Reweighted ℓ_1 minimization for audio inpainting

Ondřej Mokrý *Pavel Rajmic*

Signal Processing Laboratory, Brno University of Technology, Brno, Czech Republic
Email: {170583, rajmic}@vutbr.cz

***Abstract*—Reweighted $\boldsymbol{\ell_1}$ minimization has been successfully applied to the task of audio declipping by Weinstein and Wakin (2011). We adapt the reweighting to the audio inpainting problem, for both the synthesis and the analysis models. It is shown that the reweighting provides a significant improvement in terms of SNR, especially in the analysis model.**

I. Introduction

In the past, the recovery of missing audio samples, i. e. the *audio inpainting* problem has been addressed by different means. The first sparsity-based method utilized the OMP greedy algorithm [1]. Another way of approximating the sparse prior was the convex minimization (the so-called ℓ_1 relaxation), see for example [2]. Furthermore, this approach was applied to the closely related audio *declipping* problem in [3], [4]. In [5], it was proposed that the performance of the ℓ_1 relaxation might be enhanced by the so-called *reweighting*. Since then, reweighted ℓ_1 methods found their use in different areas of signal processing [6]–[8]

Sparsity-based formulation of audio inpainting with ℓ_1 relaxation, i. e. using the (weighted) ℓ_1 norm, attains the form

$$\arg\min_{\mathbf{z}} \|\mathbf{w} \odot \mathbf{z}\|_1 \quad \text{s.t.} \quad D\mathbf{z} \in \Gamma, \tag{1a}$$
$$\arg\min_{\mathbf{x}} \|\mathbf{w} \odot A\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{x} \in \Gamma, \tag{1b}$$

where the formulations are referred to as the *synthesis* and *analysis* variant, respectively. In (1), let $D : \mathbb{C}^P \to \mathbb{C}^N, P \geq N$ be the synthesis operator of a Parseval frame and let $A = D^*$ be its analysis counterpart [9]. For audio inpainting, the set of feasible solutions is the (convex) set Γ of signals that are equal to the reliable parts of the observed signal. The vector $\mathbf{w} \in \mathbb{R}_+^P$ is the vector of (positive) weights and the symbol $\odot$ denotes the element-wise product. The bigger the weight assigned to a given coefficient, the more it contributes to the objective function, therefore it is more penalized in the minimization.

II. Inpainting algorithm

In [3], the synthesis variant of audio declipping with reweighted ℓ_1 minimization was presented. The idea of reweighting is that the restoration task is solved repeatedly, where each time, the ℓ_1 norm is weighted differently, based on the inverted absolute values of the coefficients from the previous solution. The benefit is that by such a weighting, the significant coefficients are encouraged, while the small coefficients are even more pushed towards zero, which leads to a better approximation of sparsity, i. e. the ℓ_0 (pseudo)norm.

The described approach can be easily adapted to the task of audio inpainting. The algorithm, as presented in [3], is shown in Alg. 1. In the audio inpainting task, only the set of feasible solutions Γ is different compared to the declipping case.

As the second contribution, we include reweighting into the *analysis* variant (1b), which was proposed in [5], but not presented in the application to audio declipping in [3]. The main difference to (1a) is that the solution is the signal in time domain, thus to compute the weights for subsequent iteration, it is necessary to perform one additional analysis. The resulting algorithm is shown in Alg. 2.

Algorithm 1: Synthesis reweighted ℓ_1 audio inpainting [3]

require: $D : \mathbb{C}^P \to \mathbb{C}^N$, $\Gamma \subset \mathbb{C}^N$, $K, \varepsilon, \delta > 0$
1 $k = 1$, $w_i^{(1)} = 1$, $i = 1, \dots, P$
2 **repeat**
3 $\quad \mathbf{z}^{(k)} = \arg\min_{\mathbf{z}} \|\mathbf{w}^{(k)} \odot \mathbf{z}\|_1$ s. t. $D\mathbf{z} \in \Gamma$
4 $\quad w_i^{(k)} = 1/(|z_i^{(k)}| + \varepsilon)$, $i = 1, \dots, P$
5 $\quad k \leftarrow k + 1$
6 **until** $k > K$ **or** $\|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|_2 < \delta$
7 **return** $\mathbf{x} = D\mathbf{z}^{(k-1)}$

