

## :: Background ::

- Non-negative signal decomposition methods aim to decompose a matrix  $V^+$ , into two matrices,  $W^+$  &  $H^+$  such that  $V = WH$ , where  $W$  can be seen as a dictionary of basis atoms and  $H$  contains the supports for these atoms. Non-negative Matrix Factorisation (NMF) [1] is the most commonly used of these methods. There exist also non-negative dictionary learning methods, such as NN-K-SVD [2], a non-negative variant on the K-SVD algorithm.

- Non-negative signal decompositions are useful for audio processing, as a magnitude spectrogram can be decomposed into spectrum basis atoms, and their time supports. Non-negative decomposition methods have been applied to source separation and automatic music transcription (AMT). In [3], NMF and NN-K-SVD were compared for the purpose of AMT. NMF shown to perform slightly better for the purpose.

- Much subsequent AMT research done with NMF. In [4] the atoms were constrained to be harmonic through filtering resulting in improved transcription. Bayesian NMF allows use of different constraints, such as sparsity and time smoothness [5] through the use of appropriate priors.

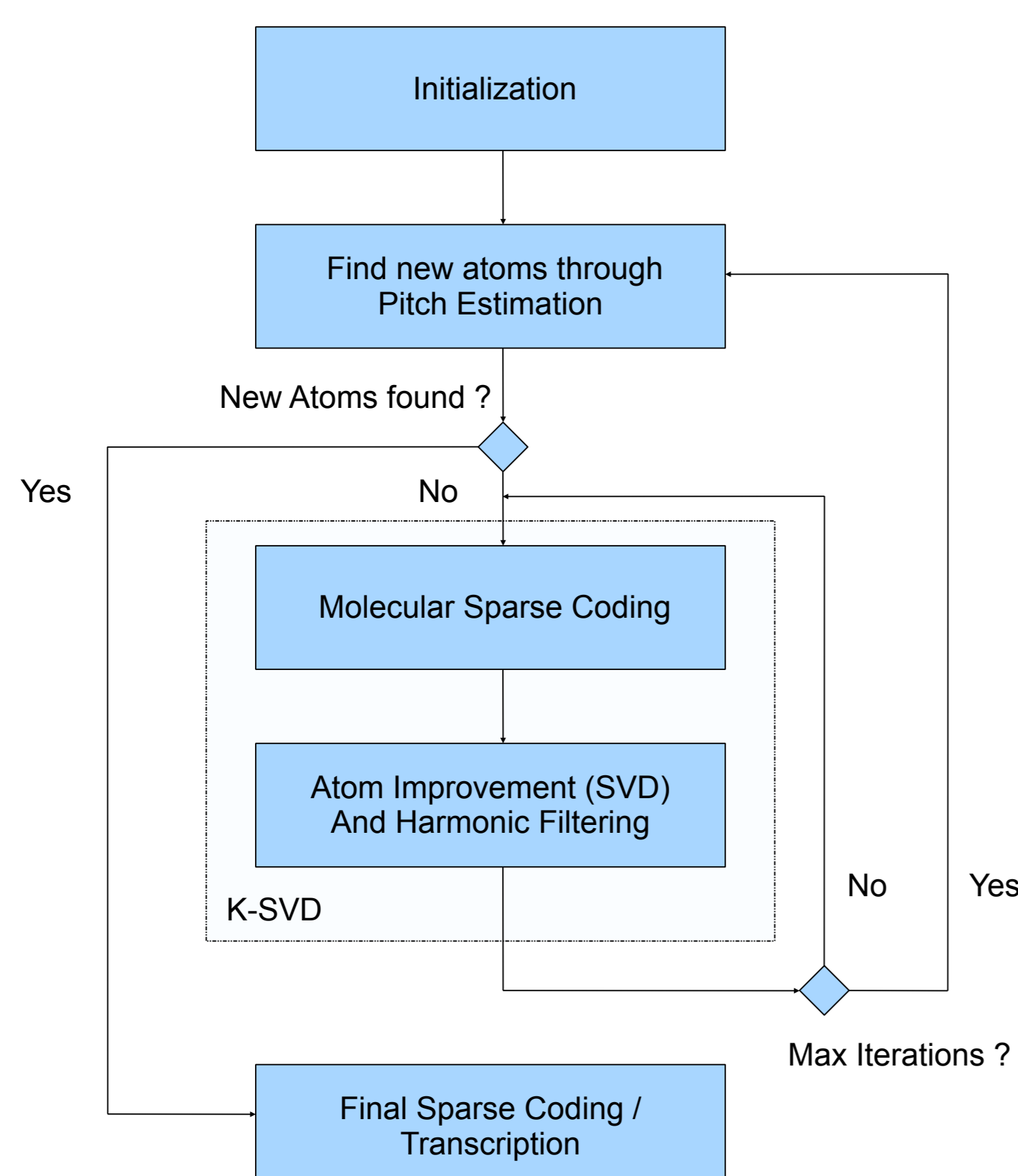
## :: Our Approach ::

- We propose a sparse dictionary learning method, based on the NN-K-SVD for the purpose of AMT, incorporating structured sparsity methods to create a decomposition consisting of time-persisting, harmonic molecules representing notes.

