Dequantized signal from two parallel quantized observations

Vojtěch Kovanda
Dept. of telecommunications
Brno University of Technology
Czech Republic
xkovan07@vutbr.cz

***Abstract*—We propose a technique for signal acquisition that uses a combination of two devices with different sampling rates and quantization accuracies. Subsequent processing involving sparsity-based regularization enables us to reconstruct the input signal at such a sampling frequency and with such a bit depth that were not possible using the two devices independently. Objective tests show the evaluation of the proposed method in comparison with the alternatives.**

***Index Terms*—Dequantization, bit depth, multichannel, audio, optimization, sparsity, analog-to-digital conversion.**

I. Introduction

In analog-to-digital (A/D) conversion, two qualities play a crucial role. The sampling frequency determines how broad the signal spectrum is that can be acquired, while the number of bits used for representing signal samples governs their accuracy [1]. In any particular application, a combination of a sampling frequency and a quantization step that are adequate is required. For demanding applications, however, a proper combination of the parameters can imply a high sale price of the A/D conversion unit.

In this paper, we propose a system that overcomes the described property via employing two parallel signal acquisition branches. One branch consists of an A/D converter with a high sampling frequency but a coarse quantization. The second branch involves a significantly more accurate quantizer; nevertheless, it operates at a low sampling frequency. The scheme of the system is in Fig. 1. The goal is to reconstruct a signal as close to the original signal as possible, given the two different observations $\mathbf{y}_1, \mathbf{y}_2$.

If the proposed concept proves beneficial, it could allow employing cheaper components to provide an acquisition quality comparable to high-end devices. Seen from another side, the approach could even make possible acquisitions that would not be accessible with a single converter; in this sense, our approach can be understood as superresolution in time and/or in sample value domains. Naturally, estimating the original signal from $\mathbf{y}_1, \mathbf{y}_2$ is not straightforward and comes at the cost of computation.

Related work. Increasing the sampling frequency and the quantization resolution of an A/D converter has always been of interest to the signal processing community. The parallel branches in Fig. 1 do not represent an entirely new idea: As regards the sampling speed, the physical limit of a sampling device can be bypassed via involving multiple A/D converters that are time-interleaved. Such an idea is actually a special case of the so-called multirate filterbank A/D conversion concept [2], [3]. To increase the resolution in value (i.e., the final quantization accuracy) within the A/D conversion, a similar trick of an array of quantizers with different offsets in value can be applied [4]. In the converters of classical design, such as Σ–Δ, the resolution is exchanged with speed [1].

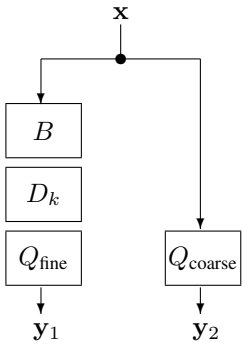

Fig. 1. Scheme of the parallel conversion. In one branch, the original signal $\mathbf{x}$ goes through an anti-aliasing filter B, then the subsampling operator D_k, and finally the quantization with a high-resolution Q_{fine}, producing the observation $\mathbf{y}_1$. The second branch does not alter the sampling frequency and only quantizes the signal with a low-resolution Q_{coarse}. Note that in practice, $\mathbf{x}$ would come in as an analog signal; in our treatment, $\mathbf{x}$ already is assumed in the discrete time.

In the mentioned approaches, no property of the analog signal is utilized, except its bandwidth. Additional signal characteristics can yet be exploited for the increase of sampling frequency or resolution, as a postprocessing step. This has been demonstrated in various signal processing fields, such as image superresolution [5], audio dequantization [6] or compressive sampling [7], [8]. An approach to increasing the sampling frequency of the A/D conversion has been presented in [9]. The authors showed that when the observed signal from a low-frequency A/D converter is understood as the subsampled version of a desired signal sampled at a high frequency, it can be estimated via optimization involving the signal sparsity

The work was supported by the Czech Science Foundation (GAČR) Project No. 23-07294S. I would like to thank prof. Mgr. Pavel Rajmic Ph.D. for supervising this work.

assumption in a proper representation system. A side effect of increasing the effective bit depth is even demonstrated, thanks to inherent oversampling. The dequantization of audio has also been studied using different prior assumptions [6], [10]–[13]. However, these methods only rely on a single channel, which is in contrast to our approach, which combines two parallel sources of quantized audio information to achieve dequantization. We are not aware of any other multichannel dequantization method.