Algorithm 2: Analysis reweighted ℓ_1 audio inpainting

require: $D : \mathbb{C}^P \to \mathbb{C}^N$, $A = D^*$, $\Gamma \subset \mathbb{C}^N$, $K, \varepsilon, \delta > 0$
1 $k = 1$, $w_i^{(1)} = 1$, $i = 1, \dots, P$
2 **repeat**
3 $\quad \mathbf{x}^{(k)} = \arg\min_{\mathbf{x}} \|\mathbf{w}^{(k)} \odot A\mathbf{x}\|_1$ s. t. $\mathbf{x} \in \Gamma$
4 $\quad \mathbf{z}^{(k)} = A\mathbf{x}^{(k)}$
5 $\quad w_i^{(k)} = 1/(|z_i^{(k)}| + \varepsilon)$, $i = 1, \dots, P$
6 $\quad k \leftarrow k + 1$
7 **until** $k > K$ **or** $\|\mathbf{z}^{(k)} - \mathbf{z}^{(k-1)}\|_2 < \delta$
8 **return** $\mathbf{x}^{(k-1)}$

III. Experiments

A test set of 10 musical audio signals sampled at 44.1 kHz with approximate length of 7 seconds was selected from the EBU SQAM database [10] to be diverse in tonal character and sparsity with respect to the used time-frequency transform (see Tab. I). For each test signal and given length of the gaps between 5 and 50 ms, 10 gaps were created and restored using Alg. 1 and 2. The particular inpainting problems for fixed weights—step 3 in both algorithms—were solved by proximal algorithms: the Douglas-Rachford [11] and Chambolle-Pock [12] for the synthesis and analysis variants, respectively.

The results were evaluated with the common *signal-to-noise ratio* (SNR) [1]. Average values from all music samples for given gap lengths are shown in Fig. 1. It can be seen that the reweighting in synthesis model provides consistent, but not quite significant improvement. The analysis model with reweighting, on the other hand, outperforms the simple model significantly for gaps larger than 10 ms. The statistical significance is illustrated by bootstrap 95% confidence intervals [16] in Fig. 2.

IV. Conclusion

This work demonstrates the utilization of the reweighted ℓ_1 norm for audio inpainting. It is shown that when reweighting is combined with the analysis model, a significant improvement of the reconstruction quality in terms of the SNR is observed.

The work was supported by the joint project of the FWF and the Czech Science Foundation (GAČR): numbers I 3067-N30 and 17-33798L, respectively. Research described in this paper was financed by the National Sustainability Program under grant LO1401. Infrastructure of the SIX Center was used.

TABLE I: Algorithm settings for the experiment

parameter	value
transform	DGT [13]
window type	Hann
window length	2800 (64 ms)
window shift	700 (16 ms)
frequency channels	2800
K	10
ε	0.001
δ	0.01

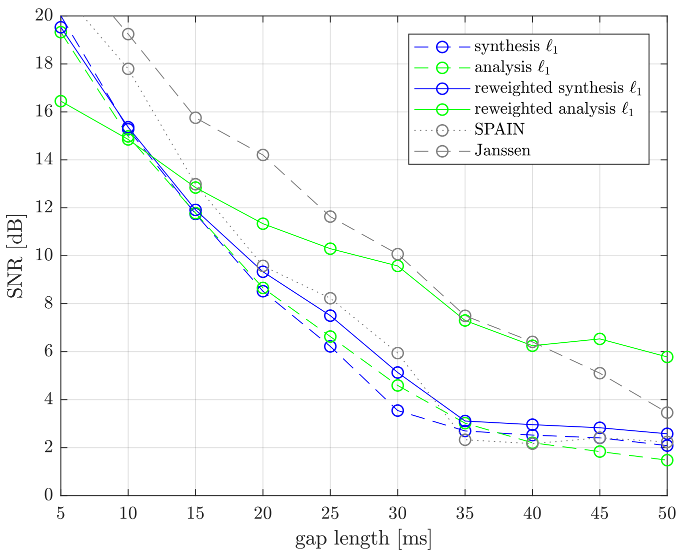

Fig. 1: Values of SNR computed for each of the restored gaps and then averaged in dB. For the ℓ_1 synthesis and the analysis model, single realizations of Douglas-Rachford and Chambolle-Pock algorithms (with no weighting) are used as a reference, respectively. These algorithms, both as a reference and as part of the reweighted approach, were limited to 1 000 iterations, or stopped by relative change of the norm of the main variable in subsequent iterations lower than 10^{-4}. Furthermore, the results are compared to SPAIN [14] and Janssen algorithm based on linear prediction [15], both applied frame-wise with window parameters according to Tab. I.

