reads

$$\boldsymbol{x}_{\lambda} = \boldsymbol{z} - \left(\mathbf{I} - (1-\lambda)\mathfrak{M}_{\Omega}\right)^{-1}\lambda\mathfrak{M}_{\Omega}\boldsymbol{z} \ , \tag{8}$$

the invertibility of $\mathbf{I} - (1-\lambda)\mathfrak{M}_{\Omega}$ being ensured by the fact that the spectrum of $\mathfrak{M}_{\Omega}$ is included in $[0,1]$ (see below).

As mentioned in Properties 2.1, the mask being binary, the Gabor multiplier $\mathfrak{M}_{\Omega}$ can be diagonalized and thus factored in matrix form as $\mathfrak{M}_{\Omega} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H$ where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma})$ is diagonal. Here $\boldsymbol{\sigma} = (\boldsymbol{\sigma}[1], \dots \boldsymbol{\sigma}[L]) \in \mathbb{R}^L$ is the vector of eigenvalues, which are sorted in decreasing order $1 \geq \boldsymbol{\sigma}[1] \geq \cdots \geq \boldsymbol{\sigma}[L] \geq 0$. We can then write

$$\boldsymbol{x}_{\lambda} = \boldsymbol{z} - \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^H\boldsymbol{z} \tag{9}$$

where $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$, with $\boldsymbol{\gamma} = (\boldsymbol{\gamma}[1], \dots \boldsymbol{\gamma}[L])$ and

$$\boldsymbol{\gamma}[i] = \frac{\lambda\boldsymbol{\sigma}[i]}{1-(1-\lambda)\boldsymbol{\sigma}[i]} \ .$$

It is easily seen that the sequence $\boldsymbol{\gamma} = (\gamma[1], \dots \gamma[L])$ is decreasing and included in $[0,1]$, which we will use for truncation purposes.

The approximate solution when truncating to the K largest eigenvalues ($K < L$) is given by:

$$\boldsymbol{x}_{\lambda}^{(K)} = \boldsymbol{z} - \boldsymbol{U}^{(K)}\boldsymbol{\Gamma}^{(K)}(\boldsymbol{U}^{(K)})^H\boldsymbol{z} \tag{10}$$

where $\boldsymbol{U}^{(K)}$ is the $L \times K$ submatrix consisting in the first K columns of $\boldsymbol{U}$, and $\boldsymbol{\Gamma}^{(K)}$ is the submatrix consisting of the K first rows and columns of $\boldsymbol{\Gamma}$. The truncation error can be bounded as [10]

$$\left\|\boldsymbol{x}_{\lambda} - \boldsymbol{x}_{\lambda}^{(K)}\right\|_2 \leq \boldsymbol{\gamma}[K+1]\,\|\boldsymbol{z}\|_2 \ . \tag{11}$$

3) Approximate solution in the general case $P \geq 1$: In the general case where P sub-regions are considered, we propose an approximation, given in eq. (13), of the practically-untractable closed-form solution (7). The approximation relies on the assumption that the composition of Gabor multipliers related to two different sub-regions may be neglected provided the regions are sufficiently far apart:

$$p \neq q \Rightarrow \left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\| \approx 0, \forall p, q \ . \tag{12}$$

Let us assume for the sake of the argument that for all $p \neq q$, $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\| = 0$, in other words the range of $\mathfrak{M}_{\Omega_q}$ is in the null space of $\mathfrak{M}_{\Omega_p}$; as a result, the eigenvectors of $\mathfrak{M}_{\Omega_q}$ with the largest eigenvalues are orthogonal to the eigenvectors of $\mathfrak{M}_{\Omega_p}$ with the largest eigenvalues.

Hence one may jointly diagonalize all the Gabor multipliers $\left\{\mathfrak{M}_{\Omega_p}\right\}_{p=1}^{P}$ in a common orthobasis of eigenvectors $\boldsymbol{U}^{(\boldsymbol{K})} := \left[\boldsymbol{U}_1^{(K_1)}, \dots, \boldsymbol{U}_P^{(K_P)}, \boldsymbol{U}_0^{(K_0)}\right]$ where $\boldsymbol{U}_p$ is the matrix of orthonormal eigenvectors associated with the K_p largest eigenvalues $\boldsymbol{\sigma}_p^{(K_p)}$ of $\mathfrak{M}_{\Omega_p}$ sorted in decreasing order, and where $\boldsymbol{U}_0^{(K_0)}$ is composed of $K_0 = L - \sum_{p=1}^{P} K_p$ orthonormal vectors to complete the basis. It follows that the closed form solution (7) is approximated by

$$\begin{aligned}\boldsymbol{x}_{\boldsymbol{\lambda}}^{(\boldsymbol{K})} &= \boldsymbol{z} - \left(\mathbf{I} - \boldsymbol{U}^{(\boldsymbol{K})}\,\mathrm{diag}\begin{bmatrix}\lambda_1\boldsymbol{\sigma}_1^{(K_1)} \\ \vdots \\ \lambda_P\boldsymbol{\sigma}_P^{(K_P)} \\ \mathbf{0}_{K_0}\end{bmatrix}\left(\boldsymbol{U}^{(\boldsymbol{K})}\right)^{-1}\right) \\ &\quad \times \sum_{p=1}^{P}\lambda_p\boldsymbol{U}_p^{(K_p)}\,\mathrm{diag}\,\boldsymbol{\sigma}_p^{(K_p)}\left(\boldsymbol{U}_p^{(K_p)}\right)^H\boldsymbol{z} \\ &= \boldsymbol{z} - \sum_{p=1}^{P}\boldsymbol{U}_p^{(K_p)}\,\mathrm{diag}\left(1-\lambda_p\boldsymbol{\sigma}_p^{(K_p)}\right)^{-1}\left(\boldsymbol{U}_p^{(K_p)}\right)^H \\ &\quad \times \sum_{p=1}^{P}\lambda_p\boldsymbol{U}_p^{(K_p)}\,\mathrm{diag}\,\boldsymbol{\sigma}_p^{(K_p)}\left(\boldsymbol{U}_p^{(K_p)}\right)^H\boldsymbol{z}\end{aligned}$$

