Convex regularizations for the simultaneous recording of room impulse responses

Alexis Benichoux, Laurent S. R. Simon, Emmanuel Vincent, Rémi Gribonval

RESEARCH REPORT

N° 8130

November 13, 2012

Project-Teams METISS

ISSN 0249-6399 ISRN INRIA/RR--8130--FR+ENG

Convex regularizations for the simultaneous recording of room impulse responses

Alexis Benichoux, Laurent S. R. Simon, Emmanuel Vincent, Rémi Gribonval

Project-Teams METISS

Research Report n° 8130 — November 13, 2012 — 19 pages

Abstract: We propose to acquire large sets of room impulse responses (RIRs) by simultaneously playing known source signals on multiple loudspeakers. We then estimate the RIRs via a convex optimization algorithm using convex penalties promoting sparsity and/or exponential amplitude envelope. We validate this approach on real-world recordings. The proposed algorithm makes it possible to estimate the RIRs to a reasonable accuracy even when the number of recorded samples is smaller than the number of RIR samples to be estimated, thereby leading to a speedup of the recording process compared to state-of-the-art RIR acquisition techniques. Moreover, the penalty promoting both sparsity and exponential amplitude envelope provides the best results in terms of robustness to the choice of its parameters, thereby consolidating the evidence in favor of sparse regularization for RIR estimation. Finally, the impact of the choice of the emitted signals is analyzed and evaluated.

Key-words: Room impulse response recording, convex optimization, compressed sensing

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Régularisations convexes pour l'enregistrement simultané de réponses acoustiques de salles

Résumé : Nous proposons d'acquérir un grand nombre de réponses de salles (RIRs) en émettant simultanément des signaux connus depuis plusieurs haut-parleurs. Nous estimons ensuite les RIRs via un algorithme d'optimisation convexe muni de pénalités convexes qui favorisent la parcimonie et/ou l'enveloppe exponentielle décroissante. Nous validons cette approche sur des enregistrements réels. L'algorithme proposé permet d'estimer les RIR avec une précision raisonnable, même quand le nombre d'échantillons enregistrés est plus petit que le nombre de d'échantillons des RIRs à estimer, aboutissant à une accélération du processus d'enregistrement par rapport aux méthodes d'acquisition de l'état de l'art. De plus, la pénalité qui force la parcimonie et l'enveloppe exponentielle décroissante donne les meilleurs résultats en terme de robustesse au choix des paramètre, ce qui justifie d'autant plus la régularisation parcimonieuse pour l'estimation des RIRs. Finalement, l'impact du choix des signaux sources est analysé et évalué.

Mots-clés : Enregistrement des réponses acoustiques de salles, optimisation convexe, compressed sensing

1 Introduction

The estimation of room impulse responses (RIRs) is a central problem in audio signal processing. The calibration of modern audio rendering systems such as wavefield synthesis (WFS) [1] requires the knowledge of the RIRs between the loudspeakers and several possible listener positions in order to compensate for the so-called room effect [2]. For example, the study in [3] considered a WFS system of 48 loudspeakers and 6 multi-actuator panels calibrated in 96 different microphone positions. Similarly, a typical measure of binaural room impulse responses (BRIRs) involves the acquisition of RIRs in up to several hundred possible source and listener spatial configurations [4]. A larger number of loudspeakers or microphone positions would be welcome in many settings, but it is limited so far due in particular to the large total recording time implied by state-of-the-art RIR acquisition techniques, which is inconvenient in real-world scenarios.

While these techniques typically consist of activating each loudspeaker in turn, we propose in this paper a way to record RIRs from multiple simultaneously active loudspeakers. We introduce a convex optimization algorithm for RIR estimation which exploits convex penalties on the RIRs in the spirit of compressed sensing [5]. We consider the classical ℓ^1 sparsity-promoting penalty [6, 7, 8] as well as new penalties accounting for the fact that RIRs have an exponentially decaying envelope. This algorithm makes it possible to estimate the RIRs to a reasonable accuracy from an amount of recorded data that would otherwise be insufficient to estimate them at all, thereby leading to a speedup of the recording process.

In our preliminary study [9], we validated this approach on a set of synthetic RIRs using Gaussian emitted signals and assuming exact knowledge of the room reverberation time. In this paper, we perform experiments on real-world recordings instead and we analyze both the choice of the emitted signals and the robustness of the algorithm to the values of its parameters. In addition, we introduce a new evaluation methodology based on measuring the distance between the estimated RIRs and ground truth RIRs, that has not been used so far to our knowledge.

The paper is organized as follows. In Section 2, we formalize the considered problem and review existing techniques for individual and simultaneous measurement of RIRs. In Section 3, we study the characteristics of RIRs in order to design appropriate penalties. We describe the linear system corresponding to the simultaneous recording of RIRs, and the convex optimization algorithm used for its inversion. Section 4 describes the real-world acoustic setup used for the experiments and Section 5 analyzes the results. We conclude in Section 6.

2 Existing methods for individual and simultaneous measurement of RIRs

2.1 Formalization of the problem

The considered problem is formalized as follows. A set of N loudspeakers simultaneously emit N known discrete-time source signals $S_n(t)$ of duration T. Recording is performed using M microphones, leading to M discrete-time observed signals $X_m(t)$ of length T. The playback and recording processes are assumed to start at the same time. Assuming quasi-linear behavior of both the loudspeakers and the microphones, the recorded signals are classically modeled as

$$X_m(t) = \sum_{n=1}^{N} (A_{mn} \star_{[0,T-1]} S_n)(t) + E_m(t) \tag{1}$$

where $A_{mn}(k)$ is the filter (RIR) of length K between source n and microphone m, $E_m(t)$ represents the background noise and the nonlinear contribution of the system, and $\star_{[0,T-1]}$ denotes convolution truncated to the discrete time interval $[\![0, T-1]\!]$ as defined in the Appendix.

In the following, we shall always assume that the emitted signals are centered and normalized according to their maximum amplitude, i.e., $\|\mathbf{S}\|_\infty = 1$. As shorthands we will denote by $\mathbf{X} \in \mathbb{R}^{M\times T}$ the matrix of recorded signals, $\mathbf{S} \in \mathbb{R}^{N\times T}$ the matrix of emitted sources, $\mathbf{A} \in \mathbb{R}^{M\times N\times K}$ the array of RIRs and $\mathbf{E} \in \mathbb{R}^{M\times T}$ the matrix of noise samples. Using a matrix convolution notation, the recording process becomes

$$\mathbf{X} = \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S} + \mathbf{E}. \tag{2}$$

The objective is to estimate the RIRs $\mathbf{A}$. It can be decomposed into two complementary problems:

- estimate $\mathbf{A}$ given the set of emitted signals $\mathbf{S}$ and a set of recorded signals $\mathbf{X}$,
- design the set of emitted signals $\mathbf{S}$ so as to maximize the estimation accuracy.

The estimation problem is a linear inverse problem consisting in finding $\mathbf{A}$ that satisfies approximately the equality $\mathbf{X} = \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}$. Assuming that the RIRs have a finite length K, the system is composed of MT equations for MNK unknown variables. Therefore it can be linearly inverted only if the recording duration in samples exceeds the critical recording duration

$$T \geq T^{\text{crit}} := NK. \tag{3}$$

This is the *overdetermined* regime exploited by state-of-the-art RIR recording techniques as detailed below.

By contrast, the main contribution of this paper is to explore the regime where shorter recordings are targeted, i.e., $T < T^{\text{crit}}$. In this case the system is necessarily singular. Recovering the RIRs from the recordings becomes an *underdetermined* linear inverse problem, which requires non linear estimation techniques based on prior knowledge on the RIRs. The resulting measurement scheme can then be seen as a multichannel compressed sensing approach.

2.2 State of the art

2.2.1 Dirac impulses

The most straightforward way to measure RIRs is to emit Dirac pulses. Ideally, when measuring the RIR for a single source, the emitted signal has duration $D = 1$ and is followed by silence for a recording duration $T = K$. For N sources, N Diracs are emitted every K samples, so that the total recording duration is $T = NK = T^{\text{crit}}$.

