Rewriting the Attack-Decay model

as a Convolutive NMF with large convolution filter size and rank-two patterns.

Below we detail how to manipulate the Attack-Decay original model formulation to make the link with CNMF clearer.

First we start from the exact AD formula described in [1]. Using the notations from the CNMF manuscript, (we use different naming conventions than [1], T_t is denotes τ, τ is denoted i)

$$V_{ft} = V^a_{ft} + V^d_{ft}$$

where V^a is the attack spectrogram while V^d is the decay spectrogram. These are further modeled as

$$V^a_{ft} = \sum_{q=1}^{r} W^a_{fq} \sum_{i=t-\tau}^{t+\tau} H_{qi} P_{t-i}$$

and

$$V^d_{ft} = \sum_{q=1}^{r} W^d_{fq} \sum_{i=1}^{t} H_{qi} e^{-(t-i)\alpha_q}$$

where it appears from the index definitions that $H_{q:}$ ranges in $[-\tau, m+\tau]$ with m the number of time samples in the audio exerpt. Similarly, $P_:$ has range $[-\tau, \tau]$, and the exponential has input values in $[0, 1-t]\alpha_k$. Note how this last dependence is quite counter-intuitive: say for $t=1$ (beginning of the song), there is no decay, but for the last sample all the song is used as a decay time. There seems to be an inversion here which was probably unintended.

First we need to change the matrices H, P and the exponential argument to match the CNMF convention that H has columns indexes from -2τ to m (instead of from $m-\tau$ to $m+\tau$ in AD). This means modifying the bonds of i by a τ shift, such that

$$V^a_{ft} = \sum_{q=1}^{r} W^a_{fq} \sum_{i=t-2\tau}^{t} H_{qi} P_{t-i}$$

and

$$V^d_{ft} = \sum_{q=1}^{r} W^d_{fq} \sum_{i=1-\tau}^{t-\tau} H_{qi} e^{-(t-i-\tau)\alpha_q}.$$

Note that P is not centered anymore and has indices in $[-2\tau, 0]$; this means that in practice we should look at $t-\tau$ for the activation reported at time t in Attack Decay. This is also why intuitively the exponential term has an additionnal $-\tau$.

Now we perform the change of variable $i := t - i$, which yields

$$V^a_{ft} = \sum_{q=1}^{r} W^a_{fq} \sum_{i=0}^{2\tau} H_{q(t-i)} P_{-i}$$

and

$$V^d_{ft} = \sum_{q=1}^{r} W^d_{fq} \sum_{i=\tau}^{t+\tau-1} H_{q(t-i)} e^{-(i-\tau)\alpha_q}.$$

which are the equations reported in the manuscript.

We can now make the link with CNMF completely explicit. Indeed by grouping the attack and decay terms together,

$$V_{ft} = \sum_{q=1}^{r} \sum_{i=0}^{\max(2\tau,t+\tau-1)} \left[W^a_{fq} P_{-i} \mathbf{1}_{i\in[0,2\tau]} + W^d_{fq} e^{-(i-\tau)\alpha_q} \mathbf{1}_{i\in[\tau,t+\tau-1]} \right] H_{q(t-i)}$$

By grouping $W^a_{fq}P_{-i}$ together as well as $W^d_{fq}e^{-(i-\tau)\alpha_k}$ into a tensor $\tilde{W}_{fqi}$, it appears that V^a_{fk} follows a CNMF model with convolution kernel size 2τ and with structured (in particular rank two) patterns.

[1] T. Cheng et. al., An Attack/Decay model for piano transcription, ISMIR 2016