Low-Rank Acoustic Room Impulse Response Estimation

Martin Jälmby[1], Filip Elvander[1], and Toon van Waterschoot[1]

[1]KU Leuven, Dept. of Electrical Engineering, ESAT-STADIUS, Leuven, Belgium *

***Abstract*— We consider low-rank estimation of acoustic room impulse responses (RIRs). We propose an estimator of that promotes a low-rank structure for a matricization of the estimated RIR. The low-rank prior acts as a regularizer for the inverse problem of estimating RIRs from noisy input-output relations, preventing overfitting and improving estimation performance. We consider two different relaxations, one using the nuclear norm, and one using the recently introduced concept of quadratic envelopes.**

1 Introduction

Room impulse responses (RIRs) describe the impact of the room on a sound signal, between a point source and a point receiver. They play a vital part in algorithms within a multitude of acoustic signal processing tasks, see e.g. [1, 2, 3, 4, 5]. Among the more popular ways to model the RIR are the infinite impulse response (IIR) (see e.g. [6]) and finite impulse response (FIR) models (see e.g. [7]). An FIR filter can be several thousands of taps long [2], which can be prohibitive from a computational and memory requirement point for view. We will demonstrate in this work that by exploiting an (approximate) low-rank structure of RIRs, the estimation performance, when identifying an RIR from input-output data, may be improved as compared to state-of-the-art estimators with and without regularization. We exploit that an RIR vector when reshaped into a matrix allow for low-rank approximations, something we have taken advantage of in recent work [8] (see also [9, 10]). We propose an estimator where the rank of the estimate is implicitly controlled by means of hyperparameters. The proposed estimator poses RIR estimation as an inverse problem, promoting low-rank by the use of a regularization term. Directly penalizing the rank of an RIR matrix yields an NP-hard problem [11]. A popular remedy to this is nuclear norm regularization. This yields a convex optimization problem that can be solved by standard methods. This approach has, however, been shown to introduce a bias towards zero of the singular values (see e.g. [12]). Here we propose to counter this bias by utilizing the so-called quadratic envelope of the rank penalty, see e.g. [13]. We here provide a framework for low-rank estimation of RIRs. We present an easy-to-use algorithm and show that low-rank regularization outperforms benchmark methods, that is, ordinary least squares and Tikhonov regularized least squares.

2 Low-rank Estimation

We assume an observed acoustic pressure signal $\mathbf{y} \in \mathbb{R}^{n_y}$. This observed signal is the result of a known signal $\mathbf{x} \in \mathbb{R}^{n_x}$ being generated by a point source in the room, and therefore being convolved with the RIR $\mathbf{h} \in \mathbb{R}^{n_h}$, $\mathbf{x} * \mathbf{h}$, where $(\mathbf{x} * \mathbf{h})_n = \sum_{k=1}^{n_h} \mathbf{h}_k \mathbf{x}_{n+n_h-k}$, $n = 1, 2, \ldots, n_y$, i.e. we consider only the part of the observed signal where the system is fully excited. We assume that $n_x \geq n_h$, to not have to consider initial conditions. We further assume white Gaussian measurement noise $\mathbf{e} \in \mathbb{R}^{n_y}$, with variance σ^2, i.e. $\mathbf{e} \in \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_y})$. Taken together, $\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{e}$, with $n_y = n_x - n_h + 1$. We define the linear operator $\mathcal{A} : \mathbb{R}^{n_h} \to \mathbb{R}^{n_y}$, $\mathcal{A}(\mathbf{h}) = \mathbf{x} * \mathbf{h}$. We also define the adjoint operator $\mathcal{A}^* : \mathbb{R}^{n_y} \to \mathbb{R}^{n_h}$, as the unique operator for which $\langle \mathbf{h}, \mathcal{A}^*(\mathbf{y}) \rangle_{n_h} = \langle \mathcal{A}(\mathbf{h}), \mathbf{y} \rangle_{n_y}$, $\forall\, \mathbf{h} \in \mathbb{R}^{n_h}, \mathbf{y} \in \mathbb{R}^{n_y}$.