In practice, electrical sparks, popping balloons, pistol and cannon shots have been used in the past to approximate Dirac pulses [10]. However with these techniques the shape and spectrum of the emitted signal is not well controlled, leading to imprecise RIR measurements. With modern digital equipment more controlled and reproducible Dirac pulses can be achieved, but these still yield RIR estimates with limited quality because of a poor signal-to-noise ratio (SNR).

The SNR of the recordings can be directly related to the root mean square (RMS) amplitude of the emitted signals

$$\text{RMS}(\mathbf{S}) = 10 \log_{10} \frac{\|\mathbf{S}\|_2^2}{T} \tag{4}$$

expressed in decibels (dB). Dirac pulses have low RMS, due to the fact that most of the emitted signal consists of the silence following the impulses. Although recent studies have tried to adapt RIR estimation to particular types of impulses [11], the acoustic community often prefers signals with higher RMS as we shall see in the following.

A common technique to increase the SNR is to repeat the measurement r times and to average the results. The resulting recording time is $T = rNK$.

2.2.2 Maximum length sequences (MLS)

The MLS method introduced by Schroeder in 1979 [12] was initially designed to recover the RIR during an opera performance using an inaudible signal. A tutorial on both theoretical and practical aspects can be found in [13]. Besides having the greatest possible RMS, MLS signals exhibit two key properties: their autocorrelation function is close to a Dirac function, and their inverse in the sense of circular convolution is known in closed form.

MLS sequences $s \in \mathbb{R}^D$ are defined for lengths $D = 2^d - 1$ where $d \in \mathbb{N}$. The approximate decorrelation property of their circularly shifted versions allowed authors to conceive a simultaneous measurement technique provided that $D \geq NK$ [14]. The trick consists in sending simultaneously N versions of the MLS: on the n-th loudspeaker, one sends the MLS sequence time-shifted by nK.

The emitted sequences may be periodically repeated every D samples. Overall $r + 2$ repeated periods make it possible to obtain r noisy instances of the circularly convolved output that can be averaged to reduce the noise level. The first and last period can be truncated to emit only K coefficients of the shifted sequences. The recording time achieved with this method is $T = rD + 2K \geq rNK + 2K$.

One problem is the constrained duration of the signal: in order to take advantage of the closed-form expression of the inverse, D cannot be reduced to NK unless $NK = 2^d - 1$ for some $d \in \mathbb{N}$. In addition, the nonlinearities of the speakers introduce artifact peaks in the measured RIR [15].

2.2.3 Exponential and linear sine sweeps

The latter limitation led to the introduction of sine sweep techniques by Farina in [16]. Their main advantage is that nonlinearities can easily be masked out from the recordings in the time-frequency domain. A sine sweep signal $s \in \mathbb{R}^D$ is defined by $s(t) = \sin\theta(t)$ where $\theta(t)$ is either exponential (exponential sweep) or quadratic (linear sweep). The typical sine sweep duration for RIR measurement is 1.5 s [17]. If the noise is stationary, doubling the sine sweep duration D yields similar results as averaging $r = 2$ sine sweeps. The inverse sweep has a closed form expression [18] but provides numerically less accurate RIR estimation than straightforward Fourier-domain inversion.

When measuring a single RIR with a sweep of duration D, the recording duration is typically $T = K + D$. A naive way to record N RIRs is to successively emit N sweeps of duration D, with a silence of duration K between each, yielding a total duration $T = rN(D + K)$ in the case of r repetitions.

A more clever way is to overlap the sine sweeps [19] such that their delayed versions are all disjoint in the time-frequency domain. Assuming quasi-linear behavior of both the loudspeakers and the microphones, a shift by K samples is sufficient between two successive sweeps[1]. When repeated r times with overlapping sweeps of duration D, this leads to a recording duration $T = rNK + D$.

A disadvantage compared to MLS is that because high frequencies are present only at the end of the sweep, the emission must be padded with samples of silence in order to estimate the RIR at these frequencies. Together with the replacement of values in $\{-1, 1\}$ by a sine function this results in a decrease of the RMS.

2.2.4 Role of the averaging

A measurement process typically involves an averaging step, in order to reduce the background noise. Usually the mean is taken among over $r = 20$ instances [17], and sometimes up to $r = 200$ [16] in the literature. A comparison between the durations of all methods is displayed in Table 1. Simultaneous MLS techniques and overlapped sine sweeps result in a shorter recording duration than successive sine sweeps for large values of r.

[1] The effect of nonlinearities on the choice of the shift is analyzed in [19].

In the rest of the paper, we present a faster technique and evaluate it for $r = 1$ in order to bring the recording time down to a minimum. Nevertheless, it remains possible to apply this technique to the average of $r > 1$ recordings.

	T	RMS (dB)
Dirac	rNK	$-10\log_{10}(K)$
Simultaneous MLS	$rNK + 2K$	0
Successive Sweeps	$rN(K + D)$	$\simeq -10\log_{10}(2)$
Overlapping sweeps	$rNK + D$	$\simeq -10\log_{10}(2)$
Proposed	$< rNK$	0

Table 1: Comparison of the total recording duration required by different RIR acquisition techniques.

3 Convex optimization framework

Earlier work on source separation [20] used convex optimization tools to estimate $\mathbf{S}$ given $\mathbf{X}$ when $\mathbf{A}$ is known, using a sparsity prior on the sources in the time-frequency domain. Here, we adapt the method in [20] to estimate $\mathbf{A}$ when $\mathbf{S}$ is known, by computing the minimizer of the following optimization problem

$$\mathbf{A}_0 = \min_{\mathbf{A}} \mathcal{P}(\mathbf{A}) \quad \text{s.t.} \quad \mathbf{X} = \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S} \tag{5}$$

where $\mathcal{P}(\mathbf{A})$ is a convex penalty function. To take into account the presence of background noise and small nonlinearities, it can be more relevant to solve a problem of the type

$$\min_{\mathbf{A}} \mathcal{P}(\mathbf{A}) \quad \text{s.t.} \quad \|\mathbf{X} - \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}\|_2^2 \leq \varepsilon \tag{6}$$

for some $\varepsilon > 0$, which is known to be equivalent to the unconstrained minimization problem

$$\mathbf{A}_\lambda = \text{argmin}_{\mathbf{A}} \left\{ \frac{1}{2} \|\mathbf{X} - \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}\|_2^2 + \lambda \mathcal{P}(\mathbf{A}) \right\} \tag{7}$$

for some Lagrangian parameter $\lambda > 0$ [21, p. 664]. When the penalty $\mathcal{P}$ is convex, the limit when λ tends toward zero provides the minimum of $\mathcal{P}$ subject to the equality constraint: $\lim_{\lambda\to 0} \mathbf{A}_\lambda = \mathbf{A}_0$.

3.1 Choice of the penalties

The choice of the penalty requires assumptions on the RIRs. Previous studies on dereverberation, source separation or RIR interpolation in a convex optimization framework have assumed that RIRs are formed by echoes at distinct instants, so that they are sparse [6, 7, 8]. This assumption is promoted by the non-weighted ℓ^1 norm [22] which is often related to maximum a posteriori estimation with a Laplacian prior (see, e.g., [23]), although this relation is disputable [24].

The statistical theory of room acoustics [25] assumes instead that the samples of an RIR follow a Gaussian distribution, with an exponentially decaying amplitude envelope $\rho(t)$ depending on the size and the absorption coefficients of the surfaces of the room. Given the room reverberation time RT_{60} [25], that is the time required for a sound to decay 60 dB below its first reflection, the amplitude envelope is defined by

$$\rho(t) = \sigma\, 10^{-3t/\text{RT}_{60}}, \tag{8}$$

where σ is a scaling factor. We proposed in our preliminary study [9] to promote this behavior via weighted norms. Because RT_{60} is unknown a priori, it is set to an approximate value for the considered environment.