Defining $\boldsymbol{\Gamma}_p^{(K_p)} = \operatorname{diag}\left(\boldsymbol{\gamma}_p[1], \ldots, \boldsymbol{\gamma}_p[K_p]\right)$ with $\boldsymbol{\gamma}_p[i] = \frac{\lambda_p \boldsymbol{\sigma}_p[i]}{1-\lambda_p \boldsymbol{\sigma}_p[i]}$, we then obtain

$$\begin{aligned}\boldsymbol{x}_{\boldsymbol{\lambda}}^{(\boldsymbol{K})} = \boldsymbol{z} &- \sum_{p=1}^{P} \boldsymbol{U}_p^{(K_p)} \boldsymbol{\Gamma}_p^{(K_p)} \left(\boldsymbol{U}_p^{(K_p)}\right)^H \boldsymbol{z} \\ &- \sum_{p \neq q} \boldsymbol{U}_p^{(K_p)} \operatorname{diag}\left(1-\lambda_p \boldsymbol{\sigma}_p^{(K_p)}\right)^{-1} \left(\boldsymbol{U}_p^{(K_p)}\right)^H \boldsymbol{U}_q^{(K_q)} \\ &\qquad \operatorname{diag}\left(\lambda_q \boldsymbol{\sigma}_q^{(K_q)}\right) \left(\boldsymbol{U}_q^{(K_q)}\right)^H \boldsymbol{z} \ .\end{aligned}$$

Since $\left(\boldsymbol{U}_p^{(K_p)}\right)^H \boldsymbol{U}_q^{(K_q)} \approx 0$ the approximate solution reads

$$\boldsymbol{x}_{\boldsymbol{\lambda}}^{(\boldsymbol{K})} = \boldsymbol{z} - \sum_{p=1}^{P} \boldsymbol{U}_p^{(K_p)} \boldsymbol{\Gamma}_p^{(K_p)} \left(\boldsymbol{U}_p^{(K_p)}\right)^H \boldsymbol{z} \ , \qquad (13)$$

which is actually a truncated version of a sum of independent single region solutions (9)

$$\boldsymbol{x}_{\boldsymbol{\lambda}}^{L} = \boldsymbol{z} - \sum_{p=1}^{P} \left(\mathbf{I} - (1-\lambda_p)\mathfrak{M}_{\Omega_p}\right)^{-1} \lambda_p \mathfrak{M}_{\Omega_p} \boldsymbol{z} \ . \qquad (14)$$

In fact the norms $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\|$ are only approximately zero. We show in the Appendix that the error induced by the approximation can actually be controlled, assuming $0 < \lambda_p < 1$ for all p.

Setting $\tilde{\lambda}_p = \min_{q \neq p} \lambda_q$ and $\Delta_p = \sum_{q \neq p} \left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\|$ and assuming that for $1 \leq p \leq P$, $\Delta_p < \frac{\lambda_p}{1-\lambda_p} \frac{\tilde{\lambda}_p}{1-\tilde{\lambda}_p}$, i.e., that the products of pairs of different Gabor multipliers are small enough compared to the values of regularization parameters, we obtain

$$\left\|\boldsymbol{x}_{\boldsymbol{\lambda}} - \boldsymbol{x}_{\boldsymbol{\lambda}}^{L}\right\|_2 \leq \sum_{p=1}^{P} \frac{(1-\tilde{\lambda}_p)\Delta_p}{\lambda_p \tilde{\lambda}_p - (1-\lambda_p)(1-\tilde{\lambda}_p)\Delta_p} \left\|\boldsymbol{z}\right\|_2 \ , \qquad (15)$$

An error bound for $\|\boldsymbol{x}_{\boldsymbol{\lambda}} - \boldsymbol{x}_{\boldsymbol{\lambda}}^{(\boldsymbol{K})}\|_2$ can be obtained from (15) and (11) using the triangle inequality.

C. All-in-one algorithms

Relying on solutions presented in Sections III-B2 and III-B3, we introduce Algorithms 1 and 2 that process a mixture $\boldsymbol{z}$ and a mask $\boldsymbol{\Omega}$ in order to estimate the target signal. Algorithm 1 (resp. 2), named `TFF-1` (resp. `TFF-P`), implements the solution of Section III-B2 (resp. III-B3), by considering Ω as a whole region (resp. as multiple subregions). In particular, it includes the creation of the Gabor multipliers (steps 2 in Alg. 1 and 4 in Alg. 2), the computation of their approximate EVD (steps 3 in Alg. 1 and 5 in Alg. 2), and the construction of the solutions (steps 6 in Alg. 1 and 9 in Alg. 2).

In order to make them computationally efficient and convenient for real-world usage, those algorithms also rely on three additional aspects :

- in `TFF-P`, TF region Ω is partitioned into subregions (step 1 in Alg. 2); this can be done automatically without specifying the number of subregions, as described in Section III-C1;
- the Gabor multipliers are diagonalized using fast randomized algorithms, as described in Section III-C2
- efficient hyperparameter tuning is obtained by estimating target energy values in the (sub-)regions (steps 1-4-5 in Alg. 1 and 2-6-7 in Alg. 2), as detailed in Section III-C3.

