Signal-Dependent Penalty Functions for Robust Acoustic Multi-Channel Equalization

Ina Kodrasi, *Member, IEEE* and Simon Doclo, *Senior Member, IEEE*

Abstract—Acoustic multi-channel equalization techniques, which aim to achieve dereverberation by reshaping the room impulse responses (RIRs) between the source and the microphone array, are known to be highly sensitive to RIR perturbations. In order to increase the robustness against RIR perturbations, several *signal-independent* methods have been proposed, which only rely on the available perturbed RIRs and do not incorporate any knowledge about the output signal. This paper presents a novel *signal-dependent* method to increase the robustness of equalization techniques by enforcing the output signal to exhibit spectro-temporal characteristics of a clean speech signal. Motivated by the sparse nature of clean speech, we propose to extend the cost function of state-of-the-art least-squares equalization techniques, i.e., the multiple-input/output inverse theorem (MINT), relaxed multi-channel least-squares (RMCLS), and partial multi-channel equalization based on MINT (PMINT), with a signal-dependent penalty function promoting sparsity of the output signal in the short-time Fourier transform domain. Three conventionally used sparsity-promoting penalty functions are investigated, i.e., the l_0-norm, the l_1-norm, and the weighted l_1-norm, and the sparsity-promoting reshaping filters are iteratively computed using the alternating direction method of multipliers. Simulation results for several acoustic systems and RIR perturbations demonstrate that incorporating sparsity-promoting penalty functions significantly increases the robustness of MINT, RMCLS, and PMINT, with the weighted l_1-norm typically outperforming the l_0-norm and the l_1-norm. Furthermore, it is shown that the weighted l_1-norm sparsity-promoting PMINT technique outperforms the other sparsity-promoting techniques in terms of perceptual speech quality. Finally, it is shown that the signal-dependent weighted l_1-norm sparsity-promoting PMINT technique yields a similar or better dereverberation performance than the signal-independent regularized PMINT technique, confirming the advantage of using signal-dependent penalty functions for robust dereverberation filter design.

Index Terms—acoustic multi-channel equalization, sparsity, signal-dependent penalty function, ADMM

I. INTRODUCTION

IN many hands-free speech communication applications such as teleconferencing, voice-controlled systems, or hearing aids, the recorded microphone signals do not only contain the desired signal, but also attenuated and delayed copies due to reverberation. Reverberation causes signal degradation, typically leading to decreased speech intelligibility [1]–[3] and performance deterioration in automatic speech recognition systems [4], [5]. Hence, in the last decades several single-

and multi-channel dereverberation techniques have been proposed [6], with multi-channel techniques being generally preferred since they enable to exploit both the spectro-temporal and the spatial characteristics of the microphone signals. Existing multi-channel dereverberation techniques can be broadly classified into: i) spectral enhancement techniques aiming at suppressing the late reverberation in the spectral domain based on an estimate of the late reverberant power spectral density [7]–[10], ii) probabilistic modeling-based techniques aiming at blindly estimating the parameters of statistical models assumed to represent the clean speech signal and the room impulse responses (RIRs) [11]–[14], and iii) acoustic multi-channel equalization techniques aiming at reshaping the available (measured or estimated) RIRs between the source and the microphone array [15]–[20]. Unlike spectral enhancement and probabilistic modeling-based techniques, acoustic multi-channel equalization techniques can in theory achieve perfect dereverberation [15], [21], therefore comprising an attractive approach to speech dereverberation.

A well-known multi-channel equalization technique aiming at acoustic system inversion is the multiple-input/output inverse theorem (MINT) technique [15], which however suffers from drawbacks in practice. Since the available RIRs typically differ from the true RIRs due to fluctuations (e.g., temperature or position variations [22]) or due to the sensitivity of blind system identification (BSI) and supervised system identification (SSI) methods to near-common zeros or interfering noise [23]–[25], MINT fails to invert the true RIRs, possibly leading to severe distortions in the output signal [18], [19], [26].

In order to increase the robustness against RIR perturbations, partial multi-channel equalization techniques such as relaxed multi-channel least-squares (RMCLS) [18] and partial multi-channel equalization based on MINT (PMINT) [19] have been recently proposed. Since early reflections tend to improve speech intelligibility [27], [28] and late reflections are the major cause of speech intelligibility degradation, the objective of these techniques is to suppress only the late reflections. Although partial equalization techniques can be significantly more robust than MINT, their performance still remains rather susceptible to RIR perturbations [19]. As a result, several methods have been proposed to further increase the robustness against RIR perturbations. In [29], [30] it has been proposed to relax the constraints on the filter design by constructing approximate dereverberation filters in the subband domain. In [31] it has been proposed to use a shorter filter length than conventionally used, such that a better-conditioned optimization criterion is obtained. In [19], [26] it has been proposed to incorporate regularization in the filter design,

such that the distortion energy due to RIR perturbations is decreased. To the best of our knowledge, all proposed methods are *signal-independent*, i.e., they only rely on the available perturbed RIRs and do not incorporate any knowledge about the output signal.

The objective of this paper is to explore the potential of increasing the robustness of acoustic multi-channel equalization techniques by incorporating a *signal-dependent* penalty function in the filter design, enforcing the output signal to exhibit spectro-temporal characteristics of a clean speech signal. Given the successful exploitation of the *sparse* nature of clean speech in many speech enhancement techniques, e.g., in underdetermined blind source separation [32], [33], adaptive beamforming [34], or probabilistic modeling-based dereverberation [14], [35], [36], we propose to extend the MINT, RMCLS, and PMINT cost functions with a penalty function which promotes sparsity of the output signal in the short-time Fourier transform (STFT) domain. The incorporation of three commonly used sparsity-promoting penalty functions is investigated, i.e., the l_0-norm, the l_1-norm, and the weighted l_1-norm. Since no closed-form expression is available for the filter minimizing these sparsity-promoting cost functions, we propose to iteratively compute the sparsity-promoting filters using the alternating direction method of multipliers (ADMM), which is a well-suited method for solving large-scale optimization problems with sparsity-promoting penalty functions [37]. Some preliminary results on incorporating the l_0- and l_1-norm penalty functions in PMINT have been presented in [38].

Simulation results for several acoustic systems and RIR perturbations show by means of instrumental measures that incorporating sparsity-promoting penalty functions in MINT, RMCLS, and PMINT significantly increases the robustness against RIR perturbations. Furthermore, it is shown that incorporating the weighted l_1-norm penalty function is typically more advantageous than incorporating the l_0- or l_1-norm penalty functions, with the weighted l_1-norm sparsity-promoting PMINT technique generally yielding the best dereverberation performance. Finally, it is shown that the signal-dependent sparsity-promoting PMINT technique using the weighted l_1-norm yields a similar or better dereverberation performance than the signal-independent regularized PMINT technique.

The paper is organized as follows. In Section II the considered acoustic configuration and the used notation is introduced. In Section III state-of-the-art least-squares acoustic multi-channel equalization techniques, i.e., MINT, RMCLS, and PMINT, are briefly reviewed, and in Section IV the incorporation of regularization to increase their robustness against RIR perturbations is discussed. In Section V the novel signal-dependent cost function is formulated based on several sparsity-promoting penalty functions, and the iterative method optimizing this cost function is presented. Using instrumental measures, the dereverberation performance of all considered techniques is compared in Section VI.

II. CONFIGURATION AND NOTATION

We consider an acoustic system with a single speech source and M microphones, as depicted in Fig. 1. The m-th micro-
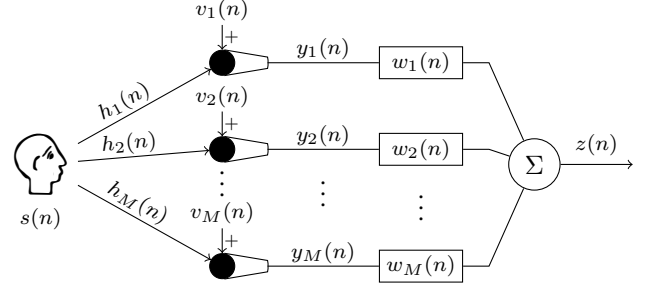

Fig. 1: Acoustic system configuration.

phone signal, $m = 1, 2, \ldots, M$, at time index n is given by

$$y_m(n) = \underbrace{h_m(n) * s(n)}_{x_m(n)} + v_m(n) = x_m(n) + v_m(n), \quad (1)$$

where $*$ denotes convolution, $s(n)$ is the clean speech signal, $h_m(n)$ is the RIR between the source and the m-th microphone, $x_m(n)$ is the reverberant speech component, and $v_m(n)$ is the noise component. Since acoustic multi-channel equalization techniques generally design dereverberation filters disregarding the presence of the additive noise, in the following it is assumed that $v_m(n) = 0$, hence, $y_m(n) = x_m(n)$. Using the filter-and-sum structure in Fig. 1, the output signal $z(n)$ is equal to the sum of the filtered microphone signals, i.e.,

$$z(n) = \sum_{m=1}^{M} x_m(n) * w_m(n) \quad (2)$$

$$= s(n) * \underbrace{\sum_{m=1}^{M} h_m(n) * w_m(n)}_{c(n)}, \quad (3)$$

where $w_m(n)$ is the filter applied to the m-th microphone signal and $c(n)$ denotes the equalized impulse response (EIR) between the source and the output of the system. In vector notation, the RIR $\mathbf{h}_m$ and the filter $\mathbf{w}_m$ are given by

$$\mathbf{h}_m = [h_m(0) \ h_m(1) \ \ldots \ h_m(L_h - 1)]^T, \quad (4)$$

$$\mathbf{w}_m = [w_m(0) \ w_m(1) \ \ldots \ w_m(L_w - 1)]^T, \quad (5)$$

where L_h and L_w denote the RIR length and the filter length, respectively. Using the ML_w–dimensional stacked filter vector $\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \ldots \ \mathbf{w}_M^T]^T$, the EIR vector $\mathbf{c}$ of length $L_c = L_h + L_w - 1$, i.e., $\mathbf{c} = [c(0) \ c(1) \ \ldots \ c(L_c - 1)]^T$, can be expressed as [15], [19]

$$\mathbf{c} = \mathbf{Hw}, \quad (6)$$

with $\mathbf{H}$ the $L_c \times ML_w$–dimensional multi-channel convolution matrix of the RIRs, i.e., $\mathbf{H} = [\mathbf{H}_1\ \mathbf{H}_2\ \ldots\ \mathbf{H}_M]$, and

$$\mathbf{H}_m = \begin{bmatrix} h_m(0) & 0 & \ldots & 0 \\ h_m(1) & h_m(0) & \ddots & \vdots \\ \vdots & h_m(1) & \ddots & 0 \\ h_m(L_h-1) & \vdots & \ddots & h_m(0) \\ 0 & h_m(L_h-1) & \ddots & h_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & h_m(L_h-1) \end{bmatrix}. \quad (7)$$

