Audio Dequantization Using (Co)Sparse (Non)Convex Methods

Pavel Záviška*, Pavel Rajmic*, Ondřej Mokrý*

*Signal Processing Laboratory, Brno University of Technology, Brno, Czech Republic

Paper Number

5279

Abstract

The paper deals with the hitherto neglected topic of audio dequantization. It reviews the state-of-the-art sparsity-based approaches and proposes several new methods. Convex as well as non-convex approaches are included, and all the presented formulations come in both the synthesis and analysis variants. In the experiments the methods are evaluated using the signal-to-distortion ratio (SDR) and PEMO-Q, a perceptually motivated metric.

Introduction

- Quantization
  - Nonlinear limitation of signal values.
  - Necessary step in signal digitization.
  - Number of quantization levels, word length w (bps).
  - Mid-riser uniform quantization

$$(x^{\mathrm{q}})_n = \operatorname{sgn}^+(x_n)\Delta\left(\left\lfloor\frac{|x_n|}{\Delta}\right\rfloor+\frac{1}{2}\right),$$

where $\Delta = 2^{-w+1}$ is the quantization step.

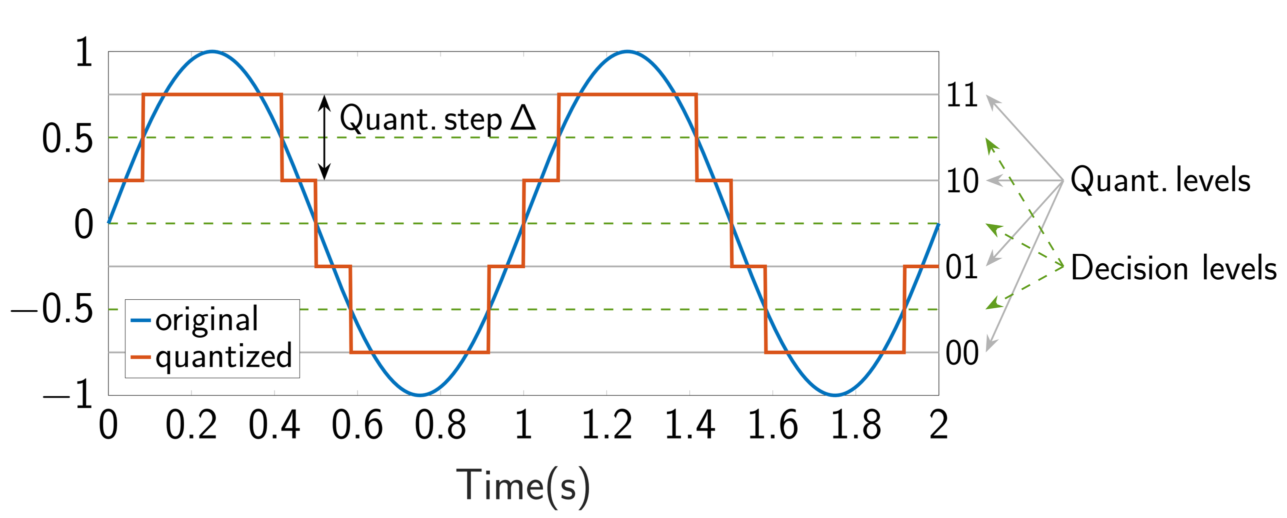


Figure 1: Demonstration of mid-riser quantization.

- Dequantization
  - Inverse problem to quantization.
  - Restore the quantized observation to be as close to the original signal as possible.
  - Ill-posed without additional knowledge.
  - Assumption of sparsity w.r.t. STFT.
- Motivation
  - Enhance standard 16-bit audio.
  - Restore audio in special cases, where less than standard bit depth had to be used (scenario for the paper).
  - Enhance audio generated by Flow-based Neural Vocoder.
  - The goal is not to compete with current lossy compression standards.