Algorithm 1 `TFF-1`: filtering out one TF region

Require: Mix $\boldsymbol{z}$, mask Ω, DGT $\mathcal{V}$, parameters $\epsilon_{\text{EVD}}, p_{\text{EVD}}$,

1: Estimate target energy $e_{\text{target}} \in \mathbb{R}$ in Ω (see III-C3)
2: $\mathfrak{M}_\Omega \leftarrow \mathcal{V}^H 1_\Omega \mathcal{V}$
3: $\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{\sigma}} \leftarrow \texttt{rand\_evd}\left(\mathfrak{M}_\Omega, \epsilon_{\text{EVD}}, p_{\text{EVD}}\right)$
4: Let

$$\widehat{\boldsymbol{x}} : \lambda \mapsto \boldsymbol{z} - \widehat{\boldsymbol{U}} \boldsymbol{\Gamma}(\lambda) \widehat{\boldsymbol{U}}^H \boldsymbol{z}$$

where

$$\boldsymbol{\Gamma} : \lambda \mapsto \operatorname{diag}\left(\frac{\lambda \widehat{\boldsymbol{\sigma}}}{1-(1-\lambda)\widehat{\boldsymbol{\sigma}}}\right)$$

5: $\widehat{\lambda} \leftarrow \operatorname{argmin}_{\lambda>0} \left| e_{\text{target}} - \|\mathcal{V}\widehat{\boldsymbol{x}}(\lambda)\|_\Omega^2 \right|$
6: **return** $\widehat{\boldsymbol{x}}\left(\widehat{\lambda}\right)$

Algorithm 2 `TFF-P`: filtering out P TF sub-regions

Require: Mix $\boldsymbol{z}$, mask Ω, DGT $\mathcal{V}$, parameters $\epsilon_{\text{EVD}}, p_{\text{EVD}}$

1: Split Ω into sub-regions $\{\Omega_p\}_{p=1}^P$ (e.g. via Algorithm 3)
2: Estimate target energies $\boldsymbol{e}_{\text{target}} \in \mathbb{R}^P$ in $\{\Omega_p\}_{p=1}^P$ (see III-C3)
3: **for** $p = 1$ to P **do**
4: $\quad \mathfrak{M}_{\Omega_p} \leftarrow \mathcal{V}^H 1_{\Omega_p} \mathcal{V}$
5: $\quad \widehat{\boldsymbol{U}}_p, \widehat{\boldsymbol{\sigma}}_p \leftarrow \texttt{rand\_evd}\left(\mathfrak{M}_{\Omega_p}, \epsilon_{\text{EVD}}, p_{\text{EVD}}\right)$
6: $\quad$ Let

$$\widehat{\boldsymbol{x}}_p : \lambda \mapsto \boldsymbol{z} - \widehat{\boldsymbol{U}}_p \boldsymbol{\Gamma}_p(\lambda) \widehat{\boldsymbol{U}}_p^H \boldsymbol{z}$$

where

$$\boldsymbol{\Gamma}_p : \lambda \mapsto \operatorname{diag}\left(\frac{\lambda \widehat{\boldsymbol{\sigma}}_p}{1-(1-\lambda)\widehat{\boldsymbol{\sigma}}_p}\right)$$

7: $\quad \widehat{\boldsymbol{\lambda}}[p] \leftarrow \operatorname{argmin}_{\lambda>0} \left| \boldsymbol{e}_{\text{target}}[p] - \|\mathcal{V}\widehat{\boldsymbol{x}}_p(\lambda)\|_{\Omega_p}^2 \right|$
8: **end for**
9: **return** $\boldsymbol{z} - \sum_{p=1}^P \widehat{\boldsymbol{U}}_p \boldsymbol{\Gamma}_p \left(\widehat{\boldsymbol{\lambda}}[p]\right) \widehat{\boldsymbol{U}}_p^H \boldsymbol{z}$

1) Partioning mask Ω into sub-regions: As introduced in Section III-B3, the solution using P sub-regions $\{\Omega_p\}_{p=1}^P$ relies on the hypothesis that the norm $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\|$ of the composition of Gabor multipliers from sub-regions Ω_p and Ω_q is small, which occurs when sub-regions are sufficiently away from each other in the time-frequency plane. We introduce Algorithm 3 that splits the whole region Ω in such sub-regions. The starting point is a set of initial sub-regions by clustering all the TF points from Ω, using a 4-connected structuring element to select immediate neighbors (step 1). Then, subregions for

which $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\|$ is greater than a tolerance threshold are merged repeatedly until no such pairs of regions exist (step 3).

Algorithm 3 Finding sub-regions for `TFF-P`

Require: Mask Ω, tolerance $\epsilon > 0$ on the spectral norm of the composition of two Gabol multipliers, DGT $\mathcal{V}$

1: Let $\mathcal{P} = \{\Omega_p\}$ be the partition of Ω into labeled connected sub-regions.
2: **while** $\exists \Omega_p, \Omega_q \in \mathcal{P}, \left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\| > \epsilon$ **do**
3: For every p, q such that $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\| > \epsilon$, replace Ω_p and Ω_q by $\Omega_p \cup \Omega_q$ in $\mathcal{P}$
4: **end while**
5: **return** $\mathcal{P}$

For the bird+car example mentioned in Figure 1, Algorithm 3 applied with $\epsilon = 10^{-5}$ leads to 9 sub-regions for which the actual values of $\left\|\mathfrak{M}_{\Omega_p}\mathfrak{M}_{\Omega_q}\right\|$ are of the order of 10^{-29}, which shows how suitable the sub-region splitting is in practice.

2) Diagonalization of Gabor multipliers using `Rand-EVD`: The algorithms proposed in this paper rely strongly on eigenvalue decompositions (EVD) of Gabor multipliers. However, most EVD algorithms have complexity $\mathcal{O}\left(L^3\right)$, which becomes prohibitive in high dimensions. This can be reduced when the numerical rank K is significantly lower than the dimension. This is the case here, at least when the region Ω is significantly smaller than the time-frequency domain Λ, as shown in Section II-C. Provided that K is known, such a low-rank approximation is computed in $\mathcal{O}\left(KL^2\right)$ by well-known algorithms based on the Lanczos algorithm such as the Implicitly Restarted Arnoldi Method available in standard librairies under the usual name `eigs`.

Even more efficient algorithms exist, as described in [27], based on random projections. In addition to be significantly faster, they are also able to estimate the numerical rank K while the previous methods needs to be given this rank.