Defining the L_w–dimensional m-th reverberant signal vector $\mathbf{x}_m(n)$ and the L_c–dimensional clean speech vector $\mathbf{s}(n)$ as

$$\mathbf{x}_m(n) = [x_m(n)\ x_m(n-1)\ \ldots\ x_m(n-L_w+1)]^T, \quad (8)$$

$$\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-L_c+1)]^T, \quad (9)$$

with $\mathbf{x}_m(n) = \mathbf{H}_m^T\mathbf{s}(n)$, the output signal $z(n)$ in (3) can now be expressed as

$$z(n) = \sum_{m=1}^{M} \mathbf{w}_m^T\mathbf{x}_m(n) = \sum_{m=1}^{M} \mathbf{w}_m^T\mathbf{H}_m^T\mathbf{s}(n) = \underbrace{\mathbf{w}^T\mathbf{H}^T}_{\mathbf{c}^T}\mathbf{s}(n). \quad (10)$$

Alternatively, (10) can also be expressed as

$$z(n) = \sum_{m=1}^{M} \mathbf{x}_m^T(n)\mathbf{w}_m = \mathbf{x}^T(n)\mathbf{w}, \quad (11)$$

with $\mathbf{x}(n) = [\mathbf{x}_1^T(n)\ \mathbf{x}_2^T(n)\ \ldots\ \mathbf{x}_M^T(n)]^T$ the ML_w–dimensional stacked reverberant signal vector. Based on (11), the output signal vector $\mathbf{z}(n)$ of length L_z, i.e.,

$$\mathbf{z}(n) = [z(n)\ z(n-1)\ \ldots\ z(n-L_z+1)]^T, \quad (12)$$

can be written as

$$\mathbf{z}(n) = \mathbf{X}(n)\mathbf{w}, \quad (13)$$

with $\mathbf{X}(n)$ the $L_z \times ML_w$–dimensional multi-channel convolution matrix of the reverberant signals, i.e., $\mathbf{X}(n) = [\mathbf{X}_1(n)\ \mathbf{X}_2(n)\ \ldots\ \mathbf{X}_M(n)]$, and

$$\mathbf{X}_m(n) = \begin{bmatrix} x_m(n) & \cdots & x_m(n-L_w+1) \\ x_m(n-1) & \cdots & x_m(n-L_w) \\ \vdots & \ddots & \vdots \\ x_m(n-L_z+1) & \cdots & x_m(n-L_w-L_z+2) \end{bmatrix}. \quad (14)$$

For conciseness, the time index n will be omitted when possible in the remainder of this paper.

III. ACOUSTIC MULTI-CHANNEL EQUALIZATION TECHNIQUES

Acoustic multi-channel equalization techniques aim at speech dereverberation by designing a filter $\mathbf{w}$ such that the EIR $\mathbf{c}$ in (6) is equal to a dereverberated target EIR $\mathbf{c}_t$. However, it should be realized that in practice only the perturbed RIRs $\hat{h}_m$ are available, i.e., $\hat{h}_m = h_m + e_m$, where e_m represents the unknown RIR perturbations arising due to fluctuations (e.g., temperature or position fluctuations [22])

or due to the sensitivity of BSI and SSI methods to near-common zeros or interfering noise [23]–[25]. Hence, for the filter design, the perturbed convolution matrix $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ is used, where $\mathbf{E}$ represents the convolution matrix of the RIR perturbations.

In this paper we will focus on state-of-the-art least-squares acoustic multi-channel equalization techniques, i.e., MINT [15], RMCLS [18], and PMINT [19], which seek to compute the filter $\mathbf{w}$ as the solution to the system of equations

$$\mathbf{W}\hat{\mathbf{H}}\mathbf{w} = \mathbf{W}\mathbf{c}_t, \quad (15)$$

with $\mathbf{W}$ an $L_c \times L_c$–dimensional diagonal weighting matrix. For the considered equalization techniques, the definition of the weighting matrix $\mathbf{W}$ and target EIR $\mathbf{c}_t$ is presented in Tables I and II, where $\mathbf{I}_c$ denotes the $L_c \times L_c$–dimensional identity matrix, τ denotes a delay, L_d denotes the length of the direct path and early reflections, and $p \in \{1,\ 2,\ \ldots,\ M\}$. The delay τ is incorporated in all techniques in order to relax the causality constraints on the filter design [26]. The length of the direct path and early reflections L_d in the RMCLS and PMINT techniques is typically considered to be between 10-50 ms [18], [19]. Based on these definitions of $\mathbf{W}$ and $\mathbf{c}_t$ it can be said that on the one hand, the MINT and PMINT techniques constrain all taps of the EIR (since $\mathbf{W} = \mathbf{I}_c$), which results in a good perceptual speech quality but high sensitivity to RIR perturbations [19], [26], while on the other hand, the RMCLS technique does not constrain all taps of the EIR (since $\mathbf{W} \neq \mathbf{I}_c$), which results in a lower sensitivity to RIR perturbations but decreased perceptual speech quality [18], [19].

For all considered equalization techniques, the filter solving (15) is computed by minimizing the least-squares cost function

$$J_{\text{LS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2. \quad (16)$$

Assuming that the RIRs do not share any common zeros [15] and using $L_w \geq \left\lceil \frac{L_h-1}{M-1} \right\rceil$ [39], the least-squares filter solv-

TABLE I: Definition of the weighting matrix $\mathbf{W}$ for state-of-the-art least-squares equalization techniques.

Technique	Weighting matrix $\mathbf{W}$
MINT	$\mathbf{I}_c$
RMCLS	$\text{diag}\{[\underbrace{1\ \ldots\ 1}_{\tau}\ \underbrace{1\ 0\ \ldots\ 0}_{L_d}\ 1\ \ldots 1]^T\}$
PMINT	$\mathbf{I}_c$

TABLE II: Definition of the target EIR $\mathbf{c}_t$ for state-of-the-art least-squares equalization techniques.

Technique	Target EIR $\mathbf{c}_t$
MINT	$[\underbrace{0\ \ldots\ 0}_{\tau}\ 1\ 0\ \ldots\ 0]^T$
RMCLS	$[\underbrace{0\ \ldots\ 0}_{\tau}\ 1\ 0\ \ldots\ 0]^T$
PMINT	$[\underbrace{0\ldots0}_{\tau}\ \underbrace{\hat{h}_p(0)\ldots\hat{h}_p(L_d-1)}_{L_d}\ 0\ldots0]^T$

ing (15) and minimizing (16) to 0 is equal to

$$\mathbf{w}_{\text{LS}} = (\mathbf{W}\hat{\mathbf{H}})^+(\mathbf{W}\mathbf{c}_t), \quad (17)$$

where $\{\cdot\}^+$ denotes the matrix pseudo-inverse. When the true RIRs are available, i.e., $\hat{\mathbf{H}} = \mathbf{H}$, the least-squares filter yields perfect dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{w}_{\text{LS}} = \mathbf{W}\mathbf{c}_t$ [19], [21]. However, in the presence of RIR perturbations, i.e., $\hat{\mathbf{H}} \neq \mathbf{H}$, the least-squares filter typically fails to achieve dereverberation, i.e., $\mathbf{W}\mathbf{H}\mathbf{w}_{\text{LS}} \neq \mathbf{W}\mathbf{c}_t$, possibly causing severe distortions in the output signal [19].

IV. REGULARIZED ACOUSTIC MULTI-CHANNEL EQUALIZATION TECHNIQUES

In order to increase the robustness of acoustic multi-channel equalization techniques against RIR perturbations, it has been proposed to incorporate regularization in the filter design, such that the energy of distortions due to RIR perturbations is reduced [19], [26]. Regularized equalization techniques design filters by minimizing the regularized least-squares cost function

$$\boxed{J_{\text{RLS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2 + \delta\mathbf{w}^T\mathbf{R}_e\mathbf{w},} \quad (18)$$

where the matrix $\mathbf{R}_e$ models $\mathcal{E}\{\mathbf{E}^T\mathbf{E}\}$, with $\mathcal{E}$ the expected value operator, and δ is a regularization parameter providing a trade-off between the least-squares error and the distortion energy due to RIR perturbations. When knowledge is available about the type of RIR perturbations (e.g., arising due to microphone position fluctuations or arising from BSI or SSI methods), the matrix $\mathbf{R}_e$ can be constructed based on an appropriate perturbation model [17], [40]. When such knowledge is not available, the RIR perturbations are often assumed to be spatially and temporally white, i.e., $\mathbf{R}_e = \mathbf{I}_w$, with $\mathbf{I}_w$ denoting the $ML_w \times ML_w$–dimensional identity matrix [19], [26]. This assumption has been used for the regularized techniques in Section VI-E.

Minimizing (18) yields the regularized least-squares filter

$$\mathbf{w}_{\text{RLS}} = [(\mathbf{W}\hat{\mathbf{H}})^T(\mathbf{W}\hat{\mathbf{H}}) + \delta\mathbf{R}_e]^{-1}(\mathbf{W}\hat{\mathbf{H}})^T(\mathbf{W}\mathbf{c}_t). \quad (19)$$

While the least-squares filter in (17) typically fails to achieve dereverberation in the presence of RIR perturbations, it has been experimentally validated in [19] that the regularized least-squares filter in (19) results in a significantly better performance. However, as can be seen from (18), the regularization term is signal-independent, since it only depends on an RIR perturbation model and does not incorporate any knowledge about the output signal.