Problem formulations

- Consistent ℓ_1 minimization

$$\operatorname*{arg\,min}_{\mathbf{c}\in\mathbb{C}^Q} \|\mathbf{c}\|_1 \quad \text{s.t.} \quad A^*\mathbf{c}\in\Gamma \qquad (1a)$$

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^P} \|A\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{x}\in\Gamma \qquad (1b)$$

- Inconsistent ℓ_1 minimization

$$\operatorname*{arg\,min}_{\mathbf{c}\in\mathbb{C}^Q} \lambda\|\mathbf{c}\|_1 + \frac{1}{2}d_\Gamma^2(A^*\mathbf{c}) \qquad (2a)$$

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^P} \lambda\|A\mathbf{x}\|_1 + \frac{1}{2}d_\Gamma^2(\mathbf{x}) \qquad (2b)$$

- Nonconvex ℓ_0 approximation

$$\operatorname*{arg\,min}_{\mathbf{c},\mathbf{z}\in\mathbb{C}^Q} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad A^*\mathbf{c}\in\Gamma,\ \|\mathbf{c}-\mathbf{z}\|_2\leq\epsilon \qquad (3a)$$

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^P,\mathbf{c}\in\mathbb{C}^Q} \|\mathbf{c}\|_0 \quad \text{s.t.} \quad \mathbf{x}\in\Gamma,\ \|\mathbf{x}-A^*\mathbf{c}\|_2\leq\epsilon \qquad (3b)$$

$$\operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^P,\mathbf{c}\in\mathbb{C}^Q} \|\mathbf{c}\|_0 \quad \text{s.t.} \quad \mathbf{x}\in\Gamma,\ \|A\mathbf{x}-\mathbf{c}\|_2\leq\epsilon \qquad (3c)$$

- Set of feasible solutions

$$\Gamma = \{\mathbf{x}\in\mathbb{R}^P \mid \|\mathbf{x}-\mathbf{x}^{\mathrm{q}}\|_\infty < \Delta/2\}$$

$\mathbf{c},\mathbf{z}\in\mathbb{C}^Q$	signal coefficients,
$\mathbf{x}\in\mathbb{R}^P$	signal in time domain,
$A\colon\mathbb{R}^P\to\mathbb{C}^Q$	analysis operator,
$d_\Gamma(\cdot)$	distance from the set Γ.

Experiments

- Audio database:
  - 10 musical audio excerpts,
  - approximate length 7 seconds,
  - sampling rate 44.1 kHz,
  - bit-depth 16 bps.
- Quantized to 7 different levels
  - $w = 2, 3, \dots, 8$ bps,
  - Mid-riser quantization.
- Signals restored using algorithms based on sparsity.
- Discrete Gabor Transform (DGT/STFT),
  - 8192 samples long Hann window,
  - 75% overlap.
- Evaluation
  - Signal-to-distortion ratio improvement (ΔSDR),
  - PEMO-Q ODG (perceptually motivated metric)

  0.0 Imperceptible,
  −1.0 Perceptible, but not annoying,
  −2.0 Slightly annoying,
  −3.0 Annoying,
  −4.0 Very annoying.

- Implementation available on GitHub:

  github.com/zawi01/audio_dequantization

Results

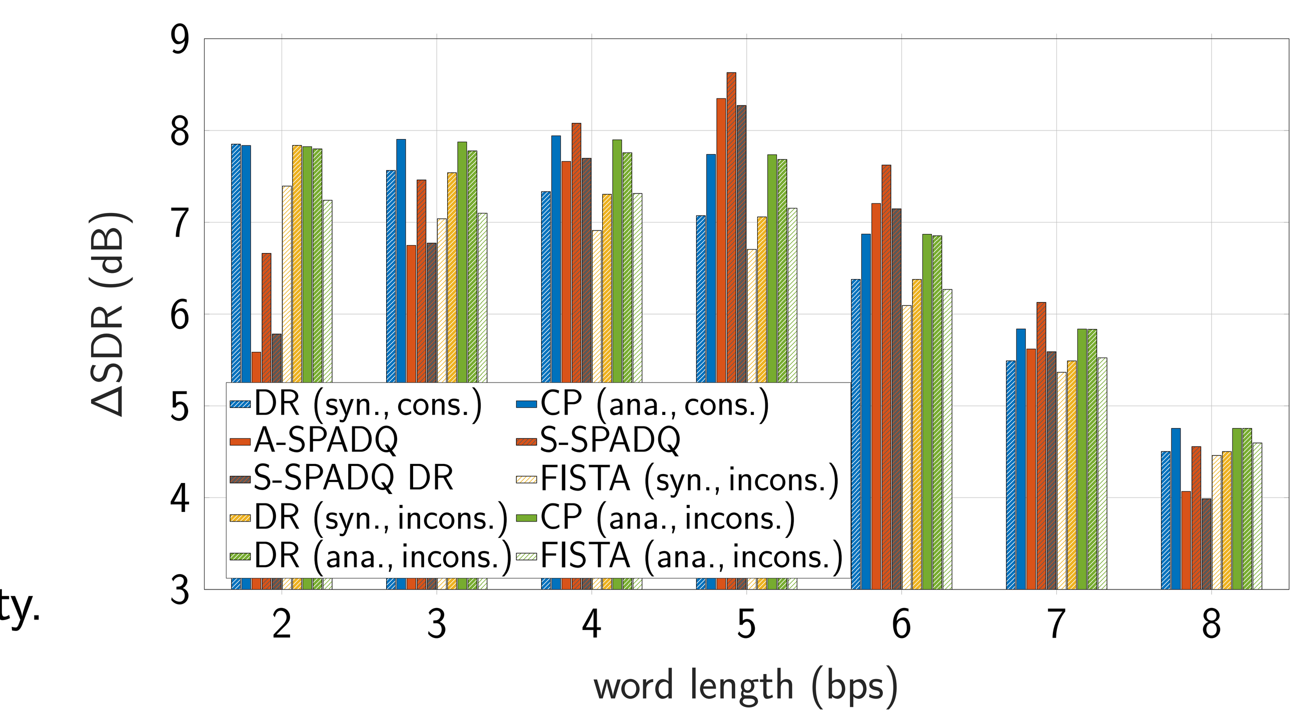


Figure 2: Average performance in terms of ΔSDR.