In the current context, the low-rank eigenvalue decomposition of Gabor multipliers is obtained by combining the Adaptive Range Finder Algorithm [27, Algorithm 4.2] followed by the so-called Eigenvalue Decomposition via Nyström Method [27, Algorithm 5.5]. The Adaptive Range Finder Algorithm provide a low-rank matrix that approximates the range of the Gabor multiplier. The rank is estimated from a target approximation error ϵ_{EVD} and a probability p_{EVD} to reach this error. This algorithm is particularly suitable in the case of fast operators with fast-decreasing eigenvalues like the DGT-based multipliers used here. Then, the Eigenvalue Decomposition via Nyström Method uses this matrix to compute the approximate EVD, taking advantage of the Hermitian, positive semidefinite properties of our Gabor multipliers.

The combination of both algorithms will be named `Rand-EVD` in the rest of the paper. In practice, we have observed that `Rand-EVD` is able to estimate the numerical rank and to compute an approximate EVD about three times faster than `eigs` in settings such as those described in Section IV with parameters $\epsilon_{\text{EVD}} < 10^{-3}$ and $p_{\text{EVD}} = 0.9999$.

3) Efficient solutions for tuning hyperparameters: We propose a simple method to estimate hyperparameter $\boldsymbol{\lambda}$. The main idea is to consider a target energy for the coefficients within Ω or any sub-region Ω_p, and to adjust the related hyperparameter consequently to fit this target value. Indeed, as can be seen in Problem (1), the higher the value of $\boldsymbol{\lambda}[p]$, the lower the energy $\|\mathcal{V}\boldsymbol{x}\|^2_{\Omega_p}$ in sub-region Ω_p.

The target energy may be adjusted by hand or by computing the energy in an unmasked region similar to the masked one, as described in [10]. Here, we propose a more systematic and yet simple strategy, assuming that the target signal has some stationarity properties: for each frequency channel f, compute the average $\bar{e}_f$ of squared-magnitude TF coefficients in the mixture localized at time t and frequency f for any t such that $(t, f) \in \overline{\Omega}$; then, estimate the energy in Ω_p as $\boldsymbol{e}_{\text{target}}[p] = \sum_{(t,f)\in\Omega_p} \bar{e}_f$; the energy in Ω is estimated as $e_{\text{target}} = \sum_{p=1}^{P} \boldsymbol{e}_{\text{target}}[p]$.

Given target energies computed by the proposed strategy or by any other method, one can tune the hyperparameters by solving the optimization problems at step 5 in Algorithm 1 or at step 7 in Algorithm 2. This is computationally efficient for two reasons. First, while usual hyperparameter tuning via cross validation generally implies to run the whole optimization procedure for model estimation at each hyperparameter evaluation, `TFF-1` and `TFF-P` requires to diagonalize the Gabor multipliers only once. Hence, the cost of computing a solution for a given λ or λ_p is linear in the signal length. Second, estimating each λ_p is done separately thanks to the fact that sub-regions are far away from each other by construction (see Section III-C1). Hence, simple line searches over each λ_p in $\boldsymbol{\lambda}$ are conducted, avoiding a costly global P-dimensional optimization over $\boldsymbol{\lambda}$.

An illustration in the case of a bird+car mixture previously mentioned is given in Section IV-B and in Figure 5.

IV. Numerical experiments

This section is focused on practical experiments with the algorithms introduced in section III. In section IV-A, the experimental setting is introduced, including the sound material, the problem generation, the performance measure, the solvers, their parameters and implementation details. In section IV-B, we show and comment on results for the time-frequency fading problem using five different approaches (2 baseline one and 3 based on the proposed approaches):

For reproducibility, all the code and data are available in the `tff2020` toolkit [2]. Simulations have been conducted both in MATLAB and Python on CPU at 2.3GHz.

A. Experimental setting

1) Sound material: all audio signals considered here have been sampled at 8 kHz and their length is $L = 16384$ samples (i.e., about 2 seconds).

The observed mixture $\boldsymbol{z} = \boldsymbol{x}^o + \boldsymbol{y}^o$ is constructed as the sum of a target signal $\boldsymbol{x}^o$ and a perturbation signal $\boldsymbol{y}^o$. The target signals have a wideband spectrogram while the perturbation signals are well localized in time-frequency. Each source is normalized so that the target signal energy is 8dB

[2] https://gitlab.lis-lab.fr/skmad-suite/tff2020

above the perturbation energy after mixing. We consider 18 mixtures obtained from all pairs of target signal (*car*, *plane* and *train* engine sounds) and perturbation signal (*beeps*, *bird* song, *chirps*, *clicks*, *finger snaps* and *modulations*).

2) Problem generation: the DGT is parameterized using 1024 frequency bins and a hop size is $a = 64$. To favor the TF localization properties of the Gabor multiplier eigenvectors as highlighted in section II, a Gaussian window for which 96% of the energy is concentrated within a width of 256 samples. Considering only non-negative frequencies, it results in a time-frequency matrix of size 513×256. With these parameters, the redundancy of the Gabor frame equals 8. Some additional experiments have also been conducted using a Hann window of size 512 samples. Results are similar to those with the Gaussian window and are not reported here in details.

The TF fading problem is based upon the knowledge of the region Ω. In our experiments, Ω was generated as follows:

- extract $\Omega_{\boldsymbol{x}^o}$ as the TF points where the TF magnitude of target is with a dynamic range of 40dB below its maximum magnitude;
- extract $\Omega_{\boldsymbol{x}^o > \boldsymbol{y}^o}$ as the TF points where the TF magnitude of target is higher than that of the perturbation;
- take the union $\Omega_{\text{raw}} = \Omega_{\boldsymbol{x}^o} \cup \Omega_{\boldsymbol{x}^o > \boldsymbol{y}^o}$;
- smooth Ω_{raw} via standard mathematical morphology techniques [28] by applying a closing followed by an opening with a structuring element of radius 3 and + shape.