In order to assess the respective impact of the above two assumptions, we consider the following five penalties:

$$\mathcal{P}_1(\mathbf{A}) = \left\| \frac{\mathbf{A}}{\sigma} \right\|_1 = \sum_{m,n,k} \frac{|A_{m,n}(k)|}{\sigma} \tag{9}$$

$$\mathcal{P}_2(\mathbf{A}) = \frac{1}{2} \left\| \frac{\mathbf{A}}{\sigma} \right\|_2^2 = \sum_{m,n,k} \frac{|A_{m,n}(k)|^2}{2\sigma^2} \tag{10}$$

$$\mathcal{P}_{1,\rho}(\mathbf{A}) = \sum_{m,n,k} \frac{|A_{m,n}(k)|}{\rho(k)} \tag{11}$$

$$\mathcal{P}_{2,\rho}(\mathbf{A}) = \sum_{m,n,k} \frac{|A_{m,n}(k)|^2}{2\rho^2(k)} \tag{12}$$

$$\mathcal{P}_{1,2,\rho}(\mathbf{A}) = \sum_{m,n} \left(\sum_{k=0}^{k_R-1} \frac{|A_{m,n}(k)|}{\rho(k)} + \sum_{k=k_R}^{K} \frac{|A_{m,n}(k)|^2}{2\rho^2(k)} \right). \tag{13}$$

The penalties $\mathcal{P}_1$ and $\mathcal{P}_{1,\rho}$ promote sparsity while the penalties $\mathcal{P}_2$ and $\mathcal{P}_{2,\rho}$ do not, and the penalties $\mathcal{P}_{1,\rho}$ and $\mathcal{P}_{2,\rho}$ promote a decaying amplitude envelope while the penalties $\mathcal{P}_1$ and $\mathcal{P}_2$ do not. The penalty $\mathcal{P}_{1,2,\rho}$ is motivated by the assumption that sparsity holds only for small delays $k < k_R$, where k_R is a parameter to be set. The solution of (5) with the penalty $\mathcal{P}_2$ is the naive Moore-Penrose pseudo-inverse [26, p. 257], which does not rely on any assumption on the RIRs.

Note that we do not claim that real-world RIRs are actually sparse or that their amplitude envelope decays according to the assumed value of RT_{60}, which is generally not true. We simply aim to evaluate the impact of these penalties on the RIR estimation accuracy. While pseudo-inversion is expected to perform well when the problem is overdetermined, we expect other penalties to yield better results in the underdetermined case even though the RIRs to be estimated do not satisfy these assumptions. This will be confirmed in Section 5.

3.2 Convex optimization algorithm

The optimization problem (7) has the form

$$\min_{\mathbf{A}} \{\mathcal{L}(\mathbf{A}) + \lambda \mathcal{P}(\mathbf{A})\}, \tag{14}$$

where $\mathcal{L} : \mathbf{A} \mapsto \frac{1}{2}\|\mathbf{X} - \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}\|_2^2$ is a differentiable loss, $\nabla\mathcal{L}$ is Lipschitz and $\mathcal{P}$ is lower convex semi-continuous. To solve it, one can thus use the Fast Iterative Soft Thresholding Algorithm (FISTA) [27]. FISTA exploits the knowledge of the Lipschitz constant L of the gradient $\nabla\mathcal{L}$ of the loss, as well as the proximal operator [28] of the penalty $\mathcal{P}$.

Definition 1 *For $\mathcal{P}$ convex lower semi-continuous the proximal operator of $\mathcal{P}$ is the function*

$$\mathrm{prox}_\varphi : x \mapsto \mathrm{argmin}_y \left\{ \varphi(y) + \frac{1}{2}\|x - y\|_2^2 \right\}$$

The general formulation of FISTA is given in Algorithm 1.

Algorithm 1 Fast Iterative Soft Thresholding Algorithm.

1: $\mathbf{A}^0 \in \mathbb{R}^{MNK}, \tau^0 = 1$
2: **for** $q \leq q_{\max}$ **do**
$\tilde{\mathbf{A}}^q = \text{prox}_{\frac{\lambda}{L}\mathcal{P}}\left(\mathbf{A}^{q-1} - \frac{\nabla\mathcal{L}(\mathbf{A}^{q-1})}{L}\right)$
$\tau^q = \frac{1+\sqrt{1+4(\tau^{q-1})^2}}{2}$
$\mathbf{A}^q = \tilde{\mathbf{A}}^q + \frac{\tau^{q-1}-1}{\tau^q}(\tilde{\mathbf{A}}^q - \tilde{\mathbf{A}}^{q-1})$
3: **end for**

3.3 Computing the proximal operators

To fully specify the algorithm, we need to know the proximal operators of the penalties $\mathcal{P}_i$ introduced above. All penalties are separable, meaning that the operators can be processed coordinate by coordinate [29]. The penalties $\mathcal{P}_1$ and $\mathcal{P}_{1,\rho}$ are associated to weighted ℓ_1 norms, and we obtain soft thresholding operators [20] as proximity operators. The proximity operators of $\mathcal{P}_2$ and $\mathcal{P}_{2,\rho}$, associated to squared weighted ℓ_2 norms, can be obtained directly using differentiation.

Overall we obtain:

$$\text{prox}_{\alpha\mathcal{P}_1}(\mathbf{A})_{m,n,k} = \frac{A_{mn}(k)}{|A_{mn}(k)|}\left(|A_{mn}(k)| - \frac{\alpha}{\sigma}\right)^+ \tag{15}$$

$$\text{prox}_{\alpha\mathcal{P}_2}(\mathbf{A})_{m,n,k} = \frac{A_{mn}(k)}{1+\frac{\alpha}{\sigma^2}} \tag{16}$$

$$\text{prox}_{\alpha\mathcal{P}_{1,\rho}}(\mathbf{A})_{m,n,k} = \frac{A_{mn}(k)}{|A_{mn}(k)|}\left(|A_{mn}(k)| - \frac{\alpha}{\rho(k)}\right)^+ \tag{17}$$

$$\text{prox}_{\alpha\mathcal{P}_{2,\rho}}(\mathbf{A})_{m,n,k} = \frac{A_{mn}(k)}{1+\frac{\alpha}{\rho^2(k)}} \tag{18}$$

where $^+$ denotes the positive part of a real number. The proximal operator of $\alpha\mathcal{P}_{1,2,\rho}$ is expressed coordinatewise as that of $\alpha\mathcal{P}_{1,\rho}$ ($k < k_R$) or $\alpha\mathcal{P}_{2,\rho}$ ($k \geq k_R$).

3.4 Gradient of the loss and its Lipschitz constant

The computation of the gradient of $\mathcal{L}$ hinges on the introduction of the adjoint operator with respect to the truncated convolution. This construction is detailed in the Appendix.

Lemma 1 *For $n \leq N$ we define $\mathbf{S}_n^* \in \mathbb{R}^T$ with $\mathbf{S}_n^*(t) = S_n(T-t-1)$, $0 \leq t \leq T-1$, and $\mathbf{S}^* = (\mathbf{S}_1^*, \ldots, \mathbf{S}_N^*)$. We have*

$$\langle \mathbf{X}, \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}\rangle = \langle \mathbf{X} \star_{[T\text{-}1,T+K\text{-}2]} \mathbf{S}^*, \mathbf{A}\rangle. \tag{19}$$

The gradient can then be expressed as

$$\nabla\mathcal{L}(\mathbf{A}) = (\mathbf{X} - \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}) \star_{[T\text{-}1,T+K\text{-}2]} \mathbf{S}^*. \tag{20}$$

The Lipschitz constant L of $\nabla\mathcal{L}$ is the modulus of the largest eigenvalue of the operator

$$\mathbf{A} \mapsto (\mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}) \star_{[T\text{-}1,T+K\text{-}2]} \mathbf{S}^*.$$

We compute it numerically using the power iteration algorithm [20, Algorithm 5].