V. SPARSITY-PROMOTING ACOUSTIC MULTI-CHANNEL EQUALIZATION TECHNIQUES

Instead of using a signal-independent regularization term, in this section we propose to increase the robustness of equalization techniques by incorporating a signal-dependent penalty function in the filter design, enforcing the output signal $z(n)$ to exhibit spectro-temporal characteristics of a clean speech signal. While in principle any penalty function imposing a well-defined characteristic of clean speech could be used, we propose to use penalty functions which promote sparsity of the output signal in the STFT domain. The advantage of promoting sparsity of the STFT coefficients of the output signal is expected to be twofold. First, it is widely accepted that clean speech is sparse in the STFT domain [32], [41], [42]. Empirical observations, e.g., in [32], have shown that when clean speech is corrupted by reverberation, the STFT coefficients of the mixture are less sparse than the STFT coefficients of clean speech (cf. Figs. 5(a) and (b)). Hence, promoting sparsity of the STFT coefficients of the output signal can yield a signal which better resembles a clean speech signal. Second, in the presence of RIR perturbations, non-robust equalization techniques introduce distortions (i.e., non-zero STFT coefficients) in the output signal [19] (cf. Fig. 5(c)). By promoting sparsity of the STFT coefficients of the output signal, it is expected that these distortions are suppressed. In the remainder of this section, the proposed signal-dependent least-squares cost function is formulated, several possible sparsity-promoting penalty functions are discussed, and the iterative method for optimizing this cost function is presented.

It should be noted that sparsity-promoting techniques have also been proposed to increase the robustness of BSI methods to additive noise [43], [44]. However, unlike the sparsity-promoting equalization techniques proposed in the following which aim at enforcing sparsity on the output speech signal, sparsity-promoting techniques for BSI aim at enforcing sparsity on the estimated RIRs.

A. Sparsity-promoting cost function

In order to incorporate the STFT of the output signal in the filter design, the STFT coefficients $Z(k,l)$ of the output signal $z(n)$ are computed as

$$Z(k,l) = \sum_{n=0}^{N-1} w_{\text{STFT}}(n)z(lR+n)e^{-\frac{j2\pi kn}{N}}, \quad (20)$$

where $w_{\text{STFT}}(n)$ denotes the STFT analysis window of length N, $k = 0,\ 1,\ \ldots,\ N-1$, denotes the frequency bin index, $l = 0,\ 1,\ \ldots,\ L-1$, denotes the frame index with L the total number of frames, and R denotes the frame shift. Based on (20), we define the STFT operator $\boldsymbol{\Psi} \in \mathcal{C}^{NL \times L_z}$ similarly as in [33], [45], transforming the L_z–dimensional time domain vector $\mathbf{z}$ in (12) into the NL–dimensional time-frequency domain vector $\tilde{\mathbf{z}}$, i.e.,

$$\tilde{\mathbf{z}} = \boldsymbol{\Psi}\mathbf{z}. \quad (21)$$

In other words, the vector $\tilde{\mathbf{z}}$ consists of all STFT coefficients $Z(k,l)$, $k = 0,\ 1,\ \ldots,\ N-1$, $l = 0,\ 1,\ \ldots,\ L-1$. Using (13), the time-frequency domain vector $\tilde{\mathbf{z}}$ can be expressed in terms of the filter $\mathbf{w}$ as

$$\tilde{\mathbf{z}} = \boldsymbol{\Psi}\mathbf{X}\mathbf{w}. \quad (22)$$

It should be noted that for a tight analysis window $w_{\text{STFT}}(n)$, i.e., the same window can be used as a synthesis window such that the perfect overlap-add constraint is satisfied, the inverse short-time Fourier transform operator $\boldsymbol{\Psi}^H \in \mathcal{C}^{L_z \times NL}$ satisfies $\boldsymbol{\Psi}^H\boldsymbol{\Psi} = \mathbf{I}_z$, with $\mathbf{I}_z$ the $L_z \times L_z$–dimensional identity matrix.

Aiming at promoting sparsity of the STFT coefficients of the output signal in the filter design, we now propose to add a signal-dependent term to the least-squares cost function, i.e.,

$$J_{\text{SLS}} = J_{\text{LS}} + \eta f_{\text{SP}}(\tilde{\mathbf{z}}) \tag{23}$$
$$= \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2 + \eta f_{\text{SP}}(\mathbf{\Psi}\mathbf{X}\mathbf{w}), \tag{24}$$

with f_{SP} a sparsity-promoting penalty function (cf. Section V-B) and η a weighting parameter providing a trade-off between the minimization of the least-squares error and the value of the penalty function.

B. Sparsity-promoting penalty functions

In this paper we will consider three commonly used sparsity-promoting norms for the penalty function f_{SP} in (24), i.e., the l_0-norm[1], the l_1-norm, and the weighted l_1-norm [46]–[49].

The l_0-norm $f_{\text{SP}}^0(\tilde{\mathbf{z}})$ counts the number of non-zero entries in $\tilde{\mathbf{z}}$, i.e.,

$$f_{\text{SP}}^0(\tilde{\mathbf{z}}) = \|\tilde{\mathbf{z}}\|_0 = |q : \tilde{z}(q) \neq 0|. \tag{25}$$

Using $f_{\text{SP}}^0(\tilde{\mathbf{z}})$ in (24) penalizes the non-zero STFT coefficients of the output signal, which can be advantageous to suppress the reverberant energy as well as the distortions introduced by non-robust least-squares equalization techniques. However, the l_0-norm is non-convex, and it is well known that optimization problems with non-convex penalty functions are typically hard (if not impossible) to solve, particularly for large scale problems [46]. In addition, iterative methods proposed to solve such optimization problems are not guaranteed to converge to the global minimum, but only to a local minimum [50].

A commonly used alternative to the l_0-norm is the l_1-norm penalty function $f_{\text{SP}}^1(\tilde{\mathbf{z}})$, defined as

$$f_{\text{SP}}^1(\tilde{\mathbf{z}}) = \|\tilde{\mathbf{z}}\|_1 = \sum_{q=0}^{NL-1} |\tilde{z}(q)|. \tag{26}$$

The l_1-norm can be viewed as a convex relaxation of the l_0-norm, and efficient methods have been proposed to solve optimization problems with l_1-norm penalty functions [47], [48]. As can be seen in (26), the l_1-norm is magnitude-dependent, differing from the l_0-norm by penalizing the coefficients of $\tilde{\mathbf{z}}$ with larger magnitude more than the coefficients with smaller magnitude. This penalization is not necessarily desirable, since it does not guarantee the preservation of the spectro-temporal structure of a typical speech signal.

To counteract the magnitude dependency of the l_1-norm, the weighted l_1-norm penalty function $f_{\text{SP}}^{w,1}(\tilde{\mathbf{z}})$ has been proposed [49], defined as

$$f_{\text{SP}}^{w,1}(\tilde{\mathbf{z}}) = \|\text{diag}\{\mathbf{u}\}\tilde{\mathbf{z}}\|_1 = \sum_{q=0}^{NL-1} |u(q)\tilde{z}(q)|, \tag{27}$$

with $\mathbf{u}$ an NL–dimensional vector of positive scalar weights, i.e., $u(q) > 0$, $q = 0, 1, \ldots, NL - 1$, selectively penalizing the coefficients of $\tilde{\mathbf{z}}$. To promote the same sparsity structure in

---

the output signal as in the desired signal, it has been proposed in [49] to select the weighting vector such that it has small values when the desired signal has non-zero magnitude and significantly larger values otherwise. In the context of speech dereverberation, it would hence be desirable to select the weights to be inversely proportional to the magnitude of the STFT coefficients of the clean speech signal, i.e.,

$$u(q) = \frac{1}{|\tilde{s}(q)|}, \quad q = 0, 1, \ldots, NL - 1, \tag{28}$$

with $\tilde{s}(q)$ the STFT coefficients of the clean speech signal computed similarly as in (21). Using the weights in (28) results in a larger penalization - and hence, a larger suppression - of the STFT coefficients of the output signal corresponding to the STFT coefficients of the clean speech signal with a small magnitude. However, since the clean speech signal is obviously not available, we propose to use one of the reverberant microphone signals and compute the weights as

$$u(q) = \frac{1}{|\tilde{x}_p(q)| + \zeta}, \quad q = 0, 1, \ldots, NL - 1, \tag{29}$$

with $\tilde{x}_p(q)$ the STFT coefficients of the p-th microphone signal computed similarly as in (21) and $\zeta > 0$ a small positive scalar included to avoid division by 0 and to ensure stability of the weighted l_1-norm sparsity-promoting techniques. When $|\tilde{x}_p(q)| \approx 0$, e.g., during speech absence, using ζ ensures that $u(q)$ is large but finite, such that the magnitude of the STFT coefficients of the reconstructed signal is also (nearly) 0. As will be experimentally shown in Section VI-B, incorporating the weighted l_1-norm penalty function using the weights in (29) typically yields a better dereverberation performance than incorporating the l_0- or l_1-norm penalty functions. The advantage of using the weighted l_1-norm penalty function lies in the fact that appropriate weights as in (29) ensure that the spectro-temporal structure of a typical speech signal is preserved.