The RIR $\mathbf{h}$ can be estimated from the input and observed signal vectors with the least squares method,

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2, \tag{1}$$

and under the assumption that the noise $\mathbf{e}$ is white Gaussian, this yields the maximum likelihood estimatior. Inverse problems like these are, however, often ill-posed [14], particularly in acoustic signal processing, where poor excitation is common [15]. The inverse mapping from $\mathbf{y}$ to $\mathbf{h}$ might therefore be unstable due to ill-conditioning, and the presence of noise could yield an over-fitted solution [14]. To have a well-posed problem, a possible approach is to use regularization [16], by appending a penalty term to the optimization problem (1),

$$\underset{\mathbf{h}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\mathbf{h})\|_2^2 + g(\mathbf{h}), \tag{2}$$

where the function $g : \mathbb{R}^{n_h} \to \mathbb{R}$ is the penalty term. This form includes the aforementioned Tikhonov regularization, $g(\mathbf{h}) = \lambda_{\mathrm{T}} \|\mathbf{h}\|_2^2$, where $\lambda_{\mathrm{T}} \geq 0$ controls the degree of regularization.

It has previously been demonstrated how the RIR can be well modeled as a sum of decaying sinusoids (see e.g. [8]),

$$h(\mathbf{r}_r, \mathbf{r}_s, n) = \sum_{m=1}^{m_s} \mu_m(\mathbf{r}_r, \mathbf{r}_s) e^{-\beta_m n} \cos(\omega_m n + \phi_m),$$

for $n = 1, 2, \ldots, n_h$. Here, μ_m denotes the initial amplitude, $\mathbf{r}_r, \mathbf{r}_s \in \mathbb{R}^3$ the position of the receiver and the source, respectively, $\beta_m \in \mathbb{R}_+$ the exponential decay constant, $\omega_m \in [0, \pi]$ angular frequency, $\phi_m \in [0, 2\pi)$ the phase, and $m_s \in \mathbb{N}$ the number of decaying sinusoids. We will from hereon refer to $h(\mathbf{r}_r, \mathbf{r}_s, n)$ as simply $h(n)$. The RIR vector is given by $\mathbf{h} = [h(1), h(2), \ldots, h(n_h)]^T$. We assume that $q = \sqrt{n_h} \in \mathbb{N}$, and define the operator $\mathcal{R}_{qq} : \mathbb{R}^{n_h} \to \mathbb{R}^{q \times q}$, that reshapes $\mathbf{h} \in \mathbb{R}^{n_h}$ into $\mathbf{H} \in \mathbb{R}^{q \times q}$. When a vector consisting of the sum of m_s sampled decaying sinusoids is reshaped into a matrix, that matrix will have rank $2m_s$, see e.g. [17]. *Low-rank* is here meant in a loose sense, i.e. that a low-rank approximation will render a small error.

In light of this, we would like to solve the problem (1), but with the added constraint that $\mathrm{rank}(\mathcal{R}_{qq}(\mathbf{h})) \leq 2m_s$). This problem, however, is and NP-hard [11]. Further, the rank, $2m_s$, is typically not known. Instead, one could consider the penalized version, on the form of (2), with $g(\mathbf{h}) = \lambda \|\sigma(\mathcal{R}_{qq}(\mathbf{h}))\|_0$. Here, $\lambda \geq 0$, analogous to λ_{T}, controls the degree of penalization, and we have used the fact that the rank of a matrix is equal to the number of non-zero singular values, given its singular value decomposition (SVD) $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Further,

*This research work was carried out at the ESAT Laboratory of KU Leuven. The research leading to these results has received funding from the Research Foundation – Flanders (FWO, grant 12ZD622N) and from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

$\sigma(\mathbf{H}) : \mathbb{R}^{q\times q} \to \mathbb{R}^q$ takes $\mathbf{H} \in \mathbb{R}^{q\times q}$ and returns a vector of its singular values in non-decreasing order. This changes the problem of not knowing the rank, into the problem of choosing λ. The NP-hardness persists, however, so to be able to find an approximate solution, the problem needs to be relaxed.

In order to achieve low-rank solutions, two different penalty functions will be used and compared. Firstly, the popular approach of using the nuclear norm [18], $g_{\text{NN}}(\mathbf{h}) = \lambda_{\text{NN}}\|\mathcal{R}_{qq}(\mathbf{h})\|_*$, where $\|\mathbf{H}\|_* = \text{trace}((\mathbf{H}^*\mathbf{H})^{1/2}) = \|\sigma(\mathbf{H})\|_1$, and $\lambda_{\text{NN}} \geq 0$. The advantage of this penalty function is that the problem will be convex and can be solved efficiently [19]. The disadvantage is that this solution will be biased, since the singular values will be shrunk [12]. Secondly, the use of a function created in attempt to closer emulate the ℓ_0-norm, the *quadratic envelope* of the ℓ_0-norm. The quadratic envelope of a function consists of pointwise, quadratic approximations, that nowhere overestimates the original function. For a function $g : \mathbb{R}^n \to \mathbb{R}$, the quadratic envelope, at $\mathbf{u} \in \mathbb{R}^n$, is defined as