3) Performance measures: to assess numerically the quality of the reconstruction, we use two performance measures. The Source to Distortion Ratio (SDR) [29]

$$\text{SDR}\left(\boldsymbol{x}^o, \widehat{\boldsymbol{x}}\right) = 20\log_{10}\left(\frac{\|\boldsymbol{x}^o\|}{\|\widehat{\boldsymbol{x}} - \boldsymbol{x}^o\|}\right)$$

quantifies the ratio between the energy of the target signal denoted by $\boldsymbol{x}^o$ and residual error with an estimate $\widehat{\boldsymbol{x}}$. In order to complement this signal-domain evaluation by a TF-domain quality measure, we compute the Itakura-Saito divergence (IS) using the discrete Fourier transform $\mathcal{F}$ of the signals, as

$$\text{IS}\left(\boldsymbol{x}^o, \widehat{\boldsymbol{x}}\right) = \sum_f \left(\left|\frac{\mathcal{F}\boldsymbol{x}^o[f]}{\mathcal{F}\widehat{\boldsymbol{x}}[f]}\right| - \log\left|\frac{\mathcal{F}\boldsymbol{x}^o[f]}{\mathcal{F}\widehat{\boldsymbol{x}}[f]}\right| - 1\right)$$

4) Solvers: for comparison, we compute 2 baseline solutions and 3 estimates obtained from the proposed approach:

- `Zero-fill`: a baseline method consisting in applying $\mathfrak{M}_{\overline{\Omega}}$ to the observed mixture $\boldsymbol{z}$, i.e., in reconstructing the signal after filling the masked regions by zeros. This is equivalent to take $\lambda_p = \lambda = 1$ in the analytical solution.
- `Interp`: a baseline method consisting in applying a linear interpolation along the frequency axis of the magnitude of observation TF matrix, and by drawing the related phase uniformly at random, as proposed in some industrial software [30].
- `TFF-1`: Time-Frequency Fading method considering the mask as one region (Algo. 1).
- `TFF-P`: Time-Frequency Fading method considering the mask as P sub-regions (Algo. 2).
- `TFF-O`: Time-Frequency Fading method considering the mask as P sub-regions and where $\boldsymbol{\lambda}$ is chosen so as to maximize the SDR in an oracle way, in order to have an upper bound on the SDR performance.

The parameters used of `Rand-EVD` are $\epsilon_{\text{EVD}} < 10^{-3}$ and $p_{\text{EVD}} = 0.9999$ and a tolerance $\epsilon = 10^{-5}$ is set for finding the subregions in Algorithm 3.

B. Results

An example of reconstruction is given in Figure 4. One can observe that the spectrograms obtained by the proposed methods are visually very similar to that of the original target signal, except for the largest sub-region on the left. On the contrary, the spectrograms obtained by the baseline methods exhibit large differences.

Detailed results are given in Table I for the SDR and in Table II for the IS divergence. Bold face is used to emphasize the best values among `TFF-1`, `TFF-P`, `Zero-fill` and `Interp`. For all mixtures, `TFF-1` and `TFF-P` clearly give SDR values higher than that of baselines `Zero-fill` and `Interp` by several dBs. `TFF-1` and `TFF-P` SDR values are close to the oracle upper bound given by `TFF-O`, by about 1dB and there is no clear difference between `TFF-1` and `TFF-P` in terms of SDR.

In terms of IS divergence, values are much more heterogeneous. `TFF-1` and `TFF-P` still outperforms the baseline methods in general. However, the `Zero-fill` strategy is the best one when the perturbation is impulsive (*clicks* and *finger snaps*). In those cases, the perturbation affects a wide frequency range: the `Interp` strategy is particularly unadapted due to the interpolation along the frequency axis and `Zero-fill` is giving the lowest (but yet very high) IS values.

Signals \ Methods		Zero-fill	Interp	TFF-O	TFF-1	TFF-P
beeps	car	20.49	20.17	24.58	**23.27**	23.26
	train	20.54	20.12	24.51	**23.69**	23.12
	plane	26.20	25.95	28.23	**27.73**	27.70
bird	car	12.32	12.08	18.32	**17.19**	15.30
	train	22.25	21.31	26.02	**25.27**	23.11
	plane	25.12	24.97	27.40	**26.01**	25.66
chirps	car	13.64	13.64	21.74	**20.03**	19.90
	train	13.82	13.58	21.96	18.16	**18.28**
	plane	18.33	17.90	24.83	22.05	**22.32**
clicks	car	9.57	9.34	11.82	10.54	**11.01**
	train	9.21	8.25	11.63	10.49	**10.67**
	plane	11.45	8.67	13.94	12.69	**12.72**
finger_snaps	car	8.50	8.29	13.36	**11.66**	11.48
	train	9.22	8.47	14.64	**12.48**	12.32
	plane	11.54	10.42	16.41	**14.66**	14.41
modulations	car	12.93	12.63	17.78	**17.40**	17.13
	train	17.14	16.89	20.59	19.34	**19.56**
	plane	21.72	21.53	23.49	22.40	**22.86**

TABLE I
SDR IN DB FOR EACH METHOD AND EACH MIXTURE.

Figure 5 gives some insights in the case of the bird+car mixture and considering only one whole region Ω. Colored points represent the performance obtained by `TFF-1`, `TFF-P`, `TFF-O` and `Zero-fill` ($\lambda = 1$). The additional points