C. Iterative optimization method

Since no closed-form expression is available for the filter minimizing the sparsity-promoting cost function in (24), iterative optimization methods are required. We propose to use the ADMM [37], since it is a well-suited method for solving large scale optimization problems of the form in (24). Within the ADMM framework, the minimization of the sparsity-promoting cost function in (24) is reformulated as

$$\min_{\mathbf{w}} \left[\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2 + \eta f_{\text{SP}}(\tilde{\mathbf{a}}) \right] \quad \text{subject to} \quad \mathbf{\Psi}\mathbf{X}\mathbf{w} = \tilde{\mathbf{a}}, \tag{30}$$

where the NL–dimensional auxiliary variable $\tilde{\mathbf{a}}$ is introduced such that the optimization problem in (24) can be split into simpler sub-problems. The augmented Lagrangian of (30) is equal to

$$\mathcal{L} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2 + \eta f_{\text{SP}}(\tilde{\mathbf{a}}) + \frac{\rho}{2}\|\mathbf{\Psi}\mathbf{X}\mathbf{w} + \boldsymbol{\lambda} - \tilde{\mathbf{a}}\|_2^2, \tag{31}$$

with $\boldsymbol{\lambda}$ the NL–dimensional dual (splitting) variable and $\rho > 0$ the ADMM penalty parameter. ADMM minimizes (31) alternately with respect to the variables $\mathbf{w}$ and $\tilde{\mathbf{a}}$, followed by

---

[1]Note that the l_0-norm is not a norm in the strict mathematical sense, since it does not satisfy all the norm properties.

a dual ascent over the variable $\boldsymbol{\lambda}$ [37]. The ADMM update rules are given by

$$\mathbf{w}^{(i+1)} = \arg\min_{\mathbf{w}} \Big[\|\mathbf{W}(\hat{\mathbf{H}}\mathbf{w} - \mathbf{c}_t)\|_2^2$$
$$+ \frac{\rho}{2}\|\boldsymbol{\Psi}\mathbf{X}\mathbf{w} + \boldsymbol{\lambda}^{(i)} - \tilde{\mathbf{a}}^{(i)}\|_2^2 \Big], \quad (32)$$

$$\tilde{\mathbf{a}}^{(i+1)} = \arg\min_{\tilde{\mathbf{a}}} \Big[\eta f_{\mathrm{SP}}(\tilde{\mathbf{a}}) + \frac{\rho}{2}\|\boldsymbol{\Psi}\mathbf{X}\mathbf{w}^{(i+1)} + \boldsymbol{\lambda}^{(i)} - \tilde{\mathbf{a}}\|_2^2 \Big], \quad (33)$$

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} + \boldsymbol{\Psi}\mathbf{X}\mathbf{w}^{(i+1)} - \tilde{\mathbf{a}}^{(i+1)}, \quad (34)$$

where $\{\cdot\}^{(i)}$ denotes the variable in the i-th iteration. The minimization problem in (24) is hence decomposed into simpler sub-problems, which are solved in an alternating fashion using the update rules in (32), (33), and (34) until a termination criterion is satisfied (cf. Section VI-A). In the following, the update rules for the filter in (32) and for the auxiliary variable in (33) are presented.

Filter update rule: In order to derive the filter update rule, the gradient of (32) is set to $\mathbf{0}$, yielding

$$\mathbf{w}^{(i+1)} = \underbrace{[2(\mathbf{W}\hat{\mathbf{H}})^T(\mathbf{W}\hat{\mathbf{H}}) + \rho\mathbf{X}^T\mathbf{X}]^{-1}}_{\mathbf{C}}$$
$$\times \underbrace{[2(\mathbf{W}\hat{\mathbf{H}})^T(\mathbf{W}\mathbf{c}_t)}_{\mathbf{b}_1} + \rho\underbrace{\mathbf{X}^T\boldsymbol{\Psi}^H(\tilde{\mathbf{a}}^{(i)} - \boldsymbol{\lambda}^{(i)})}_{\mathbf{b}_2^{(i)}}] \quad (35)$$

$$= \mathbf{C}^{-1}(\mathbf{b}_1 + \rho\mathbf{b}_2^{(i)}), \quad (36)$$

where the variables $\mathbf{C}$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are introduced to highlight the fact that only the variable $\mathbf{b}_2$ is iteration-dependent, whereas the variables $\mathbf{C}$ and $\mathbf{b}_1$ can be pre-computed. Although (36) appears to require a matrix inversion in each iteration of the ADMM, the filter update can be efficiently computed by, e.g., storing the LU-factorization of $\mathbf{C}$ and using forward and backward substitution [51].

Auxiliary variable update rule: The auxiliary variable update rule in (33) is given by the proximal mapping of the sparsity-promoting penalty function [48], [52]. The proximal mapping for the l_0-, l_1-, and weighted l_1-norm penalty functions exists in closed-form [48], [52], enabling to efficiently compute the auxiliary variable update rule in each iteration of the ADMM. Simplifying the notation by defining the operator $(T)_+ = \max(T, 0)$ and the variable $\mathbf{b}$ as

$$\mathbf{b}^{(i)} = \boldsymbol{\Psi}\mathbf{X}\mathbf{w}^{(i+1)} + \boldsymbol{\lambda}^{(i)}, \quad (37)$$

the proximal mapping for the penalty functions f_{SP}^0, f_{SP}^1, and $f_{\mathrm{SP}}^{w,1}$ is given by the following element-wise mappings

$$\tilde{a}^{(i+1)}(q) = \begin{cases} \left(\big|b^{(i)}(q)\big| - \frac{\eta}{\rho}\right)_+ \frac{b^{(i)}(q)}{\big|b^{(i)}(q)\big| - \frac{\eta}{\rho}} & \text{if } f_{\mathrm{SP}}^0 \\[2ex] \left(\big|b^{(i)}(q)\big| - \frac{\eta}{\rho}\right)_+ \frac{b^{(i)}(q)}{\big|b^{(i)}(q)\big|} & \text{if } f_{\mathrm{SP}}^1 \\[2ex] \left(\big|b^{(i)}(q)\big| - \frac{\eta}{\rho}u(q)\right)_+ \frac{b^{(i)}(q)}{\big|b^{(i)}(q)\big|} & \text{if } f_{\mathrm{SP}}^{w,1} \end{cases} \quad (38)$$

As can be observed in (38), the mapping for the l_0-norm penalty function reduces the l_0-norm of the auxiliary variable $\tilde{\mathbf{a}}$ in each iteration of the ADMM by changing all but the largest elements (with magnitude larger than $\frac{\eta}{\rho}$) to 0. The mapping for the l_1-norm penalty function reduces the l_1-norm of $\tilde{\mathbf{a}}$ in each iteration by subtracting $\frac{\eta}{\rho}$ from the magnitude of every element of $\tilde{\mathbf{a}}$. Finally, the mapping for the weighted l_1-norm penalty function is similar to the mapping for the l_1-norm, with the only difference consisting in the multiplication of the threshold $\frac{\eta}{\rho}$ with the weights $u(q)$.

In summary, using the filter update rule in (35), the auxiliary variable update rule in (38) depending on the used sparsity-promoting penalty function, and the dual variable update rule in (34), the sparsity-promoting least-squares filters can be iteratively computed until a termination criterion is satisfied (cf. Section VI-A).

VI. Simulation Results

In this section the performance of the proposed sparsity-promoting acoustic multi-channel equalization techniques is investigated by means of instrumental measures. In Section VI-A the considered acoustic systems, algorithmic settings, and instrumental measures are introduced. In Section VI-B the robustness against RIR perturbations for the MINT, RMCLS, and PMINT techniques when incorporating different sparsity-promoting penalty functions is investigated. In Section VI-C the performance of the weighted l_1-norm sparsity-promoting techniques is compared for all considered acoustic systems and RIR perturbation levels. In Section VI-D the performance of the weighted l_1-norm sparsity-promoting PMINT technique for different choices of the weighting and penalty parameters η and ρ is investigated. Finally, in Section VI-E the performance of the signal-dependent weighted l_1-norm sparsity-promoting PMINT technique is compared to the performance of the signal-independent regularized PMINT technique. Exemplary sound samples from each simulation can be found at http://bit.ly/2cPDnsN.

A. Acoustic systems, algorithmic settings, and instrumental measures

Acoustic systems

We have considered three different reverberant acoustic systems with a single speech source and $M = 4$ omni-directional microphones. The RIRs between the speech source and the microphones are measured using the swept-sine technique [53] and the reverberant signals are generated by convolving 10 sentences (approximately 17 s long) from the HINT database [54] with the measured RIRs. For each acoustic system, Table III presents the reverberation time T_{60} of the room, the source-microphone distance d_{sm}, the inter-microphone distance d_{im}, and the length of the RIRs L_h at a sampling frequency $f_s = 8$ kHz.

In order to simulate RIR perturbations, the measured RIRs are perturbed by proportional Gaussian distributed errors as proposed in [55], such that a desired level of normalized projection misalignment (NPM), i.e.,

$$\mathrm{NPM} = 20\log_{10}\frac{\left\|\mathbf{h} - \frac{\mathbf{h}^T\hat{\mathbf{h}}}{\hat{\mathbf{h}}^T\hat{\mathbf{h}}}\hat{\mathbf{h}}\right\|_2}{\|\mathbf{h}\|_2}, \quad (39)$$

TABLE III: Characteristics of the considered measured acoustic systems.

Acoustic system	T_{60} [ms]	d_{sm} [m]	d_{im} [m]	L_h
S_1	610	2	0.04	5000
S_2	450	3	0.05	3800
S_3	360	2	0.04	3000

is generated. Introducing proportional Gaussian distributed errors is a widely used technique to systematically simulate RIR perturbations. The considered NPMs for each acoustic system are

$$\text{NPM} \in \{-33 \text{ dB}, \ -27 \text{ dB}, \ \ldots, \ -3 \text{ dB}\}, \quad (40)$$

with -33 dB a moderate perturbation level and -3 dB a rather large perturbation level. The reported NPM values achieved by state-of-the-art BSI methods (for relatively short RIRs) range between -10 dB and -20 dB [25], hence, the values in (40) can be considered as rather realistic.

Algorithmic settings

For all considered equalization techniques the filter length is set to $L_w = \left\lceil \frac{L_h - 1}{M - 1} \right\rceil$, i.e., $L_w = 1667$ for S_1, $L_w = 1267$ for S_2, and $L_w = 1000$ for S_3. The length of the direct path and early reflections is set to $L_d = 0.01 \times f_s$, i.e., 10 ms, and the delay is set to $\tau = 90$ (cf. Tables I and II). The target EIR $\mathbf{c}_t$ for the PMINT technique is constructed using the first RIR, i.e., $p = 1$ (cf. Table II).