$$\mathcal{Q}_\gamma(g)(\mathbf{u}) = \sup_{\alpha\in\mathbb{R},\mathbf{v}\in\mathbb{R}^n} \left\{\alpha - \frac{\gamma}{2}\|\mathbf{u}-\mathbf{v}\|^2 : \alpha - \frac{\gamma}{2}\|\mathbf{z}-\mathbf{v}\|^2 \leq g(\mathbf{z}),\ \forall\, \mathbf{z}\in\mathbb{R}^n\right\},$$

i.e. the supremum of all functions of the form $\alpha - \frac{\gamma}{2}\|\mathbf{u}-\mathbf{v}\|^2$, that nowhere overestimates the original function. The parameter $\gamma > 0$ controls how close this envelope will be to the original function. A smaller γ means a larger degree of smoothing of the function, whereas $\mathcal{Q}_\gamma(g)$ converges point-wise to g as $\gamma \to \infty$ [13]. The quadratic envelope of $\lambda_{\text{QE}}\|u\|_0$ is given by

$$\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(u) = \begin{cases} \sqrt{2\lambda_{\text{QE}}\gamma}u - \frac{\gamma}{2}u^2 & ,\ u < \mu \\ \lambda_{\text{QE}} & ,\ \mu \leq u, \end{cases}$$

where $\mu = \sqrt{2\lambda_{\text{QE}}/\gamma}$, and $\lambda_{\text{QE}} \geq 0$ is analogous to λ_{NN} and λ_{T}. Proof can be found in [20]. The benefit of the quadratic envelope is that for values $u \geq \mu$, the quadratic envelope is constant. In (2), using $g(\mathbf{h}) = g_{\mathcal{Q}}(\mathbf{h}) = \mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(\mathcal{R}_{qq}(\mathbf{h}))$ yields the final form of the second optimization problem we are aiming to solve. For a matrix $\mathbf{X}$, we define $\mathcal{Q}(\mathbf{X}) = \mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)(\mathbf{X})$. For $\gamma > \|\mathcal{A}\|$, which will be used here, we have that if $\hat{\mathbf{h}}$ is a local minimizer of (2) with $g(\mathbf{h}) = \mathcal{Q}(\mathbf{H})$, then $\hat{\mathbf{h}}$ is also a local minimizer of (2) with $g(\mathbf{h}) = \lambda\|\sigma(\mathcal{R}_{qq}(\mathbf{h}))\|_0$, and their global minima coincide [13].

3 Algorithm

The optimization problems will be solved using the proximal gradient method, see e.g. [14]. In order to facilitate the exposition, we will introduce a few linear operators. The operator $\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^n$ reverses the order of a vector, i.e $\mathcal{H}(\mathbf{x}) = [\mathbf{x}_{n_x}, \mathbf{x}_{n_x-1}, \ldots, \mathbf{x}_1]^T$, $\mathcal{F} : \mathbb{R}^n \to \mathbb{C}^n$ denotes the Fourier transform, $\mathcal{P}^k : \mathbb{R}^n \to \mathbb{R}^{n+k}$ denotes padding a vector with k zeroes at the end, and $\mathcal{C}_k^n : \mathbb{R}^m \to \mathbb{R}^{n-k+1}$, where $n \geq k$ and $m \geq n-k+1$, picks out the kth till the nth element of a vector, i.e. $\mathcal{C}_k^n(\mathbf{y}) = [y_k,\ y_{k+1},\ \ldots,\ y_n]$. The operators $\mathcal{A}$ and $\mathcal{A}^*$ can now be expressed in terms of these operators,

$$\mathcal{A}(\mathbf{h}) = \mathcal{C}_{n_h}^{n_x}(\mathcal{F}^{-1}(\mathcal{F}(\mathcal{P}^{n_x-1}(\mathbf{h})) \odot \mathcal{F}(\mathcal{P}^{n_h-1}(\mathbf{x})))),$$
$$\mathcal{A}^*(\mathbf{y}) = \mathcal{C}_{n_y}^{n_x}(\mathcal{F}^{-1}(\mathcal{F}(\mathcal{P}^{n_x-n_h}(\mathcal{H}(\mathbf{x}))) \odot \mathcal{F}(\mathcal{P}^{2n_x-n_h-n_y}(\mathbf{y})))),$$

respectively, where $\odot$ denotes the Hadamard product.