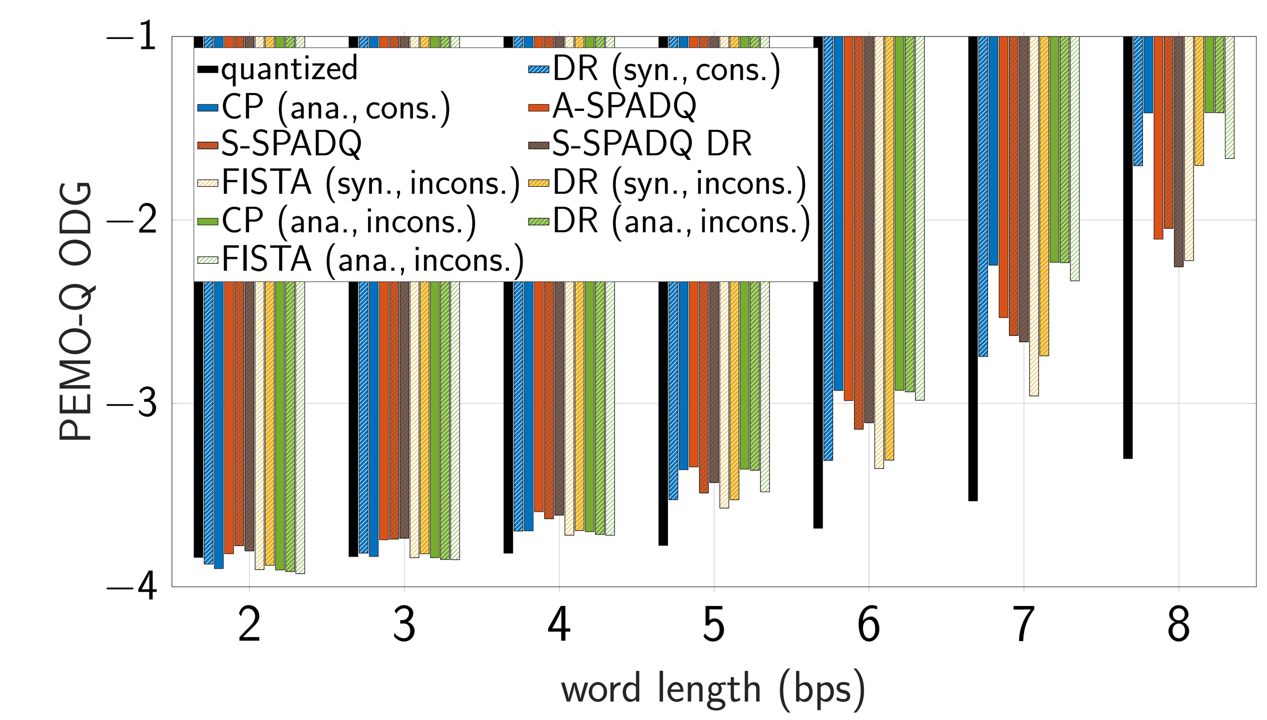
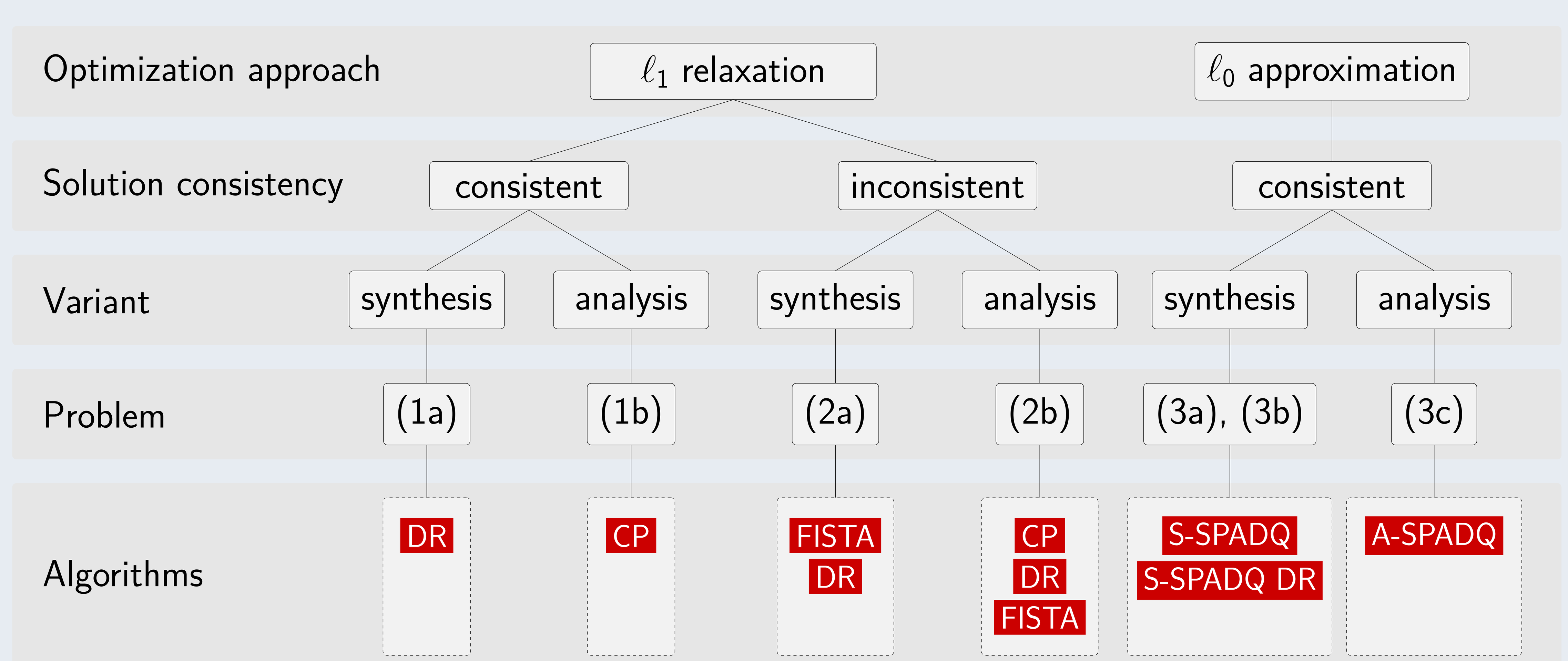


Figure 3: Average performance in terms of PEMO-Q ODG.

Overview of algorithms



Conclusion

- 10 sparsity-based approaches to audio dequantization.
  - Convex ℓ_1 relaxation, nonconvex ℓ_0 approximation.
  - Strict or only approximate compliance of the solution consistency.
  - Synthesis and analysis model.
- All methods improve the quality of the signal.
- No clear winner of all presented methods.
- Analysis model seems to outperform its synthesis counterpart.