For the sparsity-promoting filters the STFT is computed using a 32 ms Hamming window (i.e., $N = 256$) with 50% overlap between successive frames. In order to reduce the computational complexity, the sparsity-promoting filters are computed using only the first two sentences of the reverberant signals (i.e., $L = 208$), but the complete output signal has been used for the evaluation. For the weighted l_1-norm penalty function, the weights in (29) are computed using the first microphone signal, i.e., $p = 1$, and $\zeta = 10^{-8}$. The termination criterion for the ADMM is a maximum number of iterations $i_{\max} = 500$. Since the initialization of the ADMM may influence the resulting sparsity-promoting filter for the non-convex l_0-norm penalty function, we have investigated three different initializations of the filter $\mathbf{w}$:

i) $\mathbf{w}^{(0)} = [1 \ 0 \ \ldots \ 0]^T$, i.e., the filter yielding the first microphone signal,
ii) $\mathbf{w}^{(0)}$ is randomly initialized with normally distributed coefficients,
iii) $\mathbf{w}^{(0)} = \mathbf{w}_{\text{LS}}$, i.e., the filter resulting from the least-squares techniques.

In all simulations we observed that using the filter initialization i), i.e., $\mathbf{w}^{(0)} = [1 \ 0 \ \ldots \ 0]^T$, resulted in the best performance, and hence, the following simulation results are generated using this filter initialization.

Finally, we have considered several regularization parameters δ for the regularized techniques and several weighting and penalty parameters η and ρ for the sparsity-promoting techniques, i.e.,

$$\delta, \eta, \rho \in \{10^{-7}, \ 10^{-6}, \ \ldots, \ 10^{-1}\}. \quad (41)$$

The method for selecting the parameter values δ, η, and ρ for the simulation results presented in the following is described at the beginning of each section.

Performance measures

The performance of the equalization techniques is evaluated in terms of the reverberant energy suppression and perceptual speech quality improvement. The reverberant energy suppression is evaluated using the improvement in *direct-to-reverberant ratio* (ΔDRR) [6], i.e., $\Delta\text{DRR} = \text{oDRR} - \text{iDRR}$, with

$$\text{oDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} c^2(n)}{\sum_{n=n_d}^{L_c-1} c^2(n)}, \quad \text{iDRR} = 10 \log_{10} \frac{\sum_{n=0}^{n_d-1} h_1^2(n)}{\sum_{n=n_d}^{L_h-1} h_1^2(n)}, \quad (42)$$

where the first $n_d = 0.008 \times f_s$ samples of the EIR $c(n)$ and of the RIR $h_1(n)$ represent the direct path propagation and the remaining samples represent reflections. Although the ΔDRR exactly describes the reverberant energy suppression, it cannot be solely used to evaluate the dereverberation performance of equalization techniques, since it does not provide any insights on the reverberant energy decay rate. To evaluate the reverberant energy decay rate, the *energy decay curve* (EDC) [56] of the EIR $c(n)$ is compared to the energy decay curve of the true RIR $h_1(n)$. The EDC of the EIR is computed as

$$\text{EDC}_{\mathbf{c}}(n) = \frac{1}{\|\mathbf{c}\|_2^2} \sum_{j=n}^{L_c-1} c^2(j), \quad n = 0, \ 1, \ \ldots, \ L_c - 1, \quad (43)$$

and the EDC of the RIR is computed as

$$\text{EDC}_{\mathbf{h}_1}(n) = \frac{1}{\|\mathbf{h}_1\|_2^2} \sum_{j=n}^{L_h-1} h_1^2(j), \quad n = 0, \ 1, \ \ldots, \ L_h - 1. \quad (44)$$

The perceptual speech quality is evaluated using the *perceptual evaluation of speech quality* (PESQ) measure [57] and the *cepstral distance* (CD) [58]. These signal-based measures are intrusive measures, generating a similarity score between a test signal and a reference signal. The reference signal employed here is the direct path and early reflections speech signal in the first microphone, i.e., the signal obtained by convolving the clean speech signal with the direct path and early reflections (considered to be 10 ms long) of $\mathbf{h}_1$. The improvement in PESQ, i.e., ΔPESQ, is computed as the difference between the PESQ score of the output signal $z(n)$ and the PESQ score of the first microphone signal $x_1(n)$. Similarly, the improvement in CD, i.e., ΔCD, is computed as the difference between the CD of the output signal $z(n)$ and the CD of the first microphone signal $x_1(n)$. Note that a positive ΔPESQ and a negative ΔCD indicate an improvement in perceptual speech quality.

B. Sparsity-promoting acoustic multi-channel equalization techniques

In this section the performance of the considered least-squares equalization techniques, i.e., MINT, RMCLS, and
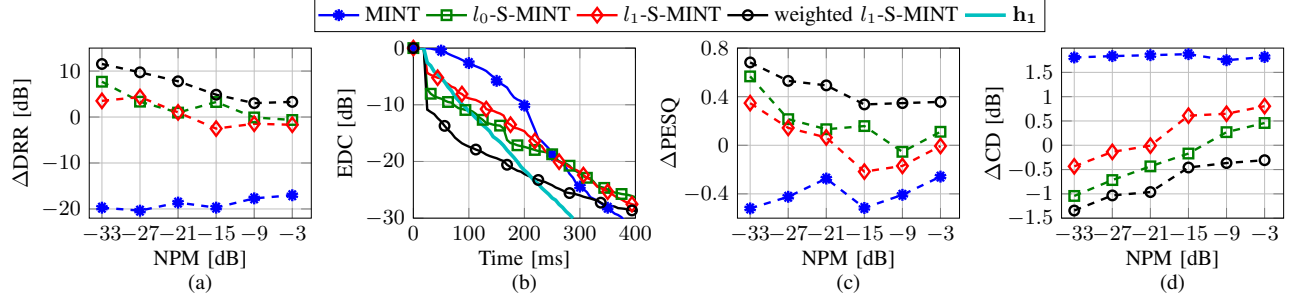
Fig. 2: Performance of the MINT technique and sparsity-promoting MINT technique using different penalty functions for acoustic system S_1 in terms of (a) ΔDRR, (b) EDC for an exemplary NPM of -33 dB, (c) ΔPESQ, and (d) ΔCD (intrusively selected weighting and penalty parameters η and ρ).
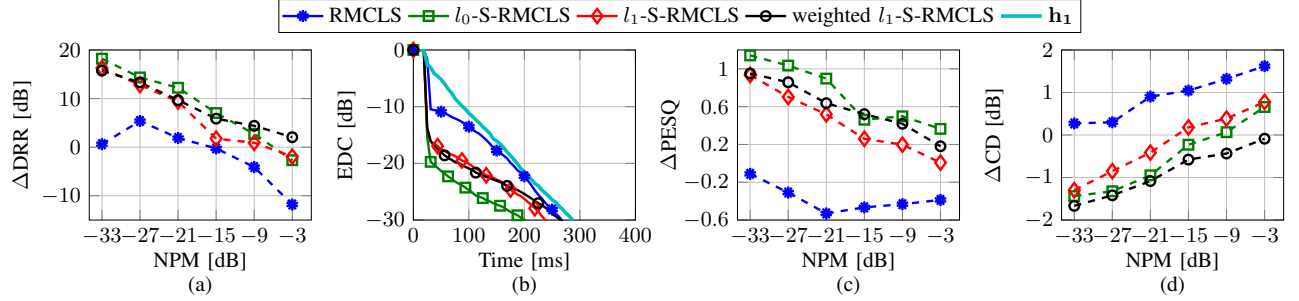

Fig. 3: Performance of the RMCLS technique and sparsity-promoting RMCLS technique using different penalty functions for acoustic system S_1 in terms of (a) ΔDRR, (b) EDC for an exemplary NPM of -33 dB, (c) ΔPESQ, and (d) ΔCD (intrusively selected weighting and penalty parameters η and ρ).

PMINT, is investigated when incorporating the sparsity-promoting penalty functions presented in Section V-B. Although similar results are obtained for all considered acoustic systems, in this section only results for acoustic system S_1 are presented. In order to illustrate the full potential of incorporating a sparsity-promoting penalty function, the weighting and penalty parameters η and ρ are selected from the set in (41) as the parameters minimizing the CD for each equalization technique, each penalty function, and each NPM (i.e., possibly different parameters are used for each equalization technique, each penalty function, and each NPM). It should be noted that selecting the weighting and penalty parameters based on the CD is an intrusive procedure that cannot be applied in practice, since knowledge of the true RIRs is required in order to compute the reference signal and the resulting EIR.

Sparsity-promoting MINT

Fig. 2 depicts the performance of the MINT technique and sparsity-promoting MINT technique using different penalty functions in terms of ΔDRR, EDC, ΔPESQ, and ΔCD. As expected, the ΔDRR values presented in Fig. 2(a) show that MINT fails to suppress the reverberant energy, even worsening the DRR by approximately 20 dB for all NPMs in comparison to $\mathbf{h_1}$. Furthermore, it can be observed that incorporating sparsity-promoting penalty functions significantly increases the robustness of MINT for all considered NPMs. While incorporating the l_0- and l_1-norm penalty functions improves the DRR in comparison to $\mathbf{h_1}$ only for low NPMs, incorporating the weighted l_1-norm penalty function yields an improvement

for all considered NPMs. These results are confirmed in Fig. 2(b), which depicts the EDC of $\mathbf{h_1}$ and the EDCs of the EIRs obtained using MINT and the sparsity-promoting MINT techniques for an exemplary NPM of -33 dB. It can be observed that the MINT technique completely fails to achieve dereverberation and even results in a slower reverberant energy decay rate than $\mathbf{h_1}$. Although the incorporation of the l_0- and l_1-norm penalty functions improves the reverberant energy decay rate in comparison to MINT, these penalty functions still yield a similar energy decay rate as $\mathbf{h_1}$. On the other hand, incorporating the weighted l_1-norm penalty function is more advantageous and results in a faster reverberant energy decay rate than $\mathbf{h_1}$. The larger reverberant energy suppression achieved by the weighted l_1-norm sparsity-promoting MINT technique is also reflected in the larger perceptual speech quality improvement illustrated by the ΔPESQ and ΔCD values in Figs. 2(c) and 2(d), where it can be observed that the weighted l_1-norm sparsity-promoting MINT technique is the only technique yielding an improvement in perceptual speech quality for all considered NPMs.