The first step in iteratively solving (2), with either $g(\mathbf{h}) = g_{\text{NN}}(\mathbf{h})$, or $g(\mathbf{h}) = g_{\mathcal{Q}}(\mathbf{h})$, is the gradient step, with respect to the data-fit term (1). The gradient of (1) is given by $\mathcal{A}^*(\mathcal{A}(\mathbf{h}^k)) - \mathcal{A}^*(\mathbf{y})$. For the backward step, the seemingly complicating fact that it is to be taken with respect to a matrix and its SVD, turns out not to be a problem. After the gradient step, the vector is reshaped into a matrix, and an SVD is performed on that matrix. The proximal operators of $\mathcal{Q}(\mathbf{H})$ and $\lambda_{\text{NN}}\|\sigma(\mathbf{H})\|_1$ are given by

$$\begin{aligned} \text{prox}_{\rho\mathcal{Q}(\cdot)}(\mathbf{H}) &= \mathbf{U}\,\mathcal{D}(\text{prox}_{\rho\mathcal{Q}_\gamma(\lambda_{\text{QE}}\|\cdot\|_0)}(\sigma(\mathbf{H})))\,\mathbf{V}^T, \\ \text{prox}_{\rho\lambda_{\text{NN}}\|\cdot\|_1}(\mathbf{H}) &= \mathbf{U}\,\mathcal{D}(\text{prox}_{\rho\lambda_{\text{NN}}\|\cdot\|_1}(\sigma(\mathbf{H})))\,\mathbf{V}^T, \end{aligned} \tag{3}$$

respectively, where $\mathcal{D} : \mathbb{R}^q \to \mathbb{R}^{q\times q}$ creates a diagonal matrix, with the argument on the diagonal. It may be noted that (3) is due to the orthogonal invariance of $\mathcal{Q}(\mathbf{H})$ and $\lambda_{\text{NN}}\|\sigma(\mathbf{H})\|_1$, respectively [21]. Proof can be found in [20]. After either of the proximal operators has acted on the vectors of singular values, the matrix is reassembled, with the adjusted singular values, and then vectorized, in order to get the current iterate of the estimated RIR. Finally, a check is made to see if the algorithm is making enough progress, or if the iterative procedure should be terminated. These steps are summarized in Algorithm 1. Solving (2) with $g(\mathbf{h}) = g_{\text{NN}}(\mathbf{h})$ or $g(\mathbf{h}) = g_{\mathcal{Q}}(\mathbf{h})$ differs only in the use of proximal operator, reflected in step 3 of Algorithm 1, where 3 a) corresponds to solving (2) with $g(\mathbf{h}) = g_{\text{NN}}(\mathbf{h})$ and 3 b) corresponds to solving (2) with $g(\mathbf{h}) = g_{\mathcal{Q}}(\mathbf{h})$.

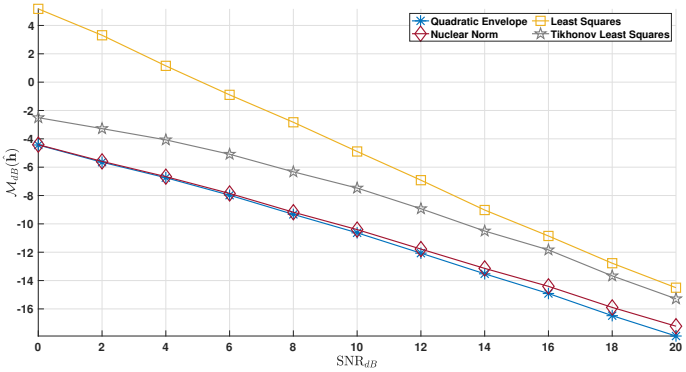

Figure 1: Normalized misalignment as a function of SNR_{dB}.