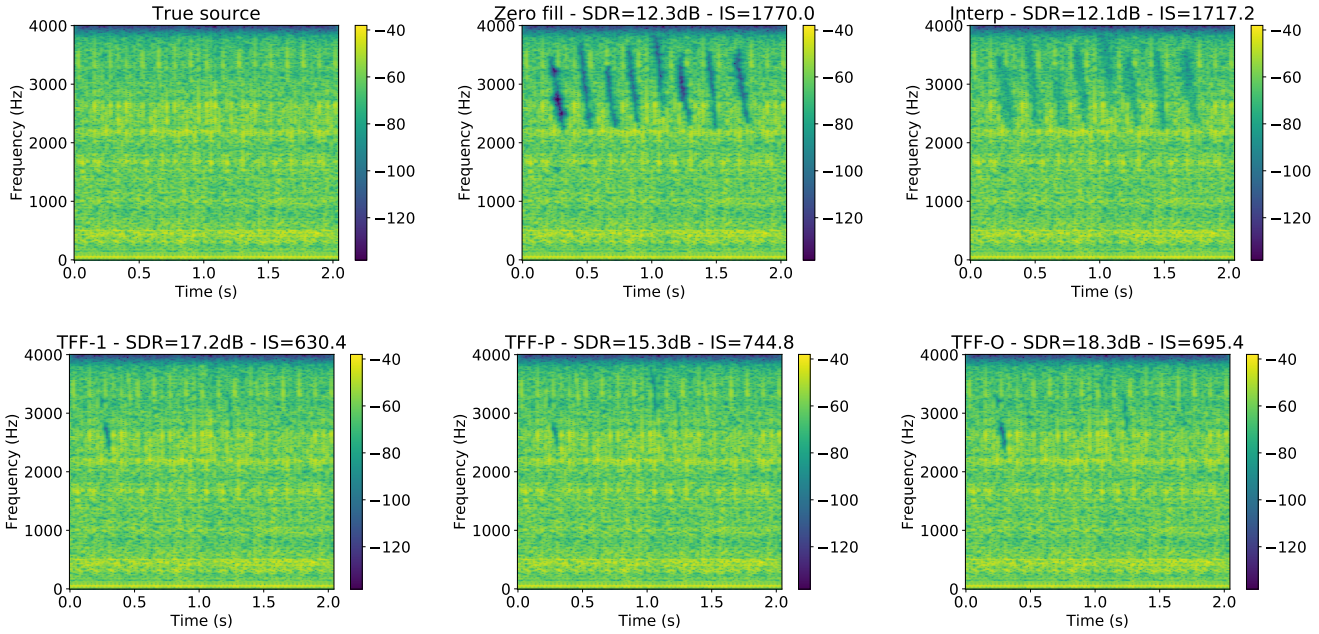

Fig. 4. Bird+car mixture: spectrograms of the target source and of the reconstruction by each method.

Signals \ Methods		Zero-fill	Interp	TFF-O	TFF-1	TFF-P
beeps	car	139.4	130.9	63.0	64.9	**63.9**
	train	270.4	214.1	255.4	**138.1**	**138.1**
	plane	452.7	235.4	142.2	**151.3**	168.3
bird	car	1770.0	1717.2	695.4	**630.4**	744.8
	train	**1870.5**	3348.3	1538.3	1943.6	2413.3
	plane	1993.6	1724.1	1383.0	**1428.7**	1439.5
chirps	car	880.0	868.4	260.3	**537.7**	613.5
	train	1205.8	1135.2	411.0	1506.8	**615.6**
	plane	989.1	982.4	450.7	1178.6	**736.1**
clicks	car	**1968.1**	2877.0	1685.8	2465.4	2152.7
	train	**2392.1**	6798.5	2111.5	5092.6	4911.6
	plane	**2273.2**	8551.7	2507.5	4262.5	4417.5
finger_snaps	car	2961.6	3286.4	1342.5	1902.8	**1769.9**
	train	**3105.5**	7668.1	2810.0	3976.7	4263.3
	plane	**2629.0**	6483.2	2396.7	3665.1	4228.4
modulations	car	832.1	877.4	235.3	282.8	**208.2**
	train	917.6	1017.7	282.2	**349.6**	412.7
	plane	1017.8	803.9	580.8	**353.5**	410.3

TABLE II
IS DIVERGENCES FOR EACH METHOD AND EACH CONFIGURATION.

named TFF−E are obtained by extracting the true energy in the subregions from the original target signal, and by using this oracle value as a target energy in the estimation of λ. Both curves show similar trends; however, the SDR curve is smooth while the IS curve behaves more erratically. One can see that the λ value estimated by TFF−1 is very close to the TFF−E value, which supports that the proposed energy estimation method is able to accurately estimate the target energy. Both TFF−1 and TFF−E points are close to the optimal TFF−O, in terms of λ value, of high SDR and of low IS divergence, while the Zero−fill points are far away from the optimal.

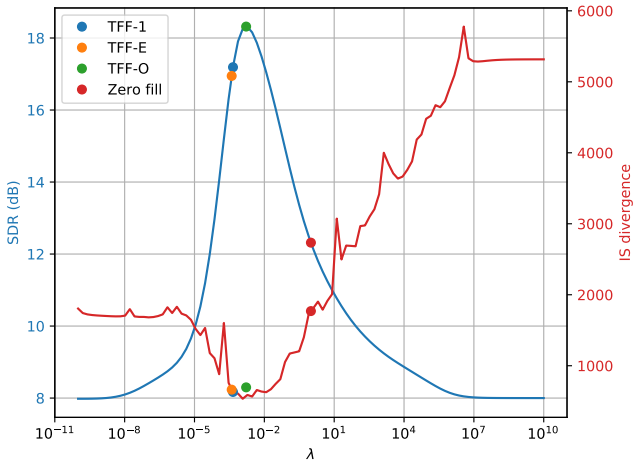

Fig. 5. Variation of the performance (SDR and IS) as a function of hyperparameter λ for the bird+car mix, considering a whole region Ω.

Due to the processing and the Gabor mutipliers diagonalization in high dimension ($L = 16384$ here), the computational complexity of the proposed methods is a key issue that has motivated several modeling and algorithmic contributions in this paper. In [10], many of those contributions were not present: processing a sound of length 8192 was lasting about 300 seconds and processing longer sounds was both untractable in time and memory. Using TFF−P, processing a twice-longer sound with similar contents takes a bit less than 200 seconds. Running times for TFF−1 and TFF−P are represented in Figure 6, where results are given both for the Hann and Gaussian windows. It shows TFF−P is about three times faster than TFF−1, while their estimation

performance are similar. Indeed, dividing the mask in P subregions results in replacing the diagonalization of a large-rank operator by P diagonalizations of lower-rank operators, with significant computational savings. One may also observe that the running times are very different from one mixture to the other: it is highly dependent on the numerical rank of the Gabor multipliers, which is related to the size of the region Ω. Eventually, the running time also depends on the DGT setting: while both windows have similar time spreading, the Hann window has a smaller support than the Gaussian window, which makes it about 30% faster here.