Sparsity-promoting RMCLS

Fig. 3 depicts the performance of the RMCLS technique and sparsity-promoting RMCLS technique using different penalty functions in terms of ΔDRR, EDC, ΔPESQ, and ΔCD. The ΔDRR values presented in Fig. 3(a) show that for low NPMs the RMCLS technique slightly suppresses the reverberant energy, whereas for NPMs larger than -15 dB the RMCLS technique fails to achieve dereverberation and worsens the
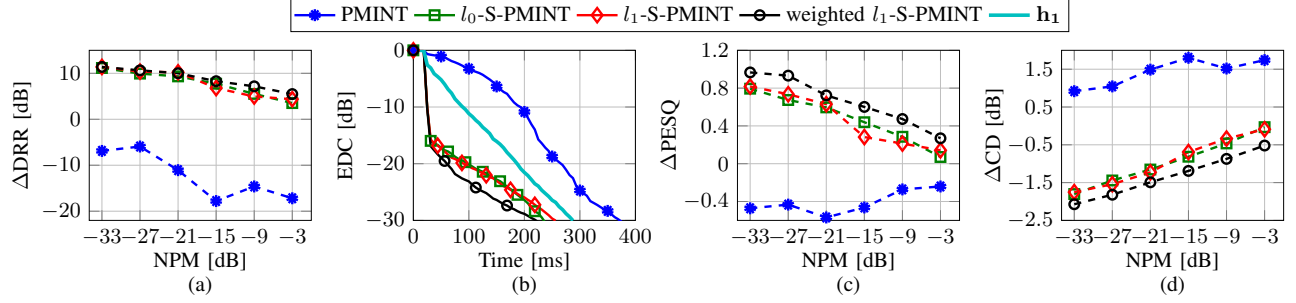
Fig. 4: Performance of the PMINT technique and sparsity-promoting PMINT technique using different penalty functions for acoustic system S_1 in terms of (a) ΔDRR, (b) EDC for an exemplary NPM of -33 dB, (c) ΔPESQ, and (d) ΔCD (intrusively selected weighting and penalty parameters η and ρ).

DRR in comparison to $\mathbf{h}_1$. Furthermore, it can be observed that incorporating sparsity-promoting penalty functions in RMCLS significantly increases the reverberant energy suppression, with all penalty functions yielding a similar ΔDRR for low NPMs. As the NPM increases to -3 dB, the weighted l_1-norm sparsity-promoting RMCLS technique is the only technique achieving an (slight) improvement in DRR in comparison to $\mathbf{h}_1$. This increase in robustness when incorporating sparsity-promoting penalty functions is confirmed in Fig. 3(b), which depicts the EDC of $\mathbf{h}_1$ and the EDCs of the EIRs obtained using RMCLS and the sparsity-promoting RMCLS techniques for an exemplary NPM of -33 dB. It can be observed that while the RMCLS technique achieves a slightly faster reverberant energy decay rate than $\mathbf{h}_1$, the incorporation of all sparsity-promoting penalty functions results in a significantly faster energy decay rate. In addition, it can be observed that for this exemplary NPM, the l_0-norm sparsity-promoting RMCLS technique yields the fastest reverberant energy decay rate, while the energy decay rate for the l_1- and weighted l_1-norm penalty functions is comparable. The larger reverberant energy suppression achieved by the sparsity-promoting RMCLS techniques is also reflected in the larger perceptual speech quality improvement illustrated by the ΔPESQ and ΔCD values in Figs. 3(c) and 3(d), where it can be observed that the l_0-norm sparsity-promoting RMCLS technique typically yields the highest ΔPESQ, whereas the weighted l_1-norm sparsity-promoting RMCLS technique always yields the lowest ΔCD.

Sparsity-promoting PMINT

Fig. 4 depicts the performance of the PMINT technique and sparsity-promoting PMINT technique using different penalty functions in terms of ΔDRR, EDC, ΔPESQ, and ΔCD. As expected, the ΔDRR values presented in Fig. 4(a) show that the PMINT technique fails to suppress the reverberant energy, even worsening the DRR in comparison to $\mathbf{h}_1$. Furthermore, it can be observed that incorporating sparsity-promoting penalty functions significantly increases the robustness of PMINT for all considered NPMs. While all penalty functions result in a similarly large ΔDRR for low NPMs, the weighted l_1-norm penalty function yields a slightly larger ΔDRR than the l_0- and l_1-norm penalty functions for high NPMs. These results are confirmed in Fig. 4(b), which depicts the EDC of $\mathbf{h}_1$ and

the EDCs of the EIRs obtained using PMINT and the sparsity-promoting PMINT techniques for an exemplary NPM of -33 dB. It can be observed that the PMINT technique completely fails to achieve dereverberation and even results in a slower reverberant energy decay rate than $\mathbf{h}_1$. The incorporation of all sparsity-promoting penalty functions results in a significantly faster reverberant energy decay rate, with the weighted l_1-norm penalty function yielding the fastest energy decay rate. Furthermore, the ΔPESQ and ΔCD values depicted in Figs. 4(c) and 4(d) show that while PMINT worsens the perceptual speech quality in comparison to the reverberant signal $x_1(n)$, the incorporation of all sparsity-promoting penalty functions results in a significant improvement, with the weighted l_1-norm penalty function always yielding the highest ΔPESQ and the lowest ΔCD.

In summary, these results show that incorporating sparsity-promoting penalty functions significantly increases the robustness of all considered equalization techniques against RIR perturbations. Furthermore, the weighted l_1-norm penalty function typically outperforms the l_0- and l_1-norm penalty functions in terms of the considered instrumental measures.

To further illustrate the advantage of promoting sparsity of the output signal in the STFT domain, Fig. 5 presents the spectrograms of the direct path and early reflections speech signal at the first microphone, the reverberant first microphone signal, the output signal obtained using the PMINT technique, and the output signals obtained using the sparsity-promoting PMINT techniques using different penalty functions for an exemplary NPM of -33 dB. Comparing Figs. 5(a) and 5(b), it can be observed that due to the spectro-temporal smearing effect of reverberation, the spectrogram of the reverberant signal is significantly less sparse than the spectrogram of the direct path and early reflections speech signal. Furthermore, Fig. 5(c) shows that the non-robust PMINT technique introduces additional distortions, such that the resulting spectrogram is significantly less sparse than the spectrogram of the reverberant signal. On the contrary, Fig. 5(d) shows that incorporating the l_0-norm penalty function leads to a significantly more sparse spectrogram, largely suppressing the distortions introduced by the PMINT technique as well as partly suppressing the reverberant energy (e.g., at $f \approx 1$ kHz). In addition, Fig. 5(e) shows that incorporating the l_1-
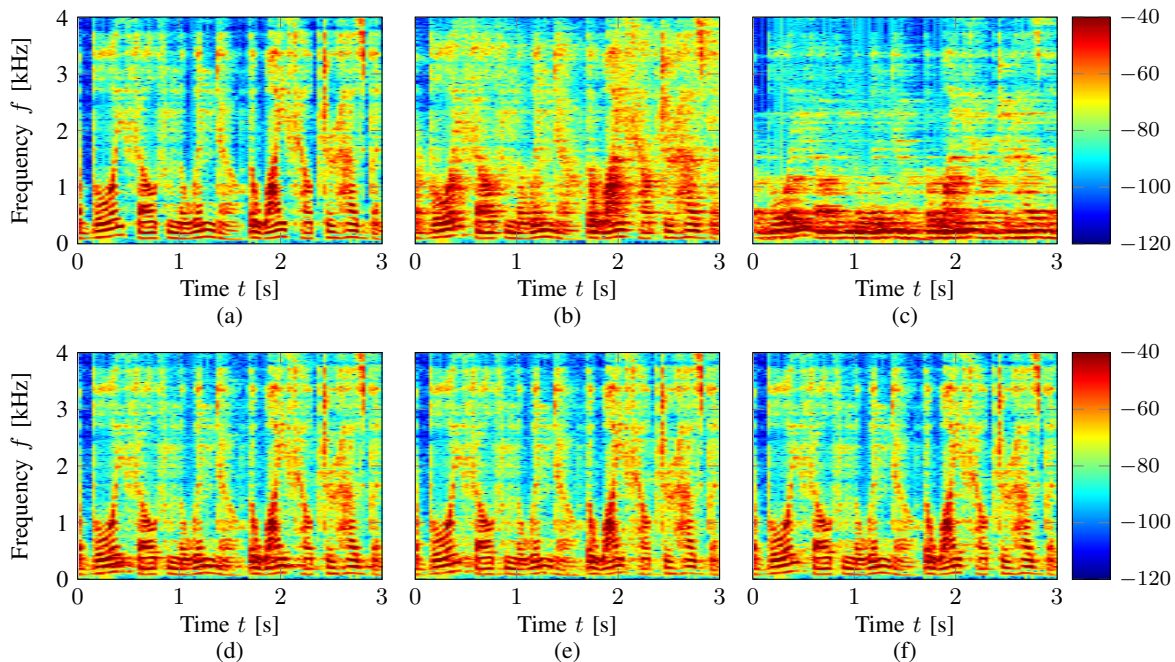
Fig. 5: Spectrogram of (a) direct path and early reflections speech signal at the first microphone, (b) reverberant first microphone signal, (c) output signal obtained using PMINT, (d) output signal obtained using l_0-norm sparsity-promoting PMINT, (e) output signal obtained using l_1-norm sparsity-promoting PMINT, and (f) output signal obtained using weighted l_1-norm sparsity-promoting PMINT (acoustic system S_1, NPM $= -33$ dB, intrusively selected weighting and penalty parameters η and ρ).

TABLE IV: Performance of the weighted l_1-norm sparsity-promoting MINT, RMCLS, and PMINT techniques (intrusively selected weighting and penalty parameters η and ρ).