- We first perform a pitch estimation, using harmonic atoms as defined in Harmonic MP [6]. For each significant pitch detected, we learn one atom, performing an SVD over the time bins in the spectrogram in which the atom is dominant. The atoms are then filtered to be harmonic based on their initial perceived pitch.

- The dictionary is then refined using several iterations of the NN-K-SVD algorithm, in which a molecular version of the NN-OMP proposed in [7] is used for sparse coding. Harmonic filtering is incorporated into the atom refining step of the K-SVD algorithm.

- This pitch estimation / H-NN-K-SVD is then repeated iteratively using the reconstruction error as input to the pitch estimation step. This is repeated until no more new atoms are detected.



## :: Harmonic Atoms ::

- We adapt the definition of a harmonic atom as proposed for pitch detection in [HMP] for use with a magnitude spectrogram, and for sub-frequency bin pitch resolution.

$$|H(f_0, t)| = \sum_{k=1}^K \max_{|f_k - k \cdot f_0| < ka.res} S(f_k, t)$$

where  $S$  is the magnitude spectrogram,  $res$  is the sub-frequencyband resolution and  $a$  is a constant.

- In the atom detection stage, the coefficients for harmonic atoms are calculated using (1) at each time bin. We record the pitch of the dominant harmonic atom at each time bin. An atom is learned for each pitch that is recorded a significant amount of times, by performing a SVD over the time bins at which the pitch is recorded
- Atoms which have been learned in this way, and atoms that have been refined in the K-SVD are subject to a harmonic comb filtering.
- The comb filter is defined as the set of spectrogram coefficients used in the calculation of the harmonic atom coefficients, and their sidelobes.

## :: Molecular NN-OMP ::

- Sparse Coding done using NN-OMP with a molecule consisting of time-correlated atoms extracted at each iteration.
- A molecule is grown by searching backwards and forwards in time from an atom selected from a smoothed coefficient matrix at each iteration (see Molecular Matching Pursuit [8]). Using the smoothed coefficient matrix and time-persisting makes the sparse coding more robust to noise, particularly in the presence of transients.

### Input

Spectrogram  $X \in \mathbb{R}_+^{f \times \tau}$ ; Dictionary  $D \in \mathbb{R}_+^{f \times k}$

Threshold  $th$ ; Smoothing factor  $\delta$

### Initialise

$i = 0$ ;  $\mathbf{R}^0 = \mathbf{X}$ ;  $S_i = \{\}$   $\forall t$ ;

$\tau_{min} = 1$ ;  $\tau_{max} = T$ ;

### Iterate

do

$i = i + 1$ ;

$\alpha(j, t) = \arg \min_{z_j} \|d_j z_j - r_t^{i-1}\|_2^2$  for  $\tau_{min} \leq t \leq \tau_{max}$ ,  $1 \leq j \leq k$ ;

$\alpha'(j, t) = \sum_{\tau=t}^{t+\delta-1} \alpha(j, \tau) / \delta$  for  $\tau_{min} - \delta + 1 \leq t \leq \tau_{max}$ ,  $1 \leq j \leq k$ ;

$(j', t') = \arg \max_{j,t} \alpha'$ ;  $max\_val = \alpha'(j', t')$ ;

$\tau_{max} = t'$ ;  $\tau_{min} = t'$ ;

while  $\alpha(j', \tau_{max}) > th$

$S_{\tau_{max}} = S_{\tau_{max}} \cup j'$ ;  $\tau_{max} = \tau_{max} + 1$ ;

while  $\alpha(j', \tau_{min}) > th$

$S_{\tau_{min}} = S_{\tau_{min}} \cup j'$ ;  $\tau_{min} = \tau_{min} - 1$ ;