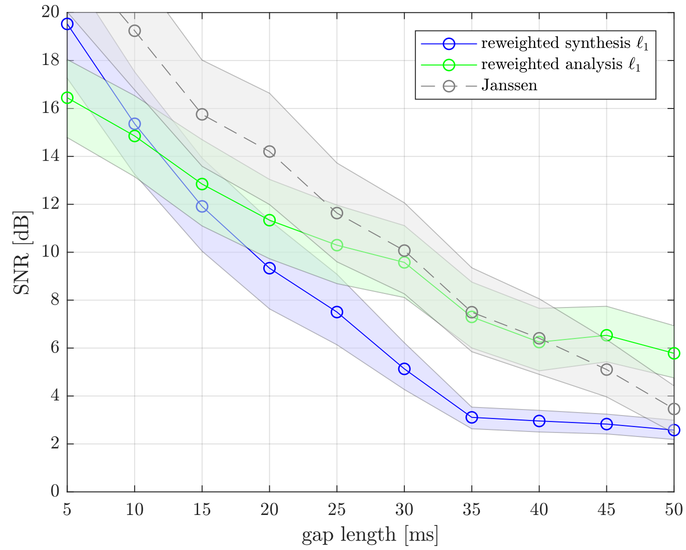

Fig. 2: Interval estimates of the expected values of selected curves from Fig. 1. The estimate with confidence level 0.05 was computed using bootstrapping [16] with 10 000 random draws from the population for each combination of algorithm and gap length.

References

[1] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. Plumbley, "Audio Inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, March 2012.

[2] F. Lieb and H.-G. Stark, "Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches," *Signal Processing*, vol. 153, pp. 291–299, December 2018.

[3] A. J. Weinstein and M. B. Wakin, "Recovering a clipped signal in sparseland," *CoRR*, vol. abs/1110.5063, 2011. [Online]. Available: http://arxiv.org/abs/1110.5063

[4] B. Defraene, N. Mansour, S. D. Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen, "Declipping of audio signals using perceptual compressed sensing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2627–2637, December 2013.

[5] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, December 2008.

[6] M. Novosadová and P. Rajmic, "Piecewise-polynomial signal segmentation using reweighted convex optimization," in *Proceedings of the 40th International Conference on Telecommunications and Signal Processing (TSP)*, Barcelona, 2017, pp. 769–774.

[7] M. Daňková and P. Rajmic, "Low-rank model for dynamic MRI: joint solving and debiasing," in *ESMRMB 2016, 33rd Annual Scientific Meeting, Vienna*. Berlin: Springer, 2016, pp. 200–201.

[8] C. M. Alaíz and J. A. K. Suykens, "Modified Frank-Wolfe algorithm for enhanced sparsity in support vector machine classifiers," *Neurocomputing*, vol. 320, pp. 47–59, 2018.

[9] O. Christensen, *An Introduction to Frames nad Riesz Bases*. Boston-Basel-Berlin: Birkhäuser, 2003.

[10] EBU SQAM CD: Sound quality assessment material recordings for subjective tests. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[11] P. Combettes and J. Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, 2011.

[12] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[13] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox 2.0," in *Sound, Music, and Motion* Springer International Publishing, 2014, pp. 419–442. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-12976-1_25

[14] O. Mokrý, P. Záviška, P. Rajmic, and V. Veselý, "Introducing SPAIN (SParse Audio INpainter)," 2018. [Online]. Available: https://arxiv.org/abs/1810.13137

[15] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 34, no. 2, pp. 317–330, April 1986.

[16] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapmann & Hall, 1994.