	ΔDRR [dB]				ΔPESQ				ΔCD [dB]			
	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average
weighted l_1-S-MINT	6.7	6.4	2.8	5.3	0.5	0.4	0.0	0.3	−0.7	−0.8	0.1	−0.5
weighted l_1-S-RMCLS	8.5	**9.3**	**8.3**	**8.7**	0.6	0.6	0.3	0.5	−0.9	−1.1	0.0	−0.7
weighted l_1-S-PMINT	**8.8**	8.7	6.3	7.9	**0.7**	**0.8**	**0.6**	**0.7**	**−1.3**	**−1.6**	**−0.8**	**−1.2**

norm penalty function results in a more sparse spectrogram than the l_0-norm penalty function, further suppressing the reverberant energy (e.g., at $t \approx 2.1$ s). Finally, Fig. 5(f) shows that an even more sparse spectrogram is obtained using the weighted l_1-norm penalty function, with the reverberant energy significantly suppressed (e.g., at $t \approx 0.5$ s and $t \approx 2.1$ s) and the spectro-temporal structure of the resulting signal closely resembling the spectro-temporal structure of the direct and early reflections speech signal in Fig. 5(a).

C. Comparison of the weighted l_1-norm sparsity-promoting acoustic multi-channel equalization techniques

Since it was shown in the previous section that incorporating the weighted l_1-norm penalty function typically yields the best performance, in this section the performance of the weighted l_1-norm sparsity-promoting MINT, RMCLS, and PMINT techniques is compared for all acoustic systems in Table III and all NPMs in (40). The weighting and penalty parameters η and ρ are intrusively determined as in Section VI-B, i.e., as the parameters minimizing CD for each equalization technique, each acoustic system, and each NPM. The performance of

the different techniques is evaluated in terms of ΔDRR, ΔPESQ, and ΔCD, and the presented performance measures are averaged over all considered NPMs.

Table IV presents the obtained ΔDRR, ΔPESQ, and ΔCD values for all considered acoustic systems and techniques. In addition, the average performance measures over all acoustic systems are also presented. First, it can be observed that the weighted l_1-norm sparsity-promoting MINT technique results in the lowest performance in terms of all performance measures. Since the MINT technique is very sensitive to RIR perturbations (cf. Fig. 2), the robustness increase that can be obtained by incorporating signal-dependent penalty functions is also limited. Second, it can be observed that the weighted l_1-norm sparsity-promoting PMINT and RM-CLS techniques yield a high reverberant energy suppression in terms of ΔDRR, with the weighted l_1-norm sparsity-promoting PMINT technique resulting in a similar ΔDRR (for S_1 and S_2) or slightly lower ΔDRR (for S_3) than the weighted l_1-norm sparsity-promoting RMCLS technique. Finally, it can be observed that for all considered acoustic systems, the weighted l_1-norm sparsity-promoting PMINT

technique yields the largest perceptual speech quality improvement, outperforming the weighted l_1-norm sparsity-promoting RMCLS technique in terms of ΔPESQ and ΔCD. While the weighted l_1-norm sparsity-promoting PMINT technique always improves the perceptual speech quality in comparison to the reverberant signal as indicated by positive ΔPESQ and negative ΔCD values, the weighted l_1-norm sparsity-promoting RMCLS technique might fail to yield an improvement, as indicated by ΔCD = 0.0 dB for acoustic system S_3. The advantage of building upon the PMINT technique lies in its control of the early reflections in the EIR, hence better preserving the perceptual speech quality of the output signal.

Summarizing these simulation results, we conclude that the weighted l_1-norm sparsity-promoting PMINT technique is a robust and perceptually advantageous equalization technique, yielding a high reverberant energy suppression and outperforming all other considered techniques in terms of the perceptual speech quality improvement.

D. Performance of the weighted l_1-norm sparsity-promoting PMINT technique for different weighting and penalty parameters

In order to gain more insight on the weighted l_1-norm sparsity-promoting PMINT technique, in this section we analyze its performance for different choices of the weighting and penalty parameters η and ρ. Fig. 6 depicts the ΔCD values obtained using the weighted l_1-norm sparsity-promoting PMINT technique with different values of η and ρ for the exemplary acoustic system S_1 and an NPM of -33 dB. It should be noted that similar results are obtained for all considered performance measures, acoustic systems, and NPMs. It can be observed that for a fixed weighting parameter η (e.g., $\eta = 10^{-5}$) changing the penalty parameter ρ can yield a significant change in the obtained ΔCD (e.g., up to 3 dB). For a fixed penalty parameter ρ (e.g., $\rho = 10^{-3}$) changing the weighting parameter η can also yield a significant change in the obtained ΔCD (e.g., up to 3 dB). Hence unfortunately, the performance of the weighted l_1-norm sparsity-promoting PMINT technique is rather dependent on the choice of the weighting and penalty parameters. However, as it will be shown in Section VI-E, once a set of optimal weighting and penalty parameters has been determined, the same parameters
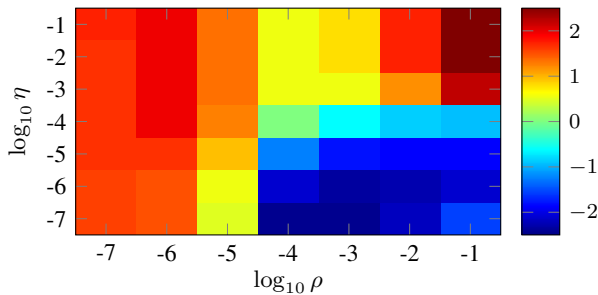

Fig. 6: Performance of the weighted l_1-norm sparsity-promoting PMINT technique in terms of ΔCD for several weighting and penalty parameters η and ρ (acoustic system S_1, NPM = -33 dB).

can be used for different acoustic systems and NPMs to obtain a near-to-optimal performance.

E. Comparison of the weighted l_1-norm sparsity-promoting PMINT and regularized PMINT techniques

In this section, we compare the performance between incorporating the proposed signal-dependent weighted l_1-norm penalty function and the signal-independent regularization term in the PMINT technique for all acoustic systems in Table III and NPMs in (40). In contrast to Sections VI-B and VI-C, for a more realistic evaluation the same weighting and penalty parameters η and ρ are now used in the weighted l_1-norm sparsity-promoting PMINT technique for all acoustic systems and NPMs. Similarly, the same regularization parameter δ is used in the regularized PMINT technique for all acoustic systems and NPMs. To systematically compare both techniques, we investigate the performance for two different settings of these parameters, i.e., the parameters η, ρ, and δ are determined as the parameters minimizing CD for acoustic system S_1 and

i) NPM = -33 dB, resulting in $\eta = 10^{-7}$, $\rho = 10^{-4}$, and $\delta = 10^{-6}$, and

ii) NPM = -3 dB, resulting in $\eta = 10^{-7}$, $\rho = 10^{-3}$, and $\delta = 10^{-1}$.

As indicated by the obtained parameter values, the weighting and penalty parameters yielding a high performance for the sparsity-promoting PMINT technique do not change significantly for different NPMs, whereas the regularization parameter yielding a high performance for the regularized PMINT technique significantly changes for different NPMs. This is not surprising since the regularized cost function in (18) explicitly incorporates an RIR perturbation model in the filter design where the appropriate regularization parameter depends on the distortion energy due to RIR perturbations, whereas the sparsity-promoting cost function in (24) only relies on characteristics of clean speech signals. The performance of both techniques is evaluated in terms of ΔDRR, ΔPESQ, and ΔCD, and the presented performance measures are averaged over all considered NPMs.

Table V presents the obtained ΔDRR, ΔPESQ, and ΔCD values for both considered techniques using the parameters in case i). In addition, the average performance measures over all acoustic systems are also presented. It can be observed that for these parameters the sparsity-promoting PMINT technique and the regularized PMINT technique yield a very similar performance in terms of all performance measures. While the sparsity-promoting PMINT technique results in a slightly higher ΔDRR and ΔPESQ, the regularized PMINT technique results in a slightly lower ΔCD.

Table VI presents the obtained ΔDRR, ΔPESQ, and ΔCD values for both considered techniques using the parameters in case ii). In addition, the average performance measures over all acoustic systems are also presented. It can now be observed that for these parameters the sparsity-promoting PMINT technique significantly outperforms the regularized PMINT technique in terms of all performance measures,

TABLE V: Performance of the weighted l_1-norm sparsity-promoting PMINT technique ($\eta = 10^{-7}$, $\rho = 10^{-4}$) and regularized PMINT technique ($\delta = 10^{-6}$).

	ΔDRR [dB]				ΔPESQ				ΔCD [dB]			
	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average
Weighted l_1-S-PMINT	**8.2**	**7.0**	**6.4**	**7.2**	**0.6**	0.5	**0.6**	**0.6**	−1.2	−1.2	−0.7	−1.0
Regularized PMINT	7.9	**7.0**	5.4	6.8	0.5	**0.6**	0.5	0.5	−1.1	**−1.3**	**−0.8**	**−1.1**

TABLE VI: Performance of the weighted l_1-norm sparsity-promoting PMINT technique ($\eta = 10^{-7}$, $\rho = 10^{-3}$) and regularized PMINT technique ($\delta = 10^{-1}$).

	ΔDRR [dB]				ΔPESQ				ΔCD [dB]			
	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average	S_1	S_2	S_3	Average
weighted l_1-S-PMINT	**8.8**	**7.4**	**6.6**	**7.6**	**0.7**	**0.6**	**0.7**	**0.7**	**−1.3**	**−1.3**	**−0.6**	**−1.1**
regularized PMINT	3.4	5.8	3.6	4.3	0.2	0.5	0.2	0.3	−0.6	−0.8	−0.3	−0.6

yielding on average a higher ΔDRR of 3.3 dB, a higher ΔPESQ of 0.4, and a lower ΔCD of 0.5 dB.

In summary, these results show that the proposed signal-dependent sparsity-promoting PMINT technique yields a similar or significantly better dereverberation performance than the signal-independent regularized PMINT technique, depending on the used weighting, penalty, and regularization parameters. This shows the potential of incorporating sparsity-promoting penalty functions which exploit well-established characteristics of clean speech signals to increase the robustness of equalization techniques against RIR perturbations. Combining signal-independent regularization based on an RIR perturbation model and signal-dependent penalty functions remains a topic for future investigation.