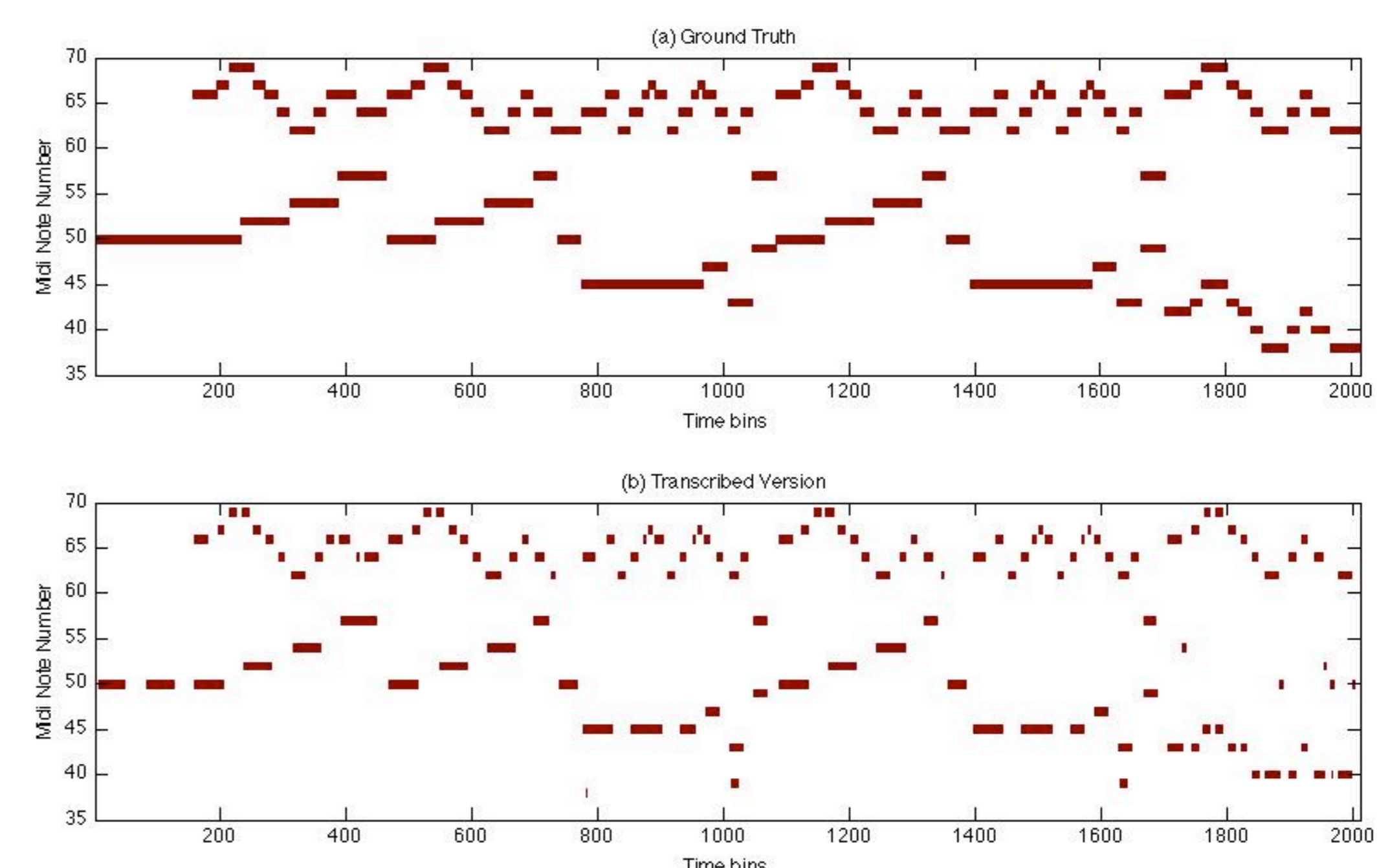
$z_t^j = \min \|D z_t - x_t\|$ , for  $\tau_{min} \leq t \leq \tau_{max}$ ,  $Support\{z_t\} = S_t$

$r_t^j = x_t - D z_t^j$

while  $max\_val > th$

## :: Work To Date ::

- Performing Transcription on simple Midi Files with low polyphony (2-3), and low-mid pitch range (2-3 octaves)
- Precision, Recall and Accuracy of 80%+
- Example shown below (excerpt from Ode To Joy on midi piano)



## :: Current & Future Work ::

- Currently looking at transcribing files from MAPS database – high quality real and synthesised piano sounds aligned to midi files.
- Audio files with higher complexity, polyphony and range than previous work
- Several problems presented to current approach
  - Iterative Pitch estimation not so robust through larger amount of iterations – may need to extract harmonic atoms in one pitch estimation.
  - Higher pitched notes hard to detect as they contain lower energy which often tends to be swamped both in atom detection and learning processes
  - Lower pitched atoms can be highly correlated – may make sense to try multi-scale approach
  - Correlation between elements of polyphonic signal makes sparse coding more difficult may need to consider interactions between atoms
- Accuracy still high (few false positives), recall and precision low due to large amount of false negatives

## :: References ::

- [1] Lee & Seung, 2001, "Algorithms for non-negative matrix factorisation".
- [2] Aharon, Elad & Bruckstein, 2005, "K-SVD and its non-negative variant for dictionary design".
- [3] Bertin, Badeau & Richard, 2007, "Blind signal decompositions for automatic music transcription:....".
- [4] Vincent, Bertin & Badeau, 2008, "Harmonic and inharmonic NMF for polyphonic pitch transcription".
- [5] Bertin, Badeau & Vincent, 2010 "Enforcing harmonicity and smoothness in Bayesian NMF ...."
- [6] Gribonval, 2003, "Harmonic decomposition of audio signals with Matching Pursuit"
- [7] Bruckstein, Elad & Zibulevsky, 2007, "On the uniqueness of non-negative sparse solutions to underdetermined systems of equations".
- [8] Daudet, 2006, "Sparse and structured decompositions of signals with the molecular MP"