II. Method

The proposed means of signal acquisition are shown in Fig. 1. The parameters affecting the observations $\mathbf{y}_1, \mathbf{y}_2$ are:

- The sampling frequency of the right-hand branch.
- The sampling frequency of the left-hand branch; in our study it is k-times lower, due to the downsampler denoted D_k. (Utilizing a non-synchronized downsampler might bring additional benefits, but our simulation is based on digitized signals at the input.)
- The bit depths of the quantizers Q_{fine} and Q_{coarse}.
- The properties of the low-pass filter B.

Due to the involvement of lossy components in the acquisition, the estimation of $\mathbf{x}$ back from $\mathbf{y}_1, \mathbf{y}_2$ is clearly an ill-posed problem. As such, a kind of regularization has to be introduced. As one of the options, we will make use of the sparsity of an audio signal in the time-frequency domain in this paper. Therefore, our recovery task can be written as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^L}{\arg\min} \quad \lambda\|A\mathbf{x}\|_1 + \imath_{\Gamma_{\text{fine}}}(D_k B\mathbf{x}) + \imath_{\Gamma_{\text{coarse}}}(\mathbf{x}). \qquad (1)$$

The operator A transforms a time-domain audio signal from $\mathbb{R}^L$ to the time-frequency domain $\mathbb{C}^P$ [14], and the sparsity of such a representation is quantified by the convex ℓ_1-norm $\|\cdot\|_1$ [15]. The functions $\imath_{\Gamma_{\text{fine}}}$ and $\imath_{\Gamma_{\text{coarse}}}$ are the indicator functions [16] taking value ∞ when $D_k B\mathbf{x} \notin \Gamma_{\text{fine}}$ and $\mathbf{x} \notin \Gamma_{\text{coarse}}$ enforcing the estimate to lie within the quantization levels corresponding to the operators Q_{fine} and Q_{coarse}, respectively. The indicator functions thus secure consistency of the solution with the observations, while the sparsity-related term promotes natural audio signals. Finally, the scalar $\lambda > 0$ is a weight which clearly could be omitted (i.e., $\lambda = 1$) in theory, but it can be used to influence the convergence of the practical numerical algorithm.

The problem (1) is convex and since it contains three terms, an advanced solver should be utilized. We make use of the Condat–Vũ algorithm (CVA) [17], [18], which utilizes proximal operators [16] corresponding to the functions involved in (1). The CVA for our problem is given in Alg. 1.

The asterisk denotes the adjoint of a linear operator; the adjoint of B is simply a filtering with the impulse response flipped in time. The operator $\text{clip}_\lambda(\cdot)$ clips the input vector elementwise such that the output samples reside in the interval $[-\lambda, \lambda]$.

As for the scalars τ, σ, the convergence of CVA is guaranteed if it holds $\tau\sigma\|A^*A + (D_k B)^*(D_k B) + Id^*Id\| \leq 1$. Using the properties of operator norms, we can arrive at a weaker (still sufficient) condition $\tau\sigma(2 + \|\mathbf{b}\|_1^2) \leq 1$, where $\|\mathbf{b}\|_1$ is the ℓ_1-norm of the impulse response corresponding to B. Also, we have utilized the assumption that A corresponds to a tight Parseval frame, which is achieved via a suitable selection of transfrom parameters [19], see below our particular choice. The parameter ρ has to satisfy $\rho \in]0, 2[$.