VII. Conclusion

In this paper we have proposed to increase the robustness of acoustic multi-channel equalization techniques against RIR perturbations using signal-dependent penalty functions, such that the output signal better resembles a clean speech signal. We have proposed to extend the cost function of least-squares equalization techniques with different sparsity-promoting penalty functions, i.e., the l_0-norm, the l_1-norm, and the weighted l_1-norm. The sparsity-promoting filters have been iteratively computed based on the ADMM. Simulation results show that incorporating sparsity-promoting penalty functions in all considered equalization techniques, i.e., MINT, RMCLS, and PMINT, significantly increases the robustness against RIR perturbations and suppresses artifacts generated by non-robust equalization techniques. In addition, it is shown that the weighted l_1-norm penalty function typically outperforms the l_0- and l_1-norm penalty functions, with the weighted l_1-norm sparsity-promoting PMINT technique yielding a high reverberant energy suppression and outperforming all other considered techniques in terms of perceptual speech quality improvement. Finally, it is shown that the proposed signal-dependent weighted l_1-norm sparsity-promoting PMINT technique yields a similar or significantly better dereverberation performance than the signal-independent regularized PMINT technique, confirming the advantage of using signal-dependent

penalty functions for robust dereverberation filter design and motivating future research in this direction.

References

[1] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, Jul. 2006.

[2] S. Goetze, E. Albertin, J. Rennies, E. A. P. Habets, and K. D. Kammeyer, "Speech quality assessment for listening-room compensation," in *Proc. AES International Conference on Sound Quality Evaluation*, Pitea, Sweden, Jun. 2010, pp. 11–20.

[3] A. Warzybok, I. Kodrasi, J. O. Jungmann, E. A. P. Habets, T. Gerkmann, A. Mertins, S. Doclo, B. Kollmeier, and S. Goetze, "Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithm," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sep. 2014, pp. 333–337.

[4] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, Nov. 2012.

[5] F. Xiong, B. T. Meyer, N. Moritz, R. Rehr, J. Anemüller, T. Gerkmann, S. Doclo, and S. Goetze, "Front-end technologies for robust ASR in reverberant environments–spectral enhancement-based dereverberation and auditory modulation filterbank features," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–18, Aug. 2015.

[6] P. A. Naylor and N. D. Gaubitch, Eds., *Speech dereverberation*. London, UK: Springer, 2010.

[7] K. Lebart and J. M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, May-Jun. 2001.

[8] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–774, Sep. 2009.

[9] S. Braun and E. A. P. Habets, "A multichannel diffuse power estimator for dereverberation in the presence of multiple sources," *EURASIP Journal on Applied Signal Processing*, vol. 2015, no. 24, Dec. 2015.

[10] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum Likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1595–1608, Sep. 2016.

[11] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1717–1731, Sep. 2010.

[12] D. Schmid, G. Enzner, S. Malik, D. Kolossa, and R. Martin, "Variational Bayesian inference for multichannel dereverberation and noise reduction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 8, pp. 1320–1335, Aug. 2014.

[13] B. Schwartz, S. Gannot, and E. A. P. Habets, "Online speech dereverberation using Kalman filter and EM algorithm," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 394–406, Feb. 2015.

[14] A. Jukić, T. Van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1509–1520, Sep. 2015.

[15] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[16] M. Kallinger and A. Mertins, "Multi-channel room impulse response shaping - a study," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006, pp. 101–104.

[17] J. O. Jungmann, R. Mazur, M. Kallinger, M. Tiemin, and A. Mertins, "Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1829–1842, Aug. 2012.

[18] F. Lim, W. Zhang, E. A. P. Habets, and P. A. Naylor, "Robust multichannel dereverberation using relaxed multichannel least squares," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1379–1390, Jun. 2014.

[19] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multi-channel equalization for speech dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, Sep. 2013.

[20] R. S. Rashobh, A. W. H. Khong, and D. Liu, "Multichannel equalization in the KLT and frequency domains with application to speech dereverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 634–646, Mar. 2014.

[21] H. Hacihabibouglu and Z. Cvetkovic, "Multichannel dereverberation theorems and robustness issues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 67–689, Feb. 2012.

[22] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

[23] A. W. H. Khong, L. Xiang, and P. A. Naylor, "Algorithms for identifying clusters of near-common zeros in multichannel blind system identification and equalization," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, USA, Mar. 2008, pp. 229–232.

[24] M. A. Haque and T. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," *IEEE Signal Processing Letters*, vol. 15, pp. 305–308, Feb. 2008.

[25] M. Hu, N. D. Gaubitch, P. A. Naylor, and D. B. Ward, "Noise robust blind system identification algorithms based on a Rayleigh quotient cost function," in *Proc. European Signal Processing Conference*, Nice, France, Sep. 2015.

[26] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[27] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 996–1005, Aug. 2011.

[28] A. Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier, "Effects of spatial and temporal integration of a single early reflection on speech intelligibility," *Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 269–282, Jan. 2013.

[29] N. D. Gaubitch and P. A. Naylor, "Equalization of multichannel acoustic systems in oversampled subbands," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1061–1070, Aug. 2009.

[30] F. Lim and P. A. Naylor, "Robust speech dereverberation using subband multichannel least squares with variable relaxation," in *Proc. European Signal Processing Conference*, Marrakech, Morocco, Sep. 2013.

[31] I. Kodrasi and S. Doclo, "The effect of inverse filter length on the robustness of acoustic multichannel equalization," in *Proc. European Signal Processing Conference*, Bucharest, Romania, Aug. 2012.

[32] S. Makino, S. Araki, and H. Sawada, "Underdetermined blind source separation using acoustic arrays," in *Handbook on array processing and sensor networks*, S. Haykin and K. J. R. Liu, Eds. Hoboken, USA: John Wiley & Sons, 2010.

[33] S. Arberet, P. Vandergheynst, R. E. Carrillo, J. Thiran, and Y. Wiaux, "Sparse reverberant audio source separation via reweighted analysis,"

[34] *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 7, pp. 1391–1402, Mar. 2013.

[34] K. Kumatani, J. McDonough, B. Rauch, D. Klakow, P. N. Garner, and L. Weifeng, "Beamforming with a maximum negentropy criterion," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 5, pp. 994–1008, Jul. 2009.

[35] T. Van Waterschoot, B. Defraene, M. Diehl, and M. Moonen, "Embedded optimization algorithms for multi-microphone dereverberation," in *Proc. European Signal Processing Conference*, Marrakech, Morocco, Sep. 2013.

[36] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "A framework for multi-channel speech dereverberation by exploiting sparsity," in *Proc. AES International Conference on Dereverberation and Reverberation of Audio, Music, and Speech*, Leuven, Belgium, Feb. 2016.

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

[38] I. Kodrasi, A. Jukić, and S. Doclo, "Robust sparsity-promoting acoustic multi-channel equalization for speech dereverberation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016, pp. 166–170.

[39] G. Harikumar and Y. Bresler, "FIR perfect signal reconstruction from multiple convolutions: minimum deconvolver orders," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 215–218, Jan. 1998.

[40] F. Lim and P. Naylor, "Statistical modelling of multichannel blind system identification errors," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sep. 2014, pp. 119–123.

[41] I. Tashev and A. Acero, "Statistical modeling of the speech signal," in *Proc. International Workshop on Acoustic, Echo, and Noise Control*, Tel Aviv, Israel, Sep. 2010.

[42] T. Gerkmann and R. Martin, "Empirical distributions of DFT-domain speech coefficients based on estimated speech variances," in *Proc. International Workshop on Acoustic, Echo, and Noise Control*, Tel Aviv, Israel, Sep. 2010.

[43] Y. Lin, J. Chen, Y. Kim, and D. D. Lee, "Blind channel identification for speech dereverberation using l_1-norm sparse learning," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., 2008, pp. 921–928.

[44] K. Kowalczyk, E. A. P. Habets, W. Kellermann, and P. A. Naylor, "Blind system identification using sparse learning for TDOA estimation of room reflections," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 653–656, July 2013.

[45] 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, Sep. 2010.

[46] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[47] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, United Kingdom: Cambridge University Press, 2004.

[48] R. Chartrand, "Shrinkage mappings and their induced penalty functions," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 2014, pp. 1026–1029.

[49] E. J. Candés, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, Oct. 2008.

[50] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, USA: Springer, 2006.

[51] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, USA: The John Hopkins University Press, 1996.

[52] N. Parikh and S. P. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, Jan. 2014.

[53] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES Convention*, Paris, France, Feb. 2000, pp. 18–22.

[54] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[55] W. Zhang and P. A. Naylor, "An algorithm to generate representations of system identification errors," *Research Letters in Signal Processing*, vol. 2008, Jan. 2008.

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

[57] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.

[58] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*. New Jersey, USA: Prentice-Hall, 1988.

Ina Kodrasi received the Master of Science degree in Communications, Systems and Electronics in 2010 from Jacobs University Bremen, Germany and the PhD degree in 2015 from the University of Oldenburg, Germany. From 2010 to 2015 she worked as an early stage researcher at the Signal Processing Group of the University of Oldenburg. From 2010 to 2011 she was also with the Fraunhofer Institute for Digital Media Technology (IDMT), Project group Hearing, Speech and Audio Technology in Oldenburg where she worked on microphone array beamforming. Since December 2015 she is a postdoctoral researcher at the Signal Processing Group of the University of Oldenburg in the field of speech dereverberation and noise reduction.

Simon Doclo (S'95, M'03) Simon Doclo received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003. From 2003 to 2007 he was a Postdoctoral Fellow with the Research Foundation Flanders at the Electrical Engineering Department (Katholieke Universiteit Leuven) and the Adaptive Systems Laboratory (McMaster University, Canada). From 2007 to 2009 he was a Principal Scientist with NXP Semiconductors at the Sound and Acoustics Group in Leuven, Belgium. Since 2009 he is a full professor at the University of Oldenburg, Germany, and scientific advisor for the project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical and biomedical applications, more specifically microphone array processing, active noise control, acoustic sensor networks and hearing aid processing. Prof. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with Marc Moonen) and the IEEE Signal Processing Society 2008 Best Paper Award (with Jingdong Chen, Jacob Benesty, Arden Huang). He was member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008-2013) and Technical Program Chair for the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Doclo has served as guest editor for several special issues (IEEE Signal Processing Magazine, Elsevier Signal Processing) and is associate editor for IEEE/ACM Transactions on Audio, Speech and Language Processing and EURASIP Journal on Advances in Signal Processing.