Algorithm 1 Low-Rank RIR Estimation Algorithm

Input: $\mathbf{h}^0$, ρ, γ, λ, $\mathbf{x}$, $\mathbf{y}$, tol, maxIter
for $k = 0, 1, ...,$ maxIter **do**
1: $\mathbf{w}^k = \mathbf{h}^k - \rho(\mathcal{A}^*(\mathcal{A}(\mathbf{h}^k)) - \mathcal{A}^*(\mathbf{y}))$ {Forward step}
2: $[\mathbf{U}^k, \mathbf{S}^k, \mathbf{V}^k] = \text{SVD}(\mathcal{R}_{qq}(\mathbf{w}^k))$
3 a): $\mathbf{s}^{k+1} = \mathcal{S}_{\lambda_{\text{NN}}}(\sigma(\mathbf{W}^k))$
3 b): $\mathbf{s}^{k+1} = \text{prox}_{\rho\mathcal{Q}(\cdot)}(\mathbf{W}^k)$
4: $\mathbf{h}^{k+1} = \mathcal{R}_{qq}^{-1}(\mathbf{U}^k\mathcal{D}(\mathbf{s}^{k+1})\mathbf{V}^{k^T})$ {Backward step}
if $\|\mathbf{h}^{k+1} - \mathbf{h}^k\|/(\|\mathbf{h}^k\| + \delta) \leq$ tol
break
end if
end for

4 Numerical Results

To demonstrate the proposed methods' applicability to actual measured RIRs, we apply it to the *single- and multichannel audio recordings database* (SMARD) [22]. We set $n_h = 3600$, $n_y = 4320$. The choice of values of γ and the different λ's is explained in [20]. With white Gaussian noise as input, we demonstrate the adequacy of the proposed modeling framework as a function of SNR. The results are shown in Fig. 1. The low-rank enforcing techniques are clearly preferable, particularly at low SNR. For more results we refer the reader to [20].

5 Conclusion

Here we have presented a method for exploiting (approximate) low-rank properties of RIRs. The low-rank penalty serves as a regularizer and prevents overfitting. It is shown that the low-rank techniques performs very well, particularly at low SNR.

References

[1] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1620–1643, 2020.

[2] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer, 2010.

[3] C. Schissler, P. Stirling, and R. Mehra, "Efficient construction of the spatial room impulse response," in *2017 IEEE Virtual Reality (VR)*, 2017, pp. 122–130.

[4] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[5] G. W. Elko, E. Diethorn, and T. Gaensler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. 2003 Int. Workshop Acoustic Echo Noise Control (IWAENC '03)*, Kyoto, Japan, Sept. 2003.

[6] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.

[7] C. Huszty, N. Bukuli, Á. Torma, and F. Augusztinovicz, "Effects of filtering of room impulse responses on room acoustics parameters by using different filter structures," *J. Acoust. Soc. Amer.*, vol. 123, pp. 3617, 2008.

[8] M. Jälmby, F. Elvander, and T. van Waterschoot, "Low-rank tensor modeling of room impulse responses," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 111–115.

[9] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics. Springer Berlin Heidelberg, 2012.

[10] M. Boussé and L. De Lathauwer, "Large-scale autoregressive system identification using kronecker product equations," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 1348–1352.

[11] Y. Xiang, S.and Zhu, X. Shen, and J. Ye, "Optimal exact least squares rank minimization," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2012, KDD '12, pp. 480–488, Association for Computing Machinery.

[12] Z. Hu, F. Nie, R. Wang, and X. Li, "Low rank regularization: A review," *Neural Networks*, vol. 136, pp. 218–232, 2021.

[13] M. Carlsson, "On convex envelopes and regularization of non-convex functionals without moving global minima," *Journal of Optimization Theory and Applications*, vol. 183, no. 1, pp. 66–84, 2019.

[14] N. Antonello, L. Stella, P. Patrinos, and T. van Waterschoot, "Proximal gradient algorithms: Applications in signal processing," arXiv: 1803.01621, 2020.

[15] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, vol. 88, no. 3, pp. 594–611, 2008.

[16] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, vol. 93, no. 2, pp. 173–180, 2020.

[17] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2017.

[18] M. Fazel, H. Hindi, and S.P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the 2001 American Control Conference. (Cat. No.01CH37148)*, 2001, vol. 6, pp. 4734–4739 vol.6.

[19] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[20] M. Jälmby, F. Elvander, and T. van Waterschoot, "Low-rank room impulse response estimation," Tech. Rep., ftp://ftp.esat.kuleuven.be/stadius/mjalmby/LRRIRE.pdf, 2022.

[21] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[22] J. K. Nielsen, J. R. Jensen, S. H. Jensen, and M. G. Christensen, "The single- and multichannel audio recordings database (SMARD)," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Antibes, France, Sept. 2014.