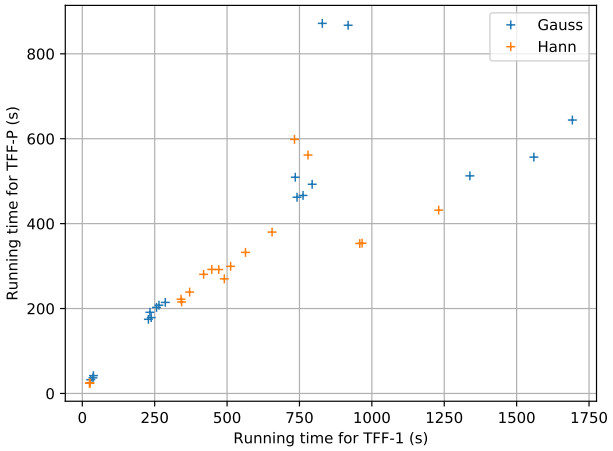

Fig. 6. Computation times of `TFF-1` versus `TFF-P`, with one point per mixture, for the Hann and Gaussian windows.

V. Conclusion

In this paper, we have addressed the problem of single- and multiple-region time-frequency fading, formulated as a variational optimization problem. Almost no assumptions are made about the signals to be processed: the only requirement is that the perturbation to be filtered out is well localized in known regions of the time-frequency plane.

We have proposed efficient algorithms based on the diagonalization of Gabor multipliers possibly coupled with fast randomized eigenvalue decomposition. The proposed methods acts in the signal domain, avoiding any time-frequency consistency issue. They are easily usable in practice, requiring very few parameters that can be set to some default values. Experiments on real sounds have shown that good reconstruction quality is obtained.

The novel time-frequency fading problem formulation and algorithmic framework may be extended in several directions. Signals with duration larger than one or two seconds may be processed easily by segmenting them, processing each segment and then reconstructing the full-length signal by overlap-add. A interesting perspective is to integrate a priori information about the sources in the optimization problem. This can lead to more specific use cases in applications like denoising, source separation, time-frequency segmentation or inpainting, as well as to developing methods to automatically estimate the mask location.

Acknowledgments

This study emerged from a real application from Ansys [30] where there was a need to remove time frequency areas in signal's spectrograms while preserving audio quality.

References

[1] J. O. Smith, *Introduction to Digital Filters: With Audio Applications*, Music signal processing series. W3K, 2007.

[2] G. F. Margrave, P. C. Gibson, J. P. Grossman, D. C. Henley, V. Iliescu, and M. P. Lamoureux, "The Gabor transform, pseudodifferential operators, and seismic deconvolution," *Integrated Comput. Aided Eng.*, vol. 12, pp. 43–55, 2005.

[3] P. Lander and E. J. Berbari, "Enhanced ensemble averaging using the time-frequency plane," in *IEEE-SP Proc. Int. Symp. on Time-Freq. and Time-Scale Analysis*, 1994, pp. 241–243.

[4] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distribution synthesis techniques," *IEEE Trans. Acous., Speech Signal Process.*, vol. 34, no. 3, pp. 442–451, 1986.

[5] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, "Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1417–1432, 2000.

[6] L. Stankovic, "On the time-frequency analysis based filtering," *Ann. Télécommun.*, vol. 55, no. 5, pp. 216–225, May 2000.

[7] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, Nov. 1946.

[8] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acous., Speech Signal Process.*, vol. 28, no. 1, pp. 55–69, 1980.

[9] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 217–220, Mar. 2013.

[10] A. M. Krémé, V. Emiya, C. Chaux, and B. Torrésani, "Filtering out time-frequency areas using Gabor multipliers," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5790–5794.

[11] K. Gröchening, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston (MA), 2011.

[12] R. Carmona, W. L. Hwang, and B. Torrésani, *Practical Time-Frequency Analysis*, vol. 9 of *Wavelet Analysis and Its Applications*, Elsevier, 1998.

[13] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Academic Press, dec 2015.

[14] P. Flandrin, *Explorations in Time-Frequency Analysis*, Cambridge University Press, 2018.

[15] H. G. Feichtinger and K. Nowak, "A first survey of Gabor multipliers," in *Advances in Gabor Analysis*, H. G. Feichtinger and T. Strohmer, Eds., pp. 99–128. Birkhaüser, 2002.

[16] W. Kozek and G. Pfander, "Identification of operators with bandlimited symbols," *SIAM J. Math. Anal.*, vol. 37, pp. 867–888, 08 2006.

[17] H. Bölcskei, "Orthogonal frequency division multiplexing based on offset QAM," in *Advances in Gabor Analysis*, pp. 321–352. 2003.

[18] H. G. Feichtinger, M. Hampejs, and G. Kracher, "Approximation of matrices by Gabor multipliers," *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 883–886, 2004.

[19] J. J. Benedetto and G. E. Pfander, "Frame expansions for gabor multipliers," *Appl. Comput. Harmon. Analysis*, vol. 20, no. 1, pp. 26 – 40, 2006, Computational Harmonic Analysis - Part 2.

[20] M. Dörfler and B. Torrésani, "On the time-frequency representation of operators and generalized Gabor multiplier approximations," *J. Fourier Anal. Appl.*, vol. 16, pp. 261–293, 2010.

[21] J. Ramanathan and P. Topiwala, "Time-frequency localization via the Weyl correspondence," *SIAM J. Math. Anal.*, vol. 24, no. 5, pp. 1378–1393, 1993.

[22] L. D. Abreu, K. Gröchenig, and J. L. Romero, "On accumulated spectrograms," in *Trans. Amer. Math. Soc.*, 2016, vol. 368, pp. 3629 – 3649.

[23] P. Soendergaard, B. Torrésani, and P. Balazs, "The linear time frequency analysis toolbox," *International Journal of Wavelets and Multiresolution Information Processing*, vol. 10, no. 4, pp. 1250032–1 – 1250032–27, 2012.