3.5 Pseudo-inversion for truncated RIRs

Although the penalties (9–13) are mathematically motivated, it must be remembered that the RIR length is manually fixed to a certain length K which is somewhat arbitrary. If we assume that only the first K' samples of the RIRs are nonzero with $K' \leq \frac{T}{N}$, the system becomes overdetermined and the solution is more easily computed by pseudo-inversion instead:

$$\mathbf{A}_{\text{cut}} = \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A} \star_{[0,T\text{-}1]} \mathbf{S}\|_2^2 \quad \text{s.t.} \quad \text{supp}(\mathbf{A}) \subset [\![0, K'-1]\!]. \tag{21}$$

In order to make sure that the proposed penalties bring some benefit compared to simply shortening the assumed length of the RIRs, we also consider in the following the solution of (21) for $K' = 0.9\frac{T}{N}$, where we found the overdeterminacy factor 0.9 to yield the best results experimentally. The first K' samples are computed using FISTA with the penalty $\mathcal{P}_2$ and with $\lambda \to 0$ and subsequently zero-padded to the total assumed length K. We will refer to this solution as $\mathcal{P}_{\text{cut}}$.

4 Experimental study

In order to evaluate our approach, we conducted a set of experiments using real-world recordings.

4.1 Setup

The recordings were made at IRISA, in the same room that was used to record certain signals of the Signal Separation Evaluation Campaign (SiSEC) [30]. The room is non rectangular and its dimensions are approximately $4\times5\times2.5$ m. The signals were emitted by $N = 4$ coaxial loudspeakers. The recordings were captured with $M = 10$ omnidirectional microphones. Both the sources and the microphones were randomly placed in the room. The sampling frequency was 44100 Hz both for playback and recording.

4.2 Ground truth

We first collected ground truth RIRs. The state-of-the-art choice is to use sine sweeps [17].

4.2.1 Acquisition process

We sent $r = 20$ linear sine sweeps from 50 Hz to 22000 Hz. Each sine sweep had a duration of 2 s, and was followed by a silence of 1 s. We then computed the average of these 20 recordings and estimated the RIRs by Fourier-domain inversion.

4.2.2 Assumed duration K of the RIRs

The obtained RIRs displayed a typical behavior: after a first part dominated by the direct path and first reflections, an exponentially decaying behavior was observed until the noise level was reached, after about 300 ms. For this reason we chose to fix the length of the RIRs to $K = 300$ ms or 13230 samples.

4.2.3 Characterization of the background noise

The acquired recordings suffered from a strong low-frequency background noise, possibly due to air conditioning in the room. This prevented the evaluation of the estimated RIRs at these frequencies, since both the estimated and the ground truth RIRs were dominated by noise. For this reason, in the rest of the study, we chose to measure the estimation accuracy by comparing the high-frequency part of the estimated RIRs with that of the ground truth RIRs. Visualization of the spectrum of the noise suggested to keep all frequencies above 100 Hz.

4.3 Performance measure

The usual "noise level" metric employed for the assessment of RIR estimation is, as stated in [17], "*the ratio expressed in dB between the average power of the signal recorded by the microphone and the average power of the noise and distortions present in the tail of the deconvolved (linear) impulse response.*" This metric implicitly assumes that the difference between the estimated RIR and the true RIR is a stationary signal, so that the amount of noise and distortion in the tail is proportional to the total estimation error.

In the underdetermined context considered in this paper, the linear inverse problem (2) admits infinitely many solutions, most of which are completely inconsistent with this assumption. Therefore we need a performance measure that reflects the estimation accuracy with respect to the ground truth. As a measurement of the error between the estimated RIRs $\hat{\mathbf{A}}$ and the ground truth RIRs $\mathbf{A}$ (in fact, the high-pass versions of $\hat{\mathbf{A}}$ and $\mathbf{A}$ as seen above), we propose to use the following SNR in dB

$$\mathrm{SNR}_{\mathbf{A}}(\hat{\mathbf{A}}) = 10 \log_{10} \frac{\|\mathbf{A}\|_2^2}{\|\hat{\mathbf{A}} - \mathbf{A}\|_2^2}. \tag{22}$$

We will conduct in Section 5.2.2 a short qualitative study showing that a $\mathrm{SNR}_{\mathbf{A}}$ on the order of 15 dB is very satisfactory and that it corresponds for the chosen penalties to a "noise level" on the order of 50 dB.

A first set of performance figures is given in Table 2, where we compare the RIRs $\mathbf{A}_r$ estimated by averaging r recorded sine sweeps to the ground truth $\mathbf{A} = \mathbf{A}_{20}$ obtained with $r = 20$.

Items averaged r	1	5	10	15	20
$\mathrm{SNR}_{\mathbf{A}}(\mathbf{A}_r)$ (dB)	26.8	28.7	32.7	38.5	∞

Table 2: Influence of averaging on the acquisition of the ground truth.

While the $\mathrm{SNR}_{\mathbf{A}}$ quantifies the RIR estimation accuracy for a given estimation technique, it is also desirable to quantify the level of noise and nonlinear distortion present in the recorded signals from which the RIRs are estimated. For this, we use the SNR of the recording $\mathbf{X}$ (in fact, its high-pass version) defined as

$$\mathrm{SNR}_{\mathbf{X}}(\mathbf{X}, \mathbf{S}) = 10 \log_{10} \frac{\|\mathbf{A} \star_{[0,T-1]} \mathbf{S}\|_2^2}{\|\mathbf{X} - \mathbf{A} \star_{[0,T-1]} \mathbf{S}\|_2^2} \tag{23}$$

where $\mathbf{A}$ are the ground RIRs.

4.4 Parameters of the proposed approach

After collecting the ground truth RIRs, we made additional recordings within the same recording session and processed them via the proposed algorithm.

4.4.1 Source signals

Signals of different durations were emitted, including silence or not. Several recordings were made, for $N = 2, 4$ sources, and 5 types of signals :

- uniform random noise in $[-1, 1]$;
- Bernoulli noise generated by a Bernouilli process on $\{-1, 1\}$ with probability $p = \frac{1}{2}$;
- MLS sequences described above;

- and human speech excerpts.

The emitted signals were normalized according to their maximum amplitude.

4.4.2 Parameters of the considered penalties

The scaling factor σ for all penalties is set to $\sigma = 1$. Given that near-optimal performance is empirically obtained for $\lambda \to 0$, this parameter has in fact essentially no impact on the performance.

We consider different values of the reverberation parameter RT_{60} in $\mathcal{P}_{1,\rho}$, $\mathcal{P}_{2,\rho}$ and $\mathcal{P}_{1,2,\rho}$ between 50 ms and 1 s. Fig. 1 shows two visualizations of one of the ground truth RIRs, rescaled to a maximum amplitude of 1. The true value of the room reverberation time computed using Schroeder's backward integration method [31] is $RT_{60} = 380$ ms. The assumption that the amplitude decays exponentially is clearly visible on the logarithmic view. Finally, the parameter k_R of $\mathcal{P}_{1,2,\rho}$ is set to $K/3 = 100$ ms.

