UNIVERSITY OF CAMBRIDGE

SEQUENTIAL MCMC METHODS FOR AUDIO SIGNAL ENHANCEMENT

RUBÉN CLAVERÍA, SIMON J. GODSILL
SIGNAL PROCESSING AND COMMUNICATIONS LABORATORY, DEPARTMENT OF ENGINEERING, UNIVERSITY OF CAMBRIDGE

icassp 2022 Singapore

SIGNAL ENHANCEMENT

Reconstructing a signal from imperfect observations is an ubiquitous problem in engineering. Sources of degradation are multiple: background noise, missing data, non-linear distortion (quantization, clipping), occlusion (image processing), to mention a few. Diverse tasks of unrelated nature can be framed as signal restoration problems: biomedical imaging, remote sensing, speech and video enhancement in streaming platforms, etc. Furthermore, signal restoration may be a crucial step in higher level problems—e.g. noise reduction may facilitate tasks source separation.

GENERAL FRAMEWORK

- Bayesian probability can describe generative models with an almost arbitrary level of detail, however its computation may be complicated.
- Gabor regression [2] is a dictionary representation endowed with structured sparsity and heavy-tailed priors on the synthesis coefficients, making it particularly suitable for audio signals.
- These priors reflect domain-specific traits that simpler optimization-based techniques ignore (e.g. as time-frequency persistence). However, inference is expensive and data is processed in a batch fashion.

CONTRIBUTION

- We formulate a state-space model that preserves the desirable traits of Gabor regression.
- Unlike original Gabor regression, the proposed model is suitable for sequential inference: A **sequential MCMC** strategy [1] is devised to estimate the filtering distributions $p(z_t|x_{1:t})$ of the synthesis coefficients, which in turn are used to produce a denoised version of the degraded input signal.
- Since sequential MCMC relies on a local analysis of the signal, the scheme has potential for real time applications.

MODEL DESCRIPTION

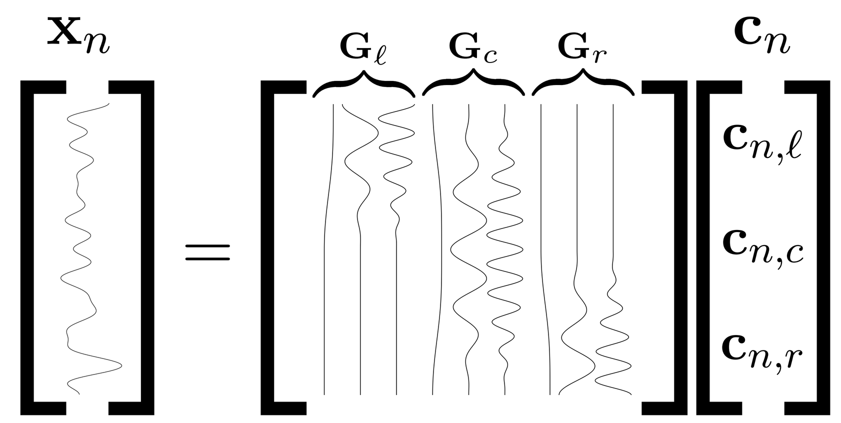
Signal represented as weighted sums of atoms:

$$x(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{Q-1} c_{m,q} \cdot g_{m,q}(t),$$

where $g_{m,q}(t)$ corresponds to time- and frequency-shifted replicas of a window function (e.g., Hann). As audio signals are normal sparse and highly structured in the spectral domain, Gabor regression imposes a spike-slab-prior on coefficients $c_{m,q}$: each $c_{m,q}$ has an indicator $\gamma_{m,q}$ that determines whether it is 0 or non-zero. Morover, a Markov chain prior is placed along the time axis n of indicators γ—this imitates the persistence of tonal components. When $c_{m,q}$ is non-zero, it is assumed to be drawn from a heavy-tailed distribution, which reflects the wide range of time-frequency coefficients in real audio signals.