[24] Z. Průša, P. L. Søndergaard, N. Holighaus, Ch. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox 2.0," in *Sound, Music, and Motion*, LNCS, pp. 419–442. Springer International Publishing, 2014.
[25] D. Arrivault and F. Jaillet, "LTFATPY, a partial python port of the Large Time-Frequency Analysis Toolbox (LTFAT)," available at https://pypi.org/project/ltfatpy/, 2018, LabEx Archimede, LIS, I2M, Aix-Marseille Université.
[26] D. T. Stoeva and P. Balazs, "A survey on the unconditional convergence and the invertibility of multipliers with implementation," in *Sampling – Theory and Applications (A Centennial Celebration of Claude Shannon)*, M. Robinson S. Casey, K. Okoudjou and B. Sadler, Eds. Springer, 2020, to appear, preprint arXiv:1803.00415.
[27] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM J. Math. Anal.*, vol. 53, no. 2, pp. 217–288, 2011.
[28] R. Van den Boomgard and R. van Balen, "Methods for fast morphological image transforms using bitmapped images," *Comput. Vision Graph. Image Process.*, vol. 54, no. 3, pp. 252–254, May 1992.
[29] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, July 2006.
[30] ANSYS VRXPERIENCE, "Sound website," https://www.ansys.com/fr-fr/products/systems/ansys-vrxperience/sound.

APPENDIX

PROOF OF THE ERROR BOUND (15)

Throughout this section we assume that $\lambda_p \in]0,1[$ fo all $p = 1, \dots P$. For the sake of simplicity we set

$$A_p = (1-\lambda_p)\mathfrak{M}_{\Omega_p} \ , \qquad \mu_p = \lambda_p/(1-\lambda_p) \ .$$

The assumption on λ_p then ensures $\|A_p\| < 1$. With these notations, we wish to approximate

$$\boldsymbol{z} - \left(\mathbf{I} - \sum_{q=1}^{P} A_q\right)^{-1} \sum_{p=1}^{P} \mu_p A_p \boldsymbol{z} \approx \boldsymbol{z} - \sum_{p=1}^{P} \mu_p (\mathbf{I} - A_p)^{-1} A_p \boldsymbol{z}$$

Defining $B_p = \sum_{q \neq p} A_q$, we then set

$$\begin{aligned} R_p &= \left(\mathbf{I} - \sum_{q=1}^{P} A_q\right)^{-1} A_p - (\mathbf{I} - A_p)^{-1} A_p \\ &= \left[\left(\mathbf{I} - (\mathbf{I} - A_p)^{-1} B_p\right)^{-1} - \mathbf{I}\right] (\mathbf{I} - A_p)^{-1} A_p \\ &= \left(\mathbf{I} - (\mathbf{I} - A_p)^{-1} B_p\right)^{-1} (\mathbf{I} - A_p)^{-1} B_p A_p (\mathbf{I} - A_p)^{-1} \end{aligned}$$

since $A_p B_p = B_p A_p$ by construction. We show below that $\|A_p\| < 1$, $\|B_p\| < 1$, and that under suitable assumptions (specified below as well), $\|(\mathbf{I} - A_p)^{-1} B_p\| < 1$. Then we have

$$\|R_p\| \leq \frac{1}{(1 - \|A_p\|)^2} \frac{\|A_p B_p\|}{1 - \left\|(\mathbf{I} - A_p)^{-1} B_p\right\|} \ .$$

Then notice that

$$\|(\mathbf{I} - A_p)^{-1} B_p\| = \left\|B_p + (\mathbf{I} - A_p)^{-1} A_p B_p\right\| \leq \|B_p\| + \frac{\|A_p B_p\|}{1 - \|A_p\|}$$

so that

$$1 - \|(\mathbf{I} - A_p)^{-1} B_p\| \geq \frac{(1 - \|A_p\|)(1 - \|B_p\|) - \|A_p B_p\|)}{1 - \|A_p\|} \ .$$

Coming back to initial notations, we have $\|A_p\| \leq 1 - \lambda_p$ so that

$$1 - \|A_p\| \geq \lambda_p > 0 \ .$$

Also, since the sub-regions Ω_q are pairwise disjoint

$$\|B_p\| = \left\|\sum_{q \neq p} A_q\right\| = \left\|\sum_{q \neq p} (1 - \lambda_q) \mathfrak{M}_{\Omega_q}\right\| \leq 1 - \min_{q \neq p} \lambda_q \ ,$$

therefore

$$1 - \|B_p\| \geq \tilde{\lambda}_p \ , \qquad \text{where} \qquad \tilde{\lambda}_p = \min_{q \neq p} \lambda_q > 0 \ .$$

Besides,

$$\|A_p B_p\| = (1 - \lambda_p) \left\|\sum_{q \neq p} (1 - \lambda_q) \mathfrak{M}_p \mathfrak{M}_q\right\| \leq (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p \ ,$$

where we have set

$$\Delta_p = \sum_{q \neq p} \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| \ .$$

Plugging this expression into the lower bound of $1 - \|(\mathbf{I} - A_p)^{-1} B_p\|$ yields

$$1 - \|(\mathbf{I} - A_p)^{-1} B_p\| \geq \frac{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p}{1 - \|A_p\|} \ .$$

From now on we assume that

$$\Delta_p < \frac{\lambda_p \tilde{\lambda}_p}{(1 - \lambda_p)(1 - \tilde{\lambda}_p)}$$

so that the lower bound is strictly positive. Putting things together we then obtain

$$\|R_p\| \leq \frac{1 - \lambda_p}{\lambda_p} \frac{(1 - \tilde{\lambda}_p) \Delta_p}{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p} \ .$$

The bound (15) then follows from

$$\left\|\boldsymbol{x}_{\boldsymbol{\lambda}} - \boldsymbol{x}_{\boldsymbol{\lambda}}^{L}\right\|_2 \leq \sum_p \frac{\lambda_p}{1 - \lambda_p} \|R_p\| \, \|\boldsymbol{z}\|_2 \ .$$