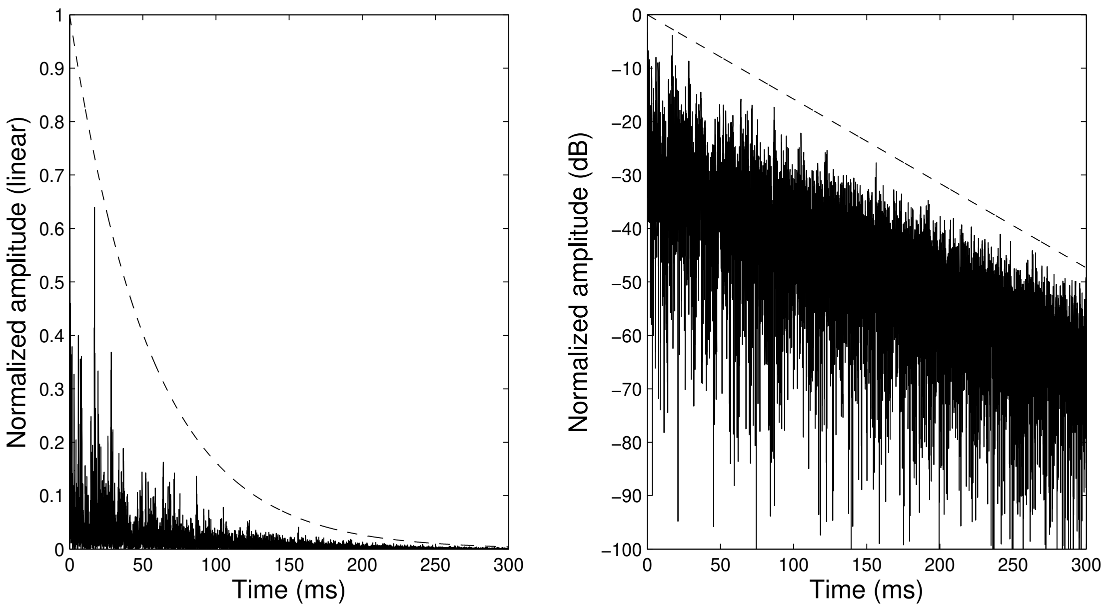

Figure 1: Linear and logarithmic view of one ground truth RIR (plain) compared to the amplitude envelope ρ (dashed) with $RT_{60} = 380$ ms. The experiments in Section 5.2.1 will show that an exact value of RT_{60} is not necessary to obtain good RIR estimates with the penalty $\mathcal{P}_{1,\rho}$.

4.4.3 Parameters of FISTA

The examination of Algorithm 1 and the expressions (15–20) reveals that the variables relating to different microphones m do not intervene with each other. This is consistent with the fact that the cost function (7) is additive with respect to m. Therefore, we equivalently apply FISTA to each microphone in turn.

The estimation of the Lipschitz constant requires 200 iterations of the power iteration algorithm. We know [32] that like many algorithms solving (7), FISTA requires a large number of iterations for small values of λ. In this situation we use the *continuation trick* also known as *warm start* : we run the algorithm for several decreasing values of λ and initialize each run at the solution of the previous run. We run 16 instances of FISTA, using decreasing values $\lambda = \{10^0, \dots, 10^{-15}\}$. The convergence of FISTA is reached for every λ in about 500 iterations.

Theoretically, we expect to achieve the best results for a specific nonzero λ but, given that the noisy low-frequency components are not taken into account in the performance measure, the noise level is low enough to neglect its influence and to consider the smallest value of λ, which approximates the limit when $\lambda \to 0$. This will be confirmed experimentally in Section 5.5.

Using Matlab on a dual-core 3.40 GHz CPU, the computation time is on the order of 20 min per microphone, per source and per second of recorded signal.

5 Experimental results

5.1 Choice of the source signals

5.1.1 Comparison between different types of sources

We first assess the impact of the choice of different source signals without silence in the case of an overdetermined system with $T = 2\,T^{\text{crit}}$ for $N = 2$ sources, inverted using FISTA with $\mathcal{P}_2$ and with $\lambda \to 0$. Table 3 shows the link between the RMS amplitude of the sources, the $\text{SNR}_{\mathbf{X}}$ of the recording, and the $\text{SNR}_{\mathbf{A}}$ of the estimated RIRs. Although Bernoulli and MLS signals potentially induce more nonlinearities than other signals, their higher RMS induces weaker noise, which altogether yields higher $\text{SNR}_{\mathbf{X}}$ and $\text{SNR}_{\mathbf{A}}$[2].

	Speech	Uniform	Bernoulli	MLS
RMS (dB)	-17.9	-4.8	0	0
$\text{SNR}_{\mathbf{X}}$ (dB)	16.2	17.1	18.2	18.3
$\text{SNR}_{\mathbf{A}}$ (dB)	16.4	18.2	22.2	22.1

Table 3: Relation between RMS, $\text{SNR}_{\mathbf{X}}$ and $\text{SNR}_{\mathbf{A}}$ for $T = 2\,T^{\text{crit}}$ depending on the chosen source signals.

5.1.2 Influence of silence within the source signals

It is common in state-of-the-art methods to leave a silence between successive recordings, to make sure that the convolution is complete. However, including a silence of length L within a signal of length T decreases $\text{SNR}_{\mathbf{X}}$ by up to $10\log_{10}(1 - L/T)$ dB. This quantity grows as the system becomes more underdetermined. As an example, for the setup studied in the next section with $T \simeq 2K$, a silence of length $L = K$ would result in a loss of up to 6 dB of $\text{SNR}_{\mathbf{X}}$. We found in a preliminary experiment that this resulted in a similar or even bigger loss of $\text{SNR}_{\mathbf{A}}$. We will therefore use Bernoulli signals without silence in all the following experiments.

5.2 Performance of our method for $T = 0.45\,T^{\text{crit}}$

5.2.1 Influence of the penalty

As an example of the results obtained in an underdetermined setting, we compare in Table 4 the performance of different penalties with $T = 544$ ms $= 0.45\,T^{\text{crit}}$ for $N = 4$ sources. This corresponds to a reduction of the recording time by a factor of 2.2 with respect to the critical time T^{crit}, which is itself smaller than the recording time required by state-of-the-art methods (see Table 1).

[2] We remind that $\text{SNR}_{\mathbf{X}}$ and $\text{SNR}_{\mathbf{A}}$ account for the effect of both nonlinearities and noise.

Penalty	$\mathcal{P}_{1,\rho}$	$\mathcal{P}_{2,\rho}$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_{1,2,\rho}$	$\mathcal{P}_{\text{cut}}$
$\text{SNR}_{\mathbf{A}}$ (dB)	15.0	15.8	12.4	0.0	4.5	12.0

Table 4: RIR accuracy depending on the chosen penalty for $T = 0.45\,T^{\text{crit}}$, given the true RT_{60}.

Unsurprisingly in this setting, naive pseudo inversion via $\mathcal{P}_2$ completely fails. The unweighted ℓ^1 norm $\mathcal{P}_1$ and the RIR shortening approach $\mathcal{P}_{\text{cut}}$ are able to recover the RIRs to a certain extent, which is a good result given that no knowledge of RT_{60} is needed. However, the best results achieved by the weighted norms $\mathcal{P}_{1,\rho}$ and $\mathcal{P}_{2,\rho}$ which provide a $\text{SNR}_{\mathbf{A}}$ on the order of 15 dB. This shows the importance of promoting an exponential decaying envelope via the penalty. The hybrid penalty $\mathcal{P}_{1,2,\rho}$ performs worse, which may be attributed to a lack of robustness to the choice of the extra parameter k_R.

5.2.2 Qualitative analysis of the resulting RIRs

Fig. 2 depicts one the RIRs estimated using $\mathcal{P}_{1,\rho}$ compared to the ground truth. The global shape of the RIR is well recovered down to -50 dB. The $\text{SNR}_{\mathbf{A}}$ of 15 dB therefore appears to correspond to a *noise level* of 50 dB, following the state-of-the-art performance measure. A zoom on the first coefficients in Fig. 3 confirms the accuracy of the estimation. In particular, the times of arrival of the first reflections are exactly estimated.