Algorithm 1: Condat–Vũ algorithm (CVA) solving (1)

Choose parameters $\tau, \sigma, \rho > 0$ and initial values $\mathbf{x}^{(0)} \in \mathbb{R}^L$, $\mathbf{u}_1^{(0)} \in \mathbb{C}^P$, $\mathbf{u}_2^{(0)} \in \mathbb{R}^{L/k}$, $\mathbf{u}_3^{(0)} \in \mathbb{R}^L$.

for $i = 0, 1, \ldots$ **do**

$\tilde{\mathbf{x}}^{(i+1)} = \mathbf{x}^{(i)} - \tau(A^*\mathbf{u}_1^{(i)} + B^* D_k^* \mathbf{u}_2^{(i)} + \mathbf{u}_3^{(i)})$

$\mathbf{x}^{(i+1)} = \rho\tilde{\mathbf{x}}^{(i+1)} + (1-\rho)\mathbf{x}^{(i)}$

$\tilde{\mathbf{u}}_1^{(i+1)} = \text{clip}_\lambda(\mathbf{u}_1^{(i)} + \sigma A(2\tilde{\mathbf{x}}^{(i+1)} - \mathbf{x}^{(i)}))$

$\mathbf{u}_1^{(i+1)} = \rho\tilde{\mathbf{u}}_1^{(i+1)} + (1-\rho)\mathbf{u}_1^{(i)}$

$\mathbf{p}_2 = \mathbf{u}_2^{(i)} + \sigma D_k B(2\tilde{\mathbf{x}}^{(i+1)} - \mathbf{x}^{(i)})$ % auxiliary

$\tilde{\mathbf{u}}_2^{(i+1)} = \mathbf{p}_2 - \sigma\,\text{proj}_{\Gamma_{\text{fine}}}(\mathbf{p}_2/\sigma)$

$\mathbf{u}_2^{(i+1)} = \rho\tilde{\mathbf{u}}_2^{(i+1)} + (1-\rho)\mathbf{u}_2^{(i)}$

$\mathbf{p}_3 = \mathbf{u}_3^{(i)} + \sigma(2\tilde{\mathbf{x}}^{(i+1)} - \mathbf{x}^{(i)})$ % auxiliary

$\tilde{\mathbf{u}}_3^{(i+1)} = \mathbf{p}_3 - \sigma\,\text{proj}_{\Gamma_{\text{coarse}}}(\mathbf{p}_3/\sigma)$

$\mathbf{u}_3^{(i+1)} = \rho\tilde{\mathbf{u}}_3^{(i+1)} + (1-\rho)\mathbf{u}_3^{(i)}$

end

III. Experiment

For the numerical experiment, we selected recordings of solo instruments. Such signals exhibit a great time-frequency sparsity, which the proposed reconstruction is regularized with. We used 83 excerpts selected from the Good-sounds dataset[1] containing various instruments, each playing several musical scales, such as violin, flute, saxophone, and clarinet. All audio is originally sampled at 48 kHz, in the 24-bit resolution (further on, we use the abbreviation bps for 'bits per sample'). The excerpts are between 10 and 20 seconds long. Audio has been peak-normalized before any processing to make the most of the available dynamic range.

Regarding the acquisition channels (see Fig. 1), Q_{coarse} operates at a bit depth varying between 4 and 16 bps, while the bit depth of Q_{fine} ranges between 10 and 24 bps. The quantization is uniform (linear PCM), as is typical in audio [1]. We use the mid-riser distribution of quantization levels, in line with [6]. In all experiments, the downsampling factor $k = 4$ is fixed, as is the low pass filter B, which has been designed by the Matlab Filter Designer as an FIR filter using the equiripple method. Note that for the purpose of the reconstruction, however, the properties of B are not crucial; actually, the proposed system would be applicable even if the filter were not present. The parameters of the operator A are: the 2048-sample-long Hann window, the window shift of 512 samples, and 4096 frequency channels. The Algorithm 1 used $\tau = 1$, $\sigma = 0.5$ and $\rho = 0.8$.

[1] https://www.upf.edu/web/mtg/good-sounds

A. Objective evaluation

We measure the performance of the algorithms using objective metrics: the common signal-to-distortion ratio (SDR) and the objective difference grade (ODG) provided by the PEMO-Q computational psychoacoustic model [20]. The ODG scale ranges from -4 to 0 (worst to best). These metrics are used for comparing our estimate $\hat{\mathbf{x}}$ with reference signal $\mathbf{x}$. We also evaluate SDR and ODG of the quantized signal observation $\mathbf{x}_\mathrm{Q}$ and an estimate given by a different dequantization method compared with the reference signal $\mathbf{x}$ to judge the effectiveness of our method. The estimate we are comparing with is obtained by using the Chambolle–Pock algorithm (CPA) considering only single channel observation $\mathbf{x}_\mathrm{Q}$; the CPA was taken from [6]. We compare the results with the same bits per sample.

