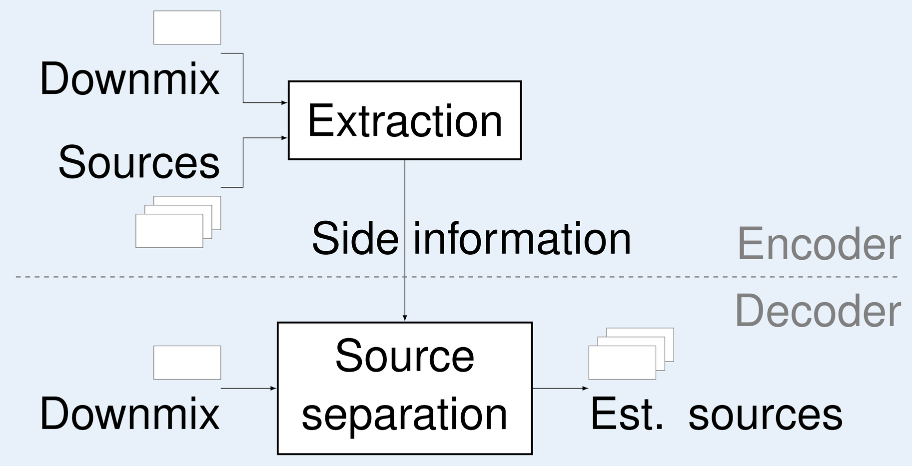
QUANTIZATION-AWARE PARAMETER ESTIMATION FOR AUDIO UPMIXING

Christian Rohlfing[1], Antoine Liutkus[2], Julian M. Becker[1]

[1]Institut für Nachrichtentechnik, RWTH Aachen University
[2]Inria, speech processing team, Villers-lès-Nancy, France

Introduction



Audio upmixing

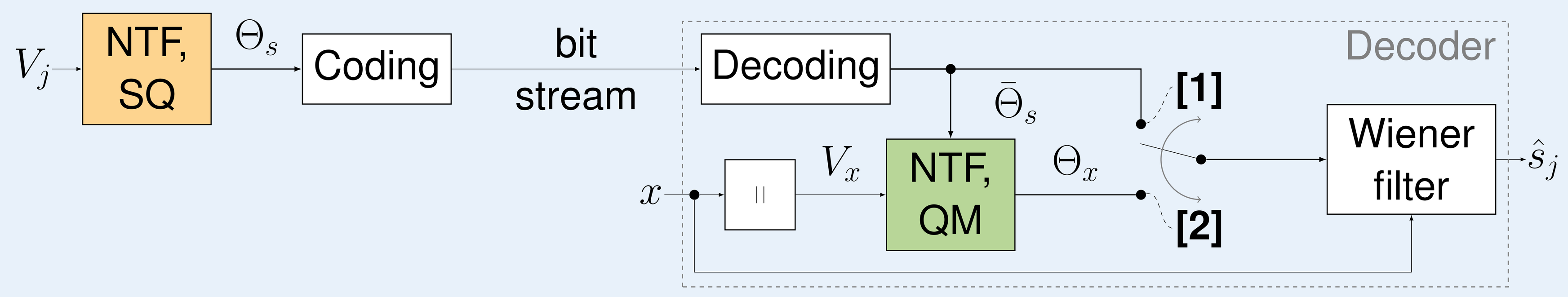
- aims at generating a multichannel signal based on a downmix
- is research topic in both audio coding and audio source separation communities
- enables numerous applications, such as adaptive rendering on loudspeakers arrays, karaoke or active listening.

Parametric Audio Upmixing

Parametric audio upmixing consists of two steps:

- **Encoder**: Sources and downmix perfectly known. Prior work **[1]**, **[2]** use
  - Nonnegative Tensor Factorization on source spectrogram V_j which computes parameters Θ_s
  - *subsequent* scalar quantization of parameters $\bar{\Theta}_s$.
- **Decoder**: Only downmix x available.
  - **[1]**: Sources are estimated with Wiener-filtering given downmix and $\bar{\Theta}_s$.
  - **[2]**: Additional NTF with mix spectrogram V_x as observation. Resulting parameters Θ_x used for Wiener-filtering. Initialization of NTF with $\bar{\Theta}_s$.

Here: Consider quantization of parameters already in design of their (re-)estimation strategy.



Self-quantization at Encoder

[1], **[2]**: NTF on sources and quantization of resulting parameters as two *independent* steps.

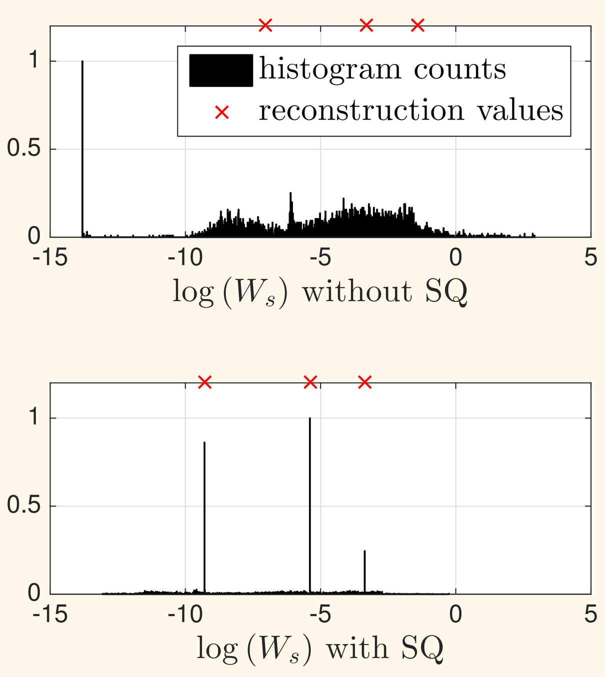
Self-quantization NTF:

Combination of signal approximation and quantization of parameters at run time.

$$\min \; d_\beta\left(V_j \mid \Theta_s\right) + \gamma_{\text{sq}}\left[d_\beta\left(\bar{W}_s \mid W_s\right) + d_\beta\left(\bar{H}_s \mid H_s\right)\right]$$

with $\bar{W}_s \triangleq \exp\left(q\left(\log W_s\right)\right)$, $\bar{H}_s \triangleq \exp\left(q\left(\log H_s\right)\right)$

Quantization $q\left(x\right)$ with Lloyd's algorithm.



Quantized-matching at Decoder

[2]: Decoder NTF is initialized with (coarsely) quantized $\bar{\Theta}_s$. Θ_x may deviate from target Θ_s too much.

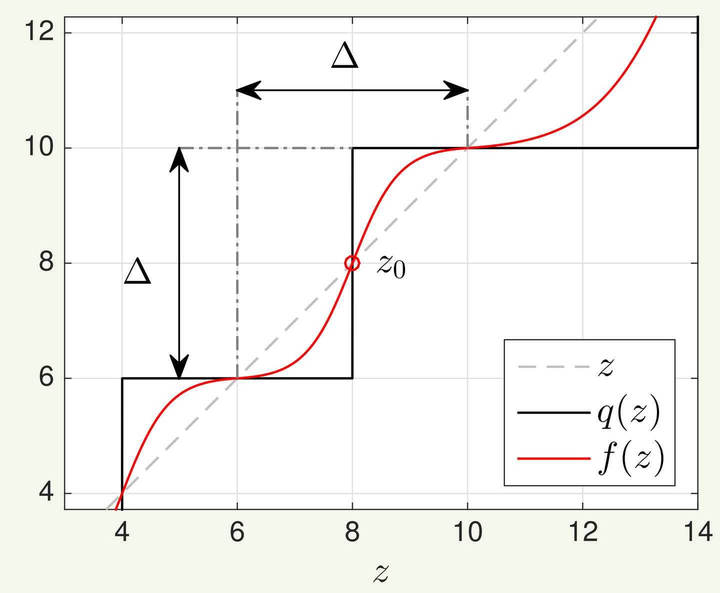
Quantized-matching NTF:

Estimate sources parameters from mix only, constraining them to have same quantization as transmitted values $\bar{\Theta}_s$.

Quantized version of Θ_x shall match $\bar{\Theta}_s$ as much as possible.

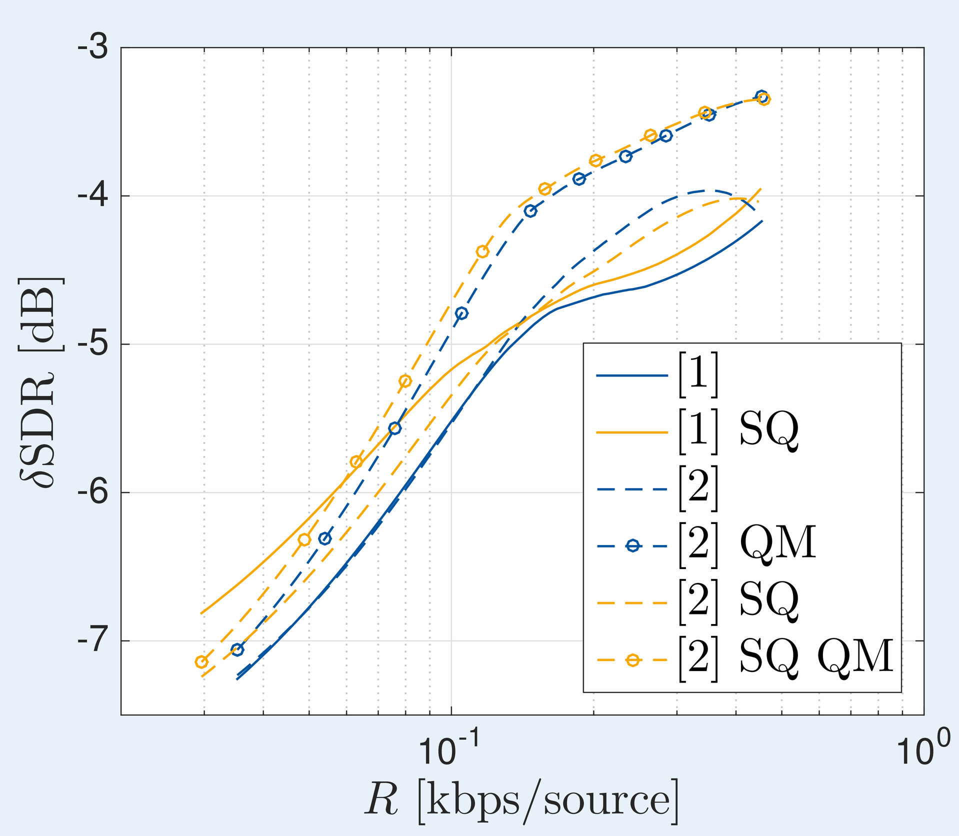
$$\min \; d_\beta\left(V_x \mid \Theta_x\right) + \gamma_{\text{qm}}\left[d_\beta\left(\bar{W}_s \mid \tilde{W}_x\right) + d_\beta\left(\bar{H}_s \mid \tilde{H}_x\right)\right]$$

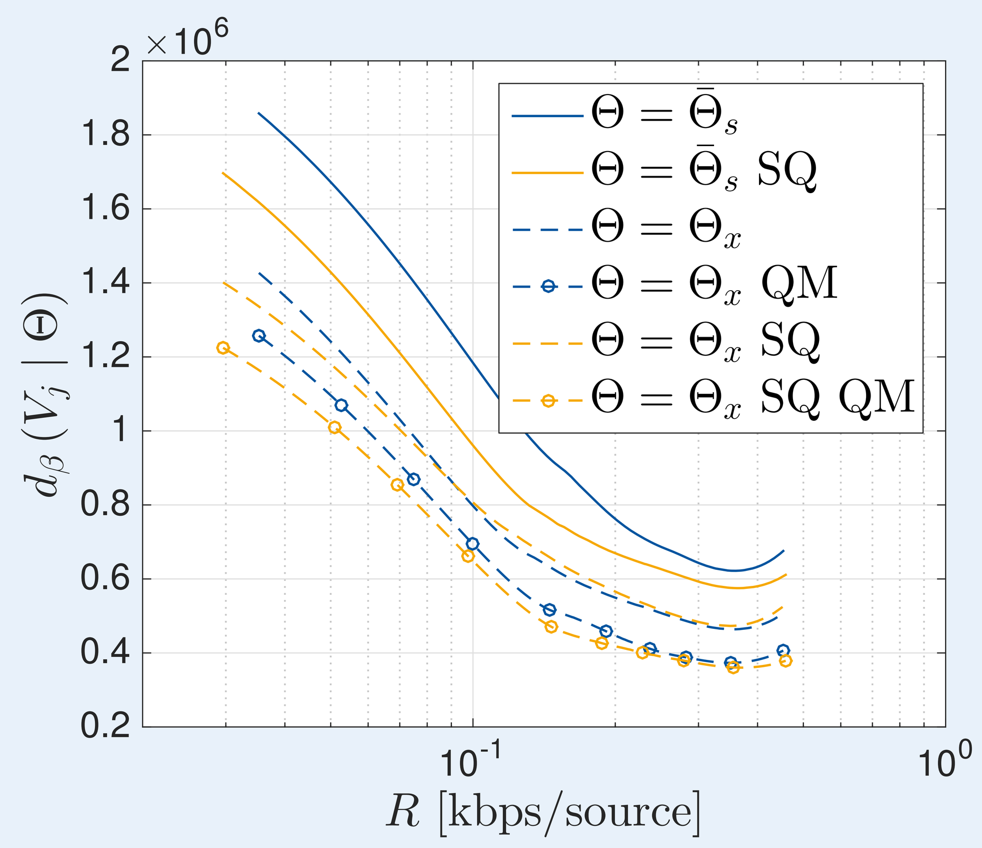
with $\tilde{W}_x \triangleq \exp\left(f\left(\log W_x\right)\right)$, $\tilde{H}_x \triangleq \exp\left(f\left(\log H_x\right)\right)$



Differentiable approximation of quantization

Experiments





Setup:

Ten mixtures (4 to 7 sources, taken from QUASI database [3]). NTF minimizes Kullback-Leibler divergence ($\beta = 1$) with number of components per source $\in \{1, 2 \ldots 10\}$ and weights $\gamma_{\text{sq}}, \gamma_{\text{qm}} \in \left\{10^{-1}, 1, 10, 10^2\right\}$.

Results:

- Self-quantizing constraint (SQ) leads simultaneously to good signal approximation and parameter quantization.
- SQ not limited to audio upmixing and may be useful whenever NTF is used for signal compression.
- Quantized-matching constraint (QM) prevents deviations from the source NTF parameters in the quantization domain.

[1] A. Liutkus, J. Pinel, R. Badeau, L. Girin, and G. Richard, "Informed source separation through spectrogram coding and data embedding," *Signal Processing*, vol. 92, no. 8, pp. 1937 – 1949, 2012.

[2] C. Rohlfing, J. M. Becker, and M. Wien, "NMF-based informed source separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 474 – 478.

[3] QUASI Database – A musical audio signal Database for Source Separation. `http://www.tsi.telecom-paristech.fr/aao/en/2012/03/12/quasi/`

rohlfing@ient.rwth-aachen.de

www.ient.rwth-aachen.de

Institut für Nachrichtentechnik, Melatener Str. 23, 52074 Aachen

ICASSP 2017