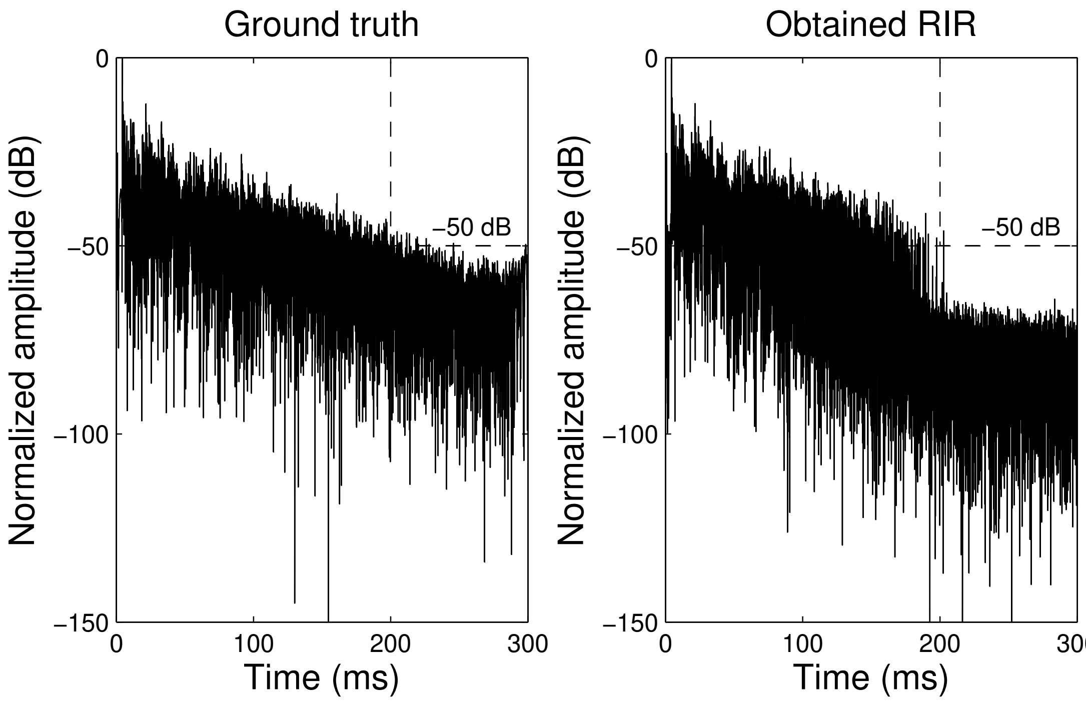

Figure 2: Logarithmic view of one the RIRs estimated using $\mathcal{P}_{1,\rho}$ for $T = 0.45\,T^{\text{crit}}$, compared to the ground truth.

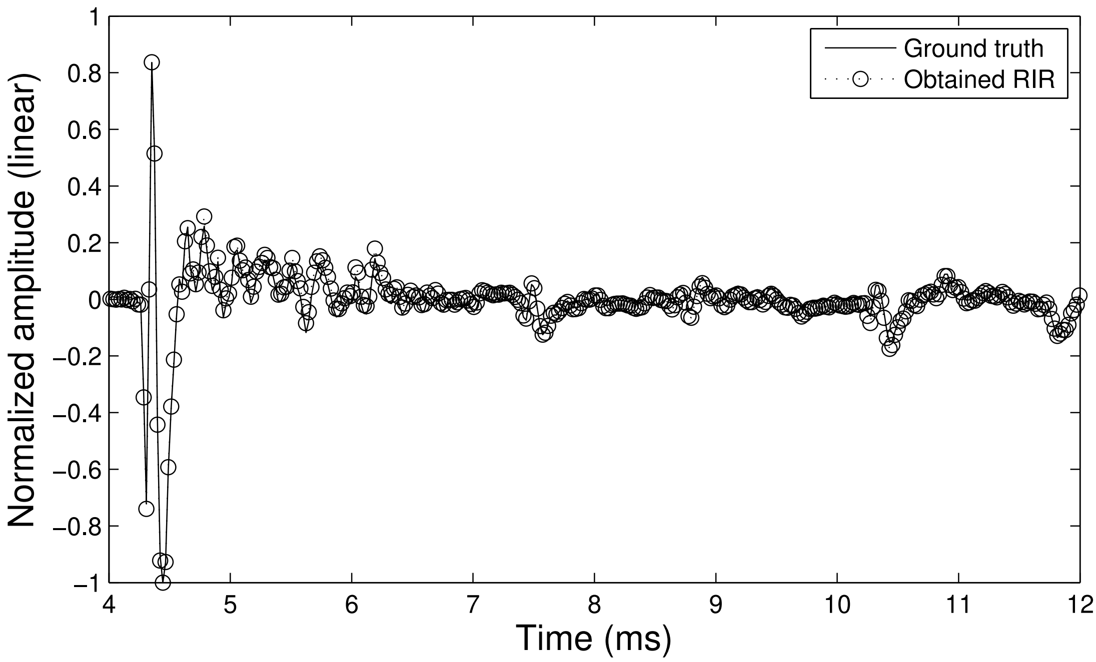

Figure 3: Linear view of the first coefficients of the estimated RIR and the ground truth in Fig. 2.

5.3 Robustness to an erroneous reverberation time

Fig. 4 bears witness to the robustness of the penalties to a bad guess of the room reverberation time RT_{60}. The weighted ℓ^2 penalty $\mathcal{P}_{2,\rho}$ outperforms the unweighted ℓ^1 penalty $\mathcal{P}_1$ for any RT_{60} between 150 ms and 600 ms. However, its performance drops quickly above that value. By contrast, the weighted ℓ^1 penalty $\mathcal{P}_{1,\rho}$, which promotes both sparsity and exponential amplitude envelope, exhibits remarkable robustness and outperforms $\mathcal{P}_1$ for all RT_{60} above 170 ms. Once again, the hybrid penalty $\mathcal{P}_{1,2,\rho}$ appears to be less robust due its extra parameter k_R. For this reason, we select $\mathcal{P}_{1,\rho}$ as the best penalty in the remaining experiments.

5.4 Influence of the recording time T

Fig. 5 shows the performance as a function of the recording length T, where $T^{\text{crit}} = 1200$ ms. While the performance of $\mathcal{P}_2$ is consistently low, that of $\mathcal{P}_{1,\rho}$ and $\mathcal{P}_{\text{cut}}$ appear to degrade gracefully as the recording time decreases. For instance, $\mathcal{P}_{1,\rho}$ still allows the recovery of the RIRs with more than 10 dB of $\mathrm{SNR}_{\mathbf{A}}$ with $T = 300$ ms $= 0.25\,T^{\text{crit}}$, which corresponds to a reduction of the recording time by a factor of 4. Note also that $\mathcal{P}_{1,\rho}$ outperforms $\mathcal{P}_{\text{cut}}$ as soon as $T \lesssim 0.6\,T^{\text{crit}}$.

5.5 Choice of the Lagrangian parameter λ

While all the above results have been shown for $\lambda \to 0$, we expect that the best results are achieved for a specific nonzero λ due to the presence of noise and nonlinearities. The analysis of the performance of $\mathcal{P}_{1,\rho}$ as a function of λ in Fig. 6 shows that, as the system becomes more underdetermined, the gain obtained by choosing the optimal λ becomes smaller. For $T = 0.45\,T^{\text{crit}}$, a gain of about 0.5 dB is obtained for the optimal $\lambda = 10^{-2}$. However, the decrease of performance is observed for larger values

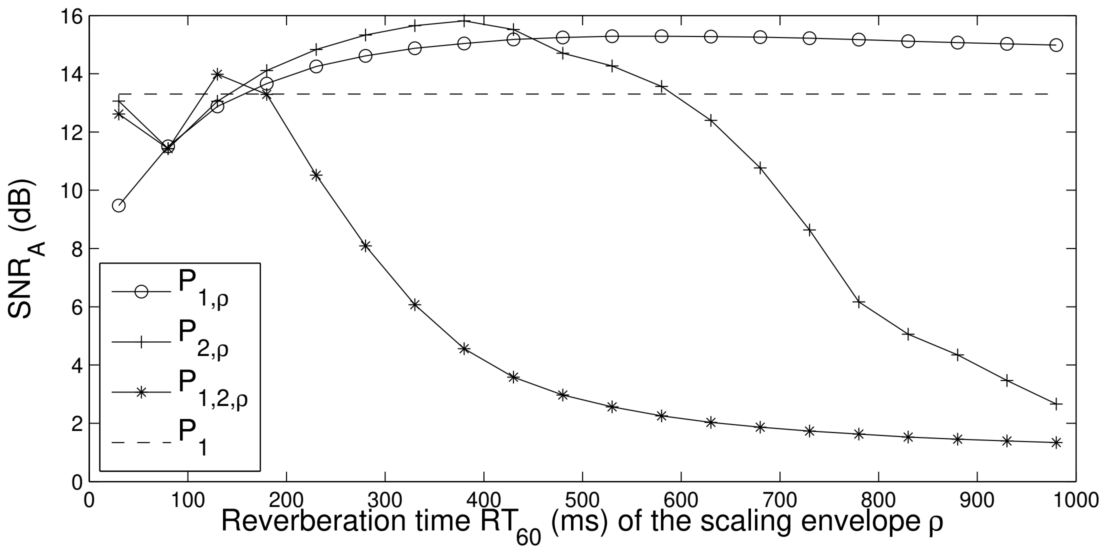

Figure 4: Influence of the parameter of the amplitude envelope on the RIR accuracy for $T = 0.45\,T^{\text{crit}}$.

of λ. Although there is theoretically a link between λ and the background noise level, there is no way to predict the optimal value of λ to our knowledge. The choice $\lambda \to 0$ therefore appears to be the most robust.