The results are presented in Fig. 2 and 3 in a condensed way. Instructions on how to read the graphs are in the captions.

As an example, take the case of 4 bps of the $\mathbf{y}_2$ branch and 16 bps of the $\mathbf{y}_1$ branch. As the $\mathbf{y}_1$ branch is downsampled by a factor of 4, the effective number of bits of this combination is $4\,\mathrm{bps} + \frac{1}{4} \cdot 16\,\mathrm{bps} = 8\,\mathrm{bps}$. Our reconstruction provides an average SDR of 43.09 dB which we compare with the reconstruction given by the CPA (i.e., CP($\mathbf{x}_\mathrm{Q}$)) using the same number of effective bits resulting in an average SDR of 37.02 dB. From the graph, we can see that the 8-bit $\mathbf{x}_\mathrm{Q}$ without processing provides an average SDR of 29.25 dB. The same combination of bit depths but this time in terms of the ODG scale, we see that our method provides an average ODG of −2.70. Using the CPA with the same bps increases the ODG of $\mathbf{x}_\mathrm{Q}$ from −3.18 to −2.51. On the other hand, combinations with a higher effective number of bits in our reconstruction, for example 14 bps, are not better than using CPA, as we can see from the graph.

B. Computational considerations

Algorithm 1 was run for 200 iterations. With regard to the SDR plots, the maximum SDR achieved within these 200 iterations is presented. In particular cases, a small additional SDR improvement can be obtained by running more than 200 iterations; nevertheless, increasing the iteration count does not bring an improvement in most cases; typically, it is the other way round. Such an interesting effect is due to the presence of the ℓ_1 norm in (1): After reaching a good SDR within the constraints given by Γ_fine and Γ_coarse, the ℓ_1 norm starts to prevail, and as a lower $\|A\mathbf{x}\|_1$ is promoted, the signal is pushed towards lower quantization decision levels, therefore worsening the SDR [6]. The same effect is observable in the strip corresponding to the CPA in Fig. 2, since the CPA utilizes the ℓ_1 norm as well. The described effect is nevertheless not present in the ODG scores, and no significant improvement typically occurred with a greater number of iterations. Thus, the ODG value was computed based on the iteration no. 200.

Note that in Alg. 1, the parameter λ appears solely in connection with the clip operator. For each particular combination of the bit depths of the quantizers, a different choice of λ is advantageous as it speeds up the convergence and yields better results. Finding a right λ requires hand-on tuning, but we took advantage of values published in [6] as the starting point.

For a 19-second-long excerpt, a single iteration of the CVA takes about 0.18 second on a PC with Intel 3.6 GHz CPU and 64 GB RAM. Thus, the excerpt was reconstructed in about 40 seconds, i.e., twice the realtime.

IV. Conclusion

We have shown the possibility of reconstructing the desired signal from two parallel acquisition branches. Our algorithm scored well in both objective and subjective tests. In the future, testing at higher sampling rates is necessary.

The advantage of our algorithm further lies in the fact that by combining common converters with bit depths such as 4, 8, or 16 bps, we are able to achieve signal quality that would correspond to a less common converter with a bit depth of, for example, 9 bps.

Further extensions come naturally. For example, our approach exploits the simplest form of sparsity available; one could think about employing advanced signal priors such as the social sparsity [21] or the phase-consistency [22]. In such cases, only parts of the CVA algorithm would change. Generalization to nonuniform quantization would be straightforward. We have demonstrated the concept in the field of audio; however, the concept is general enough to be translated to image or video processing fields, with suitable regularizers. Finally, the concept allows increasing the sampling frequency beyond the frequency physically available in a device [9], i.e., superresolution.

