Abstract—Reweighted ℓ1 minimization has been successfully applied to the task of audio declipping by Weinstein and Wakin [1]. 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 ℓ1 relaxation), see for example [2], [3]. 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 ℓ1 relaxation might be enhanced by the so-called reweighting. Since then, reweighted ℓ1 methods found their use in different areas of signal processing [6]–[8].

Sparsity-based formulation of audio inpainting with ℓ1 relaxation, i.e. using the (weighted) ℓ1 norm, attains the form

\[
\arg\min_{x} \|w \odot x\|_1 \quad \text{s.t.} \quad Dz \in \Gamma,
\]

which are the formulations referred to as the synthesis and analysis variant, respectively. In [3], let \( D : \mathbb{C}^P \rightarrow \mathbb{C}^N \), \( P \geq N \) 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 \( \Gamma \) of signals that are close to the reliable parts of the observed signal. The vector \( w \in \mathbb{R}^P_+ \) 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 ℓ1 minimization was presented. The idea of reweighting is that the restoration task is solved repeatedly, where each time, the ℓ1 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 ℓ0 (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 \( \Gamma \) is different compared to the declipping case.

As the second contribution, we include reweighting into the analysis variant [15], which was proposed in [5], but not presented in the

Algorithm 1: Synthesis reweighted ℓ1 audio inpainting

\[
\begin{align*}
&\text{require:} \quad D : \mathbb{C}^P \rightarrow \mathbb{C}^N, \Gamma \subset \mathbb{C}^N, K, \varepsilon, \delta > 0 \\
&1 \quad k = 1, \quad w_1^{(1)} = 1, \quad i = 1, \ldots, P \\
&2 \quad \text{repeat} \\
&3 \quad z^{(k)} = \arg\min_{z} \|w^{(k)} \odot z\|_1 \quad \text{s.t.} \quad Dz \in \Gamma \\
&4 \quad u_i^{(k)} = 1/(|z_i^{(k)}| + \varepsilon), \quad i = 1, \ldots, P \\
&5 \quad k \leftarrow k + 1 \\
&6 \quad \text{until} \quad k > K \quad \text{or} \quad \|z^{(k)} - z^{(k-1)}\|_2 < \delta \\
&7 \quad \text{return} \quad x = Dz^{(k-1)}
\end{align*}
\]

Algorithm 2: Analysis reweighted ℓ1 audio inpainting

\[
\begin{align*}
&\text{require:} \quad D : \mathbb{C}^P \rightarrow \mathbb{C}^N, A = D^*, \Gamma \subset \mathbb{C}^N, K, \varepsilon, \delta > 0 \\
&1 \quad k = 1, \quad w_1^{(1)} = 1, \quad i = 1, \ldots, P \\
&2 \quad \text{repeat} \\
&3 \quad x^{(k)} = \arg\min_{x} \|w^{(k)} \odot Ax\|_1 \quad \text{s.t.} \quad x \in \Gamma \\
&4 \quad z^{(k)} = Ax^{(k)} \\
&5 \quad u_i^{(k)} = 1/(|z_i^{(k)}| + \varepsilon), \quad i = 1, \ldots, P \\
&6 \quad k \leftarrow k + 1 \\
&7 \quad \text{until} \quad k > K \quad \text{or} \quad \|z^{(k)} - z^{(k-1)}\|_2 < \delta \\
&8 \quad \text{return} \quad x^{(k-1)}
\end{align*}
\]

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) [13]. 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 ℓ1 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.
TABLE I: Algorithm settings for the experiment

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>transform</td>
<td>DGT [13]</td>
</tr>
<tr>
<td>window type</td>
<td>Hann</td>
</tr>
<tr>
<td>window length</td>
<td>2800 (64 ms)</td>
</tr>
<tr>
<td>window shift</td>
<td>700 (16 ms)</td>
</tr>
<tr>
<td>frequency channels</td>
<td>2800</td>
</tr>
<tr>
<td>$K$</td>
<td>10</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>0.001</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Fig. 1: Values of SNR computed for each of the restored gaps and then averaged in dB. For the $\ell_1$ 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.

REFERENCES


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.