6 Conclusion

We proposed an algorithm to estimate RIRs from recordings of multiple active loudspeakers where the number of recorded samples is smaller than the number of RIR samples to be estimated. This algorithm relies on convex penalties incorporating knowledge about the RIRs. We investigated both existing and new penalties and showed that the penalty $\mathcal{P}_{1,\rho}$ promoting sparsity and exponential amplitude envelope is the most robust. These two assumptions on the RIRs have hence been proven to be beneficial for the purpose of regularization, although actual RIRs do not satisfy them exactly. We also introduced a new evaluation methodology based on comparing the estimated RIRs with ground truth RIRs and quantified the influence of the choice of the emitted signals.

Following the described framework, further experiments could be performed to expand this technique to other acoustic responses such as BRIRs. The estimation of RIRs is also an important problem in blind source separation, where they are called mixing filters. The proposed algorithm is a first brick towards a new algorithm for joint estimation of the source signals and the mixing filters which would make use of the proposed RIR regularization.

A Computation of the adjoint operator

The computation of $\nabla\mathcal{L}$ boils down to that of the adjoint operator of the truncated matrix convolution product $\star_{[0,T\text{-}1]}$.

The convolution with the RIR is causal. A convenient way to model its convolution is to see the

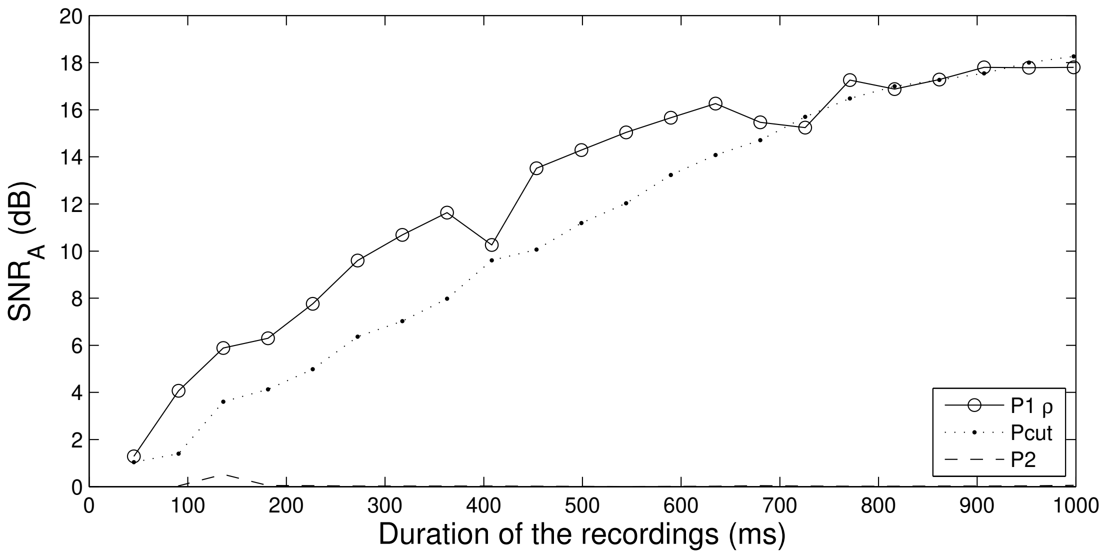

Figure 5: Performance of the different penalties as a function of the recording length T.

signals in $\ell^2(\mathbb{Z})$, with a finite support. For $x, y \in \ell^2(\mathbb{Z})$ we denote by $*$ the standard convolution

$$x * y(t) = \sum_{\tau \in \mathbb{Z}} x(t) y(\tau - t). \tag{24}$$

For $T \in \mathbb{N}$, we define the truncation operator:

$$P_T^* : \begin{array}{ccc} \mathbb{R}^{\mathbb{Z}} & \longrightarrow & \mathbb{R}^T \\ (x_t)_{t \in \mathbb{Z}} & \longmapsto & (x_t)_{0 \le t \le T-1} \end{array} \tag{25}$$

and its adjoint, the double-sided zero-padding operator

$$P_T : \begin{array}{ccc} \mathbb{R}^T & \longrightarrow & \mathbb{R}^{\mathbb{Z}} \\ (x_0, \ldots, x_{T-1}) & \longmapsto & (\ldots 0, x_0, \ldots, x_{T-1}, 0, \ldots). \end{array} \tag{26}$$

Now consider $\mathbf{x} \in \mathbb{R}^T, \mathbf{s} \in \mathbb{R}^T, \mathbf{a} \in \mathbb{R}^K$. The definition of the truncated convolution product $\star_{[0,T\text{-}1]}$ is

$$\mathbf{a} \star_{[0,T\text{-}1]} \mathbf{s} = P_T^*(P_K(\mathbf{a}) * P_T(\mathbf{s})). \tag{27}$$

For $x, s, a \in \ell^2(\mathbb{Z})$, denoting $\bar{s}(t) = s(-t)$, $t \in \mathbb{Z}$, we have:

$$\langle x, a * s \rangle = \langle x * \bar{s}, a \rangle \tag{28}$$

Then we can write

$$\begin{aligned} \langle \mathbf{x}, \mathbf{a} \star_{[0,T\text{-}1]} \mathbf{s} \rangle &= \langle \mathbf{x}, P_T^*(P_K(\mathbf{a}) * P_T(\mathbf{s})) \rangle \\ &= \langle P_T(\mathbf{x}), P_K(\mathbf{a}) * P_T(\mathbf{s}) \rangle \\ &= \langle P_T(\mathbf{x}) * \overline{P_T(\mathbf{s})}, P_K(\mathbf{a}) \rangle \\ &= \langle P_K^* \left(P_T(\mathbf{x}) * \overline{P_T(\mathbf{s})} \right), \mathbf{a} \rangle, \end{aligned} \tag{29}$$