Codes for Matlab are publicly available.[2]

[2] https://github.com/rajmic/parallel-dequantization

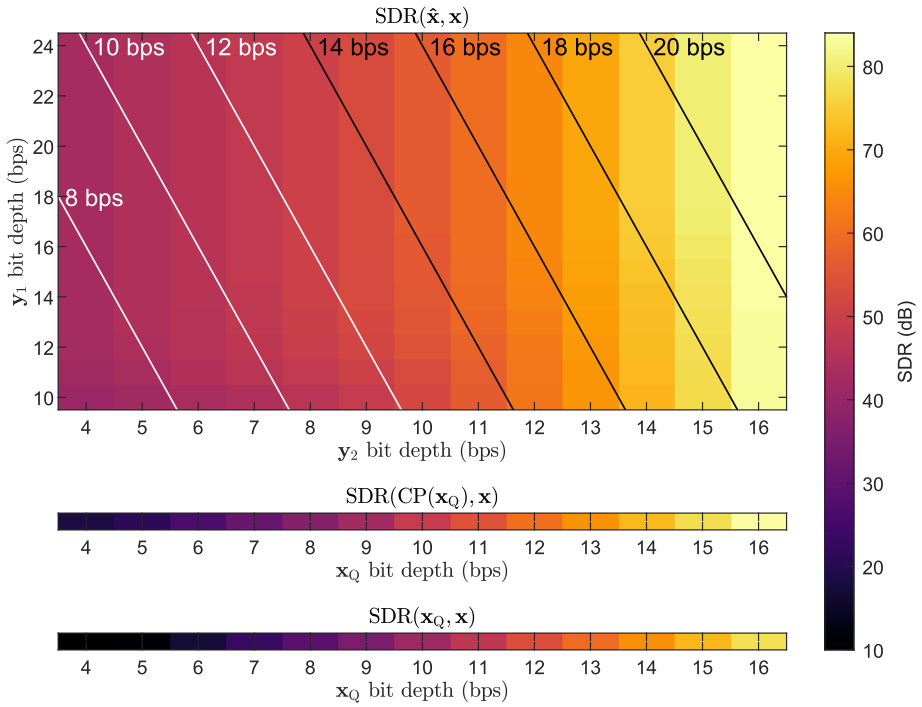

Fig. 2. Dequantization results in terms of the SDR; averages across test signals are presented. Each field in the main grid corresponds to a combination of bit depths in the two branches. The lines in the graph represent the combinations with the same number of effective bits. The colors code the SDR value of our reconstruction $\hat{\mathbf{x}}$ related to the original signal $\mathbf{x}$ (a lighter color indcates a closer estimate). The horizontal strip just below the main grid shows the SDR of the single channel reconstruction, CP($\mathbf{x}_Q$). The bottom strip presents the SDR of $\mathbf{x}_Q$ without postprocessing.

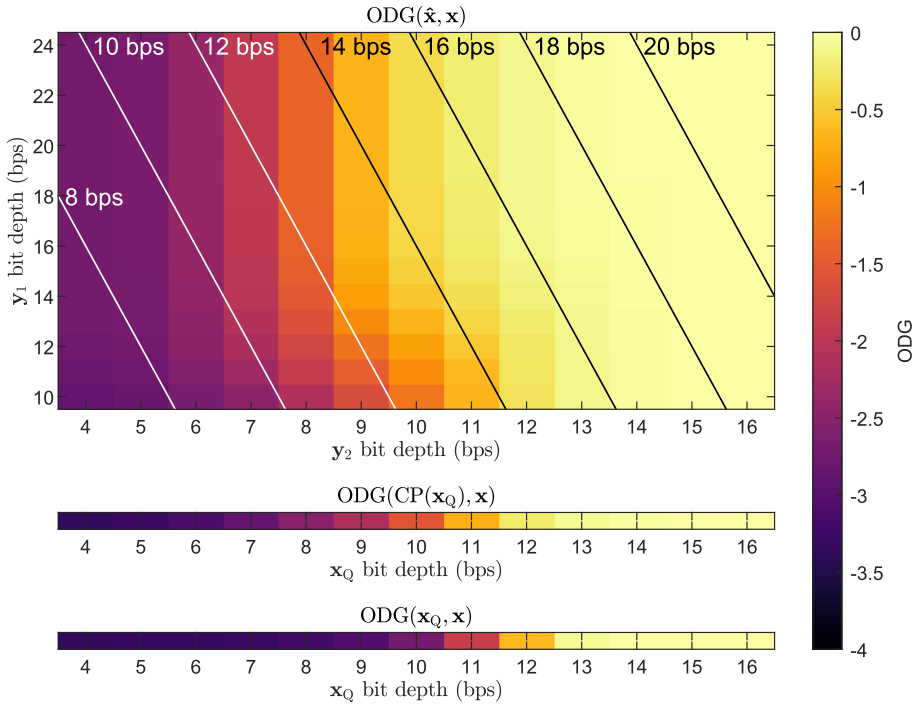

Fig. 3. Dequantization results in terms of the ODG, presented in analogy with Fig. 2.

