Multivariate Iterative Hard Thresholding for sparse decomposition with flexible sparsity patterns

Lucas Rencker, Wenwu Wang, Mark D. Plumbley - University of Surrey, UK
{l.rencker, w.wang, m.plumbley}@surrey.ac.uk

UNIVERSITY OF SURREY

MacSeNet
Machine Sensing Training Network

- Sparse coding is often performed over several consecutive vectors, i.e. audio frames or image patches.
- We propose a multivariate sparse coding algorithm to decompose several frames simultaneously.
- We compare the performance of a *global* sparsity prior over several frames, versus a frame-wise sparsity prior.

“Frame by frame” sparse coding

Sparse coding of multiple frames $x_1, \dots, x_T$:

$$\hat{x}_t = \underset{x_t}{\text{argmin}} \|y_t - Dx_t\|_2^2, \text{ s.t. } \|x_t\|_0 < K \quad \forall t = 1, \dots, T \qquad (1)$$

$\Rightarrow$ each frame is processed **individually**

$\Rightarrow$ sparsity is enforced **framewise**, with parameter K

Algorithm 1 IHT for multiple signals

for $t = 1, \dots, T$ **do** ▷ Loop over frames
  initialize: $x_t^0 = 0$
  while $\|y_t - Dx_t^n\|_2^2 > \epsilon$ **do**
    $x_t^{n+1} = x_t^n + \mu D^T(y_t - Dx_t^n)$ ▷ Gradient descent
    $x_t^{n+1} \leftarrow \mathcal{H}_K(x_t^{n+1})$ ▷ Hard Thresholding
return $\{x_t\}_{t=1..T}$

$\Rightarrow$ **Limitations**:

- Sparsity parameter K might be difficult to choose (K might vary between frames)
- vector x_t might not be sparse (ex: transients in audio)
- how to enforce sparsity *across* frames? (ex: sparsity over time)

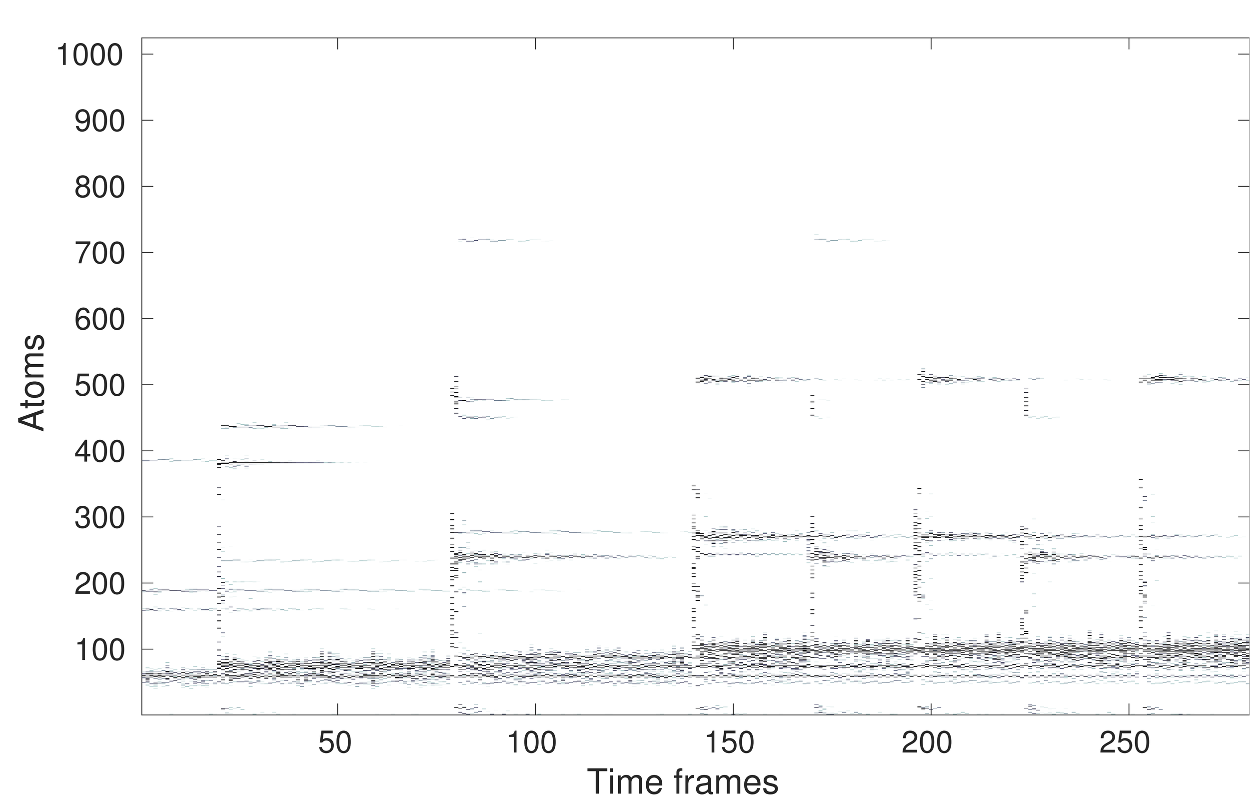

Figure: Sparse coding of $T = 280$ frames from a glockenspiel signal, using $K = 32$ atoms by frame, i.e. 8960 coefficients in total. SNR = 21.7 dB

Multivariate sparse coding

We define $Y = [y_1, \dots, y_T] \in \mathbb{R}^{N \times T}$ as the matrix containing T adjacent frames concatenated altogether, and $X = [x_1, \dots, x_T] \in \mathbb{R}^{M \times T}$ the corresponding sparse activation matrix.

Proposition:

- Reformulate the cost function in (1) in a **multivariate** way
- Introduce a **global** sparsity prior:

$$\hat{X} = \underset{X}{\text{argmin}} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|X\|_0 < K_{\text{tot}} \qquad (2)$$

$$= \underset{x_1 \dots x_T}{\text{argmin}} \sum_{t=1}^{T} \|y_t - Dx_t\|_2^2 \quad \text{s.t.} \quad \sum_{t=1}^{T} \|x_t\|_0 < K_{\text{tot}} \qquad (3)$$

- We introduce a *masking matrix* $\mathcal{M}$ in order enforce the stopping criterion $\|Y_t - DX_t^{n+1}\|_2^2 \leq \epsilon$

Algorithm 2 Multivariate IHT with global sparsity

Require: $Y, D, \epsilon, K_{\text{tot}}, \mu, \mathcal{M} = \mathbb{1}_{M \times T}$
  initialize: $X^0 = 0, n = 0$
  while $\exists\, t$ s.t. $\|Y_t - DX_t^n\|_2^2 > \epsilon$ **do**
    $X^{n+1} = X^n + \mu \mathcal{M} \otimes \left(D^T(Y - DX^n)\right)$
    $X^{n+1} \leftarrow \mathcal{H}_{K_{\text{tot}}}^{\text{global}}(X^{n+1})$ ▷ Global Hard Thresholding
    for t = 1,...,T **do** ▷ Check stopping criterion
      if $\|Y_t - DX_t^{n+1}\|_2^2 \leq \epsilon$ **then**
        $\mathcal{M}_t \leftarrow \mathbb{0}_M$ ▷ Update mask
  return X