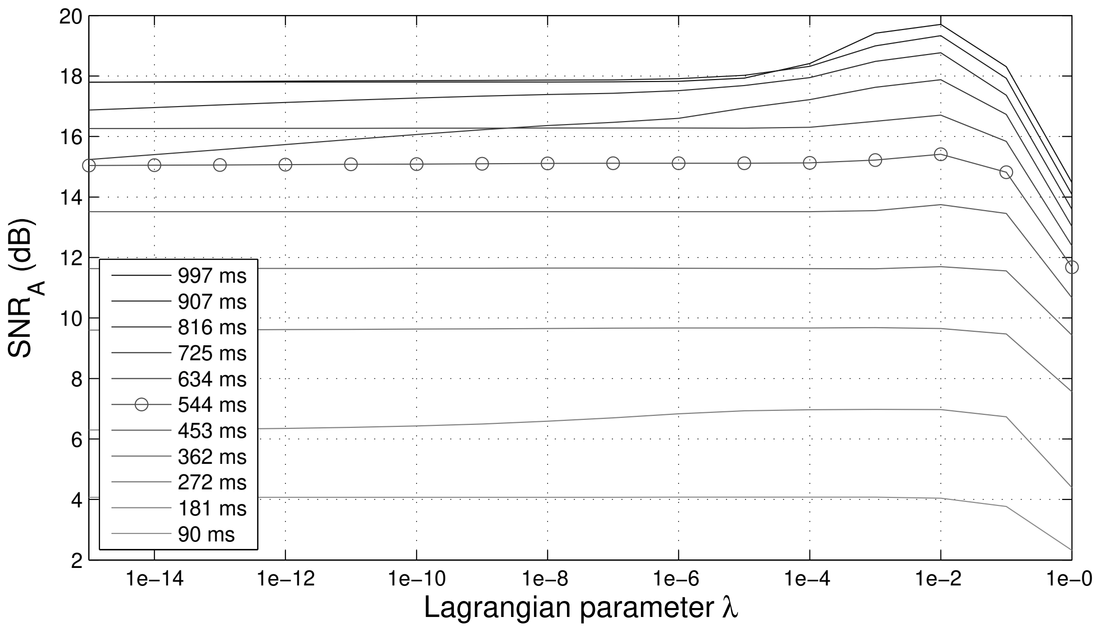

Figure 6: Influence of the Lagrangian parameter λ on the accuracy of the RIRs obtained with $\mathcal{P}_{1,\rho}$ for several recording lengths T.

where we used the notation of (28).

There remains to express $P_K^*(P_T(\mathbf{x}) * \overline{P_T(\mathbf{s})})$ as a truncated convolution. Since $P_T(\mathbf{s})$ is supported on $[\![0, T-1]\!]$, its time reversed version $\overline{P_T(\mathbf{s})}$ is supported on $[\![-(T-1), 0]\!]$. Define $\mathbf{s}^* \in \mathbb{R}^T$ by $\mathbf{s}^*(t) := \mathbf{s}(T-1-t)$, $0 \leq t \leq T-1$. We have $\overline{P_T(\mathbf{s})} = \delta_{-(T-1)} * P_T(\mathbf{s}^*)$, hence we can write

$$\begin{aligned} P_K^*\left(P_T(\mathbf{x}) * \overline{P_T(\mathbf{s})}\right) &= P_K^*\left(\delta_{-(T-1)} * P_T(\mathbf{x}) * P_T(\mathbf{s}^*)\right) \\ &= \mathbf{x} \star_{[T-1,T+K-2]} \mathbf{s}^* \end{aligned} \tag{30}$$

where the last equality comes from the fact that $P_K^*(\delta_{-(T-1)} * u)$ is the restriction of the sequence $u \in \ell^2(\mathbb{Z})$ to the interval $[\![T-1, (T-1)+(K-1)]\!]$.

The multichannel and multisource case $M, N \geq 1$ is now straightforward. For $1 \leq n \leq N$ we define $\mathbf{S}_n^* \in \mathbb{R}^T$ the time reversal of the source signal $\mathbf{S}_n$, i.e., for $0 \leq t \leq T-1$, $\mathbf{S}_n^*(t) = \mathbf{S}_n(T-1-t)$, and $\mathbf{S}^* = (\mathbf{S}_1^*, \ldots, \mathbf{S}_N^*)$. Using these notations and the previous computation the following holds

$$\begin{aligned} \langle \mathbf{X}, \mathbf{A} \star_{[0,T-1]} \mathbf{S} \rangle &= \left\langle \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_M \end{pmatrix} \star_{[T-1,T+K-2]} (\mathbf{S}_1^*, \ldots \mathbf{S}_N^*), \mathbf{A} \right\rangle \\ &= \langle \mathbf{X} \star_{[T-1,T+K-2]} \mathbf{S}^*, \mathbf{A} \rangle. \end{aligned} \tag{31}$$

References

[1] A. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, pp. 2764–2778, 1993.

[2] S. Spors, D. Seuberth, and R. Rabenstein, “Multiactuator panels for wave field synthesis: Evolution and present developments,” *Journal for Audio Engineering Society*, vol. 58, no. 12, pp. 1045–1063, 2010.

[3] E. Corteel, “Equalization in an extended area using multichannel inversion and wave field synthesis,” *Journal of the Audio Engineering Society*, vol. 54, no. 12, pp. 1140–1161, 2006.

[4] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009, article ID 298605.

[5] D. Donoho, “Compressed sensing,” *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] Y. Lin, J. Chen, Y. Kim, and D. Lee, “Blind channel identification for speech dereverberation using $\ell 1$-norm sparse learning,” in *Advances in Neural Information Processing Systems 20*. MIT Press, 2007, pp. 921–928.

[7] P. Sudhakar, S. Arberet, and R. Gribonval, “Double sparsity: Towards blind estimation of multiple channels,” *in Proc. Int. Conf. on Latent Variable Analysis and Signal Separation*, pp. 571–578, 2010.

[8] R. Mignot, L. Daudet, and F. Ollivier, “Compressed sensing for acoustic response reconstruction: interpolation of the early part,” in *Proc. 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 225–228.

[9] A. Benichoux, E. Vincent, and R. Gribonval, “A compressed sensing approach to the simultaneous recording of multiple room impulse responses,” *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 285–288, 2011.

[10] K. Bodlund, “On the use of the integrated impulse response method for laboratory reverberation measurements,” *Journal of Sound and Vibration*, vol. 56, no. 3, pp. 341–362, 1978.

[11] J. S. Abel, N. Bryan, P. Huang, M. Kolar, and B. Pentcheva, “On estimating room impulse responses from recorded balloon pops,” in *129th Audio Engineering Society Convention*, no. 8171, 2010.

[12] M. Schroeder, “Integrated-impulse method measuring sound decay without using impulses,” *The Journal of the Acoustical Society of America*, vol. 66, pp. 497–500, 1979.

[13] W. Chu, “Impulse-response and reverberation-decay measurements made by using a periodic pseudorandom sequence,” *Applied Acoustics*, vol. 29, no. 3, pp. 193–205, 1990.

[14] A. Gonzalez, P. Zuccarello, G. Pinero, and M. de Diego, “Simultaneous measurement of multichannel acoustic systems,” *Journal of the Audio Engineering Society*, vol. 52, no. 1/2, pp. 26–42, 2004.

[15] M. Wright, “Comments on aspects of MLS measuring systems,” *Journal of the Audio Engineering Society*, vol. 43, no. 1, pp. 48–49, 1995.

[16] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Proc. AES 108th Convention*, 2000, pp. 18–22.

[17] G. Stan, J. Embrechts, and D. Archambeau, “Comparison of different impulse response measurement techniques,” *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.

[18] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 8, pp. 2220–2229, 2010.

[19] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, pp. 623–637, 2007.

[20] M. Kowalski, E. Vincent, and R. Gribonval, "Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1818–1829, 2010.

[21] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[22] D. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[23] S. Winter, W. Kellermann, H. Sawada, and S. Makino, "MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and l1-norm minimization," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 81–81, 2007.

[24] Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?" *IEEE Transactions on Signal Processing*, 2010.

[25] H. Kuttruff, *Room Acoustics*, 4th ed. New York: CRC Press, 2000, no. 0419245804.

[26] G. Golub and C. F. Van Loan, *Matrix computations (3rd ed.)*. Johns Hopkins University Press, 1996.

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

[28] J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CR Acad. Sci. Paris Sér. A Math*, vol. 255, pp. 2897–2899, 1962.

[29] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2006.

[30] E. Vincent, S. Araki, F. J. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, B. V. Gowreesunker, D. Lutter, and N. Q. K. Duong, "The Signal Separation Evaluation Campaign (2007–2010): Achievements and remaining challenges," *Signal Processing*, vol. 92, pp. 1928–1936, 2012.

[31] M. R. Schroeder, "New method of measuring reverberation time," *Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965.

[32] I. Loris, "On the performance of algorithms for the minimization of 1-penalized functionals," *Inverse Problems*, vol. 25, pp. 35 008–35 023, 2009.

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399