REFERENCES

[1] J. Watkinson, *The Art of Digital Audio*, Focal Press, 2001.

[2] A. Petraglia and M.A.A. Pinheiro, "Effects of quantization noise in parallel arrays of analog-to-digital converters," in *Proceedings of IEEE International Symposium on Circuits and Systems – ISCAS '94*, 1994, vol. 5, pp. 337–340.

[3] P. Lowenborg and H. Johansson, "Quantization noise in filter bank analog-to-digital converters," in *The 2001 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2001, vol. 2, pp. 601–604.

[4] Jian Gao, Peng Ye, Hao Zeng, Zhixiang Pan, Yu Zhao, Hao Li, and Jie Meng, "Theory of quantization-interleaving adc and its application in high-resolution oscilloscope," *IEEE Access*, vol. 7, pp. 156722–156732, 2019.

[5] Filip Šroubek, Jan Flusser, and Gabriel Cristóbal, "Super-resolution and blind deconvolution for rational factors with an application to color images," *The Computer Journal*, vol. 52, no. 1, pp. 142–152, 2009.

[6] Pavel Záviška, Pavel Rajmic, and Ondřej Mokrý, "Audio dequantization using (co)sparse (non)convex methods," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, 2021, pp. 701–705.

[7] Emmanuel J. Candes and Michael B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[8] Marie Mangová, *Increasing Resolution in Perfusion Magnetic Resonance Imaging Using Compressed Sensing*, Ph.d. thesis, Brno University of Technology, 2018.

[9] Aldo Baccigalupi, Mauro D'Arco, Annalisa Liccardo, and Rosario Schiano Lo Moriello, "Compressive sampling-based strategy for enhancing ADCs resolution," *Measurement*, vol. 56, pp. 95–103, 2014.

[10] C. Brauer, Z. Zhao, D. Lorenz, and T. Fingscheidt, "Learning to dequantize speech signals by primal-dual networks: an approach for acoustic sensor networks," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 7000–7004.

[11] Lucas Rencker, Francis Bach, Wenwu Wang, and Mark D. Plumbley, "Sparse recovery and dictionary learning from nonlinear compressive measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5659–5670, Nov. 2019.

[12] P. T. Troughton, "Bayesian restoration of quantised audio signals using a sinusoidal model with autoregressive residuals," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No.99TH8452)*, Oct. 1999, pp. 159–162.

[13] Hyun-Wook Yoon, Sang-Hoon Lee, Hyeong-Rae Noh, and Seong-Whan Lee, "Audio dequantization for high fidelity audio generation in flow-based neural vocoder," in *Proc. Interspeech 2020*, Shanghai, China, Oct. 2020, pp. 3545–3549.

[14] Karlheinz Gröchenig, *Foundations of time-frequency analysis*, Birkhäuser, 2001.

[15] David L. Donoho and Michael Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of The National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[16] Patrick L. Combettes and Jean-Christophe Pesquet, "Proximal splitting methods in signal processing," *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, vol. 49, pp. 185–212, 2011.

[17] Laurent Condat, "A generic proximal algorithm for convex optimization—application to total variation minimization," *Signal Processing Letters, IEEE*, vol. 21, no. 8, pp. 985–989, Aug. 2014.

[18] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

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

[20] R. Huber and B. Kollmeier, "PEMO-Q—A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans. Audio Speech Language Proc.*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.

[21] M. Kowalski, K. Siedenburg, and M. Dörfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2498–2511, 2013.

[22] Tomoro Tanaka, Kohei Yatabe, and Yasuhiro Oikawa, "PHAIN: Audio inpainting via phase-aware optimization with instantaneous frequency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 4471–4485, 2024.