A state-space model is formulated by dividing the signal x in N disjoint segments x_n and associating a unique set of coefficients to each x_n. Thus x_n is treated as an observation and $c_n = [c_{n,\ell}, c_{n,c}, c_{n,r}]$ as its underlying state (note that they represent the left, center and right set of coefficients). To account for the fact that coefficients "in between" contribute to two different chunks x_n, we impose the restriction $p(c_{n,\ell}|c_{n-1,r}) = \delta(c_{n,\ell} - c_{n-1,r})$.

Global parameters such as noise level σ^2 are replaced by slowly-varying sequences $\sigma^2_{1:n}$.



ALGORITHM

Since the resulting model has a state-space structure, it is suitable for sequential inference. Due to the high dimensionality of the problem and the poor predictive quality of the prior (without incorporating the observations x_n, the coefficients have mean 0), standard Particle filters (e.g. bootstrap) are not a promising alternative. We instead resort to the sequential MCMC. Defining our state z_n as a vector containing all the parameters associated to a frame n (coefficient, indicators, noise level), we exploit the following factorization:

$$p(z_{n-1}, z_n|x_{1:n}) \propto p(z_{n-1}|x_{1:n-1})p(z_n|z_{n-1})p(x_n|z_n)$$

to generate samples from $p(z_{n-1}, z_n|x_{1:n})$ (marginal $p(z_n|x_{1:n})$ is obtained as a by-product). Given a sample-based approximation of the previous filtering distribution $\hat{p}(z_{n-1}|x_{1:n-1}) = \frac{1}{P}\sum_{p=1}^{P} \delta(z_{n-1} - z_{n-1}^{(p)})$ and using (i) as the iteration index, sampling is carried out performing following updates repeatedly (for each frame n):

1. $z_n^{(i+1)} \sim p(z_n|z_{n-1}^{(i)}, x_{1:n})$
2. $z_{n-1}^{(i+1)} \sim p(z_{n-1}|z_n^{(i+1)}, x_{1:n})$

Exploiting the conjugacy of the chosen priors, all parameters can be sampled from a tractable conditional distribution. To incorporate the information contained in the following observation x_{n+1} and improve the quality of the estimation, at time step $n+1$ we refresh previous state z_n by splitting it in two subsets: some components (e.g. the rightmost coefficients $c_{n,r}$) are sampled from their analytical Gibbs updates, whereas the remaining components are simply sampled from the discrete collection of particles of the approximate filtering distribution $\hat{p}(z_n|x_{1:n})$. Note that the resulting scheme still targets $p(z_{n-1}, z_n|x_{1:n})$.

EXPERIMENTS

Two experiments are performed: 1. As a proof of concept, we first investigate the effect of replacing global parameters such as noise σ^2 with sequences $\sigma^2_{1:n}$. Secondly, we test the sequential MCMC against the batch scheme in a for different excerpts (A: glockenspiel, B: jazz trumpet, C: vibraphone music, D: oratorio).

RESULTS AND DISCUSSION

Example (label)	SNR_{in} [dB]	SNR_{out} [dB] (Batch)	SNR_{out} [dB] (Sequential)
A (~3 s)	10	20.13	19.29
B (~12 s)	15	21.46	20.03
C (~8.5 s)	15	25.05	23.18
C (~8.5 s)	20	28.79	26.96
D (~8.5 s)	15	21.45	19.84

Audible results in https://rclaveria.github.io/sMCMC_audio

The SNR gap is the combined effect of partial information (we use filtering distributions rather than complete trajectories) and a mismodelling induced by the substitution σ^2 by $\sigma^2_{1:n}$. This trade-off is arguably minor, as this modification is what facilitates sequential processing.
The proposed model (+ Sequential MCMC) needs fewer iterations to converge. Reconstructions using the original model (+ batch MCMC) contain noticeable artifacts when the number of iterations is limited to 80.
The perceptual gap is less noticeable than the SNR gap: the two reconstructions are often indistinguishable to the ear.
Scheme has potential for real time applications, provided that an efficient implementation is available.

REFERENCES

References

[1] F. Septier, S. K. Pang, A. Carmi, and S. Godsill. On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking. In *IEEE (CAMSAP)*, pages 360–363. IEEE, 2009.

[2] P. J. Wolfe, S. J. Godsill, and W.-J. Ng. Bayesian variable selection and regularization for time–frequency surface estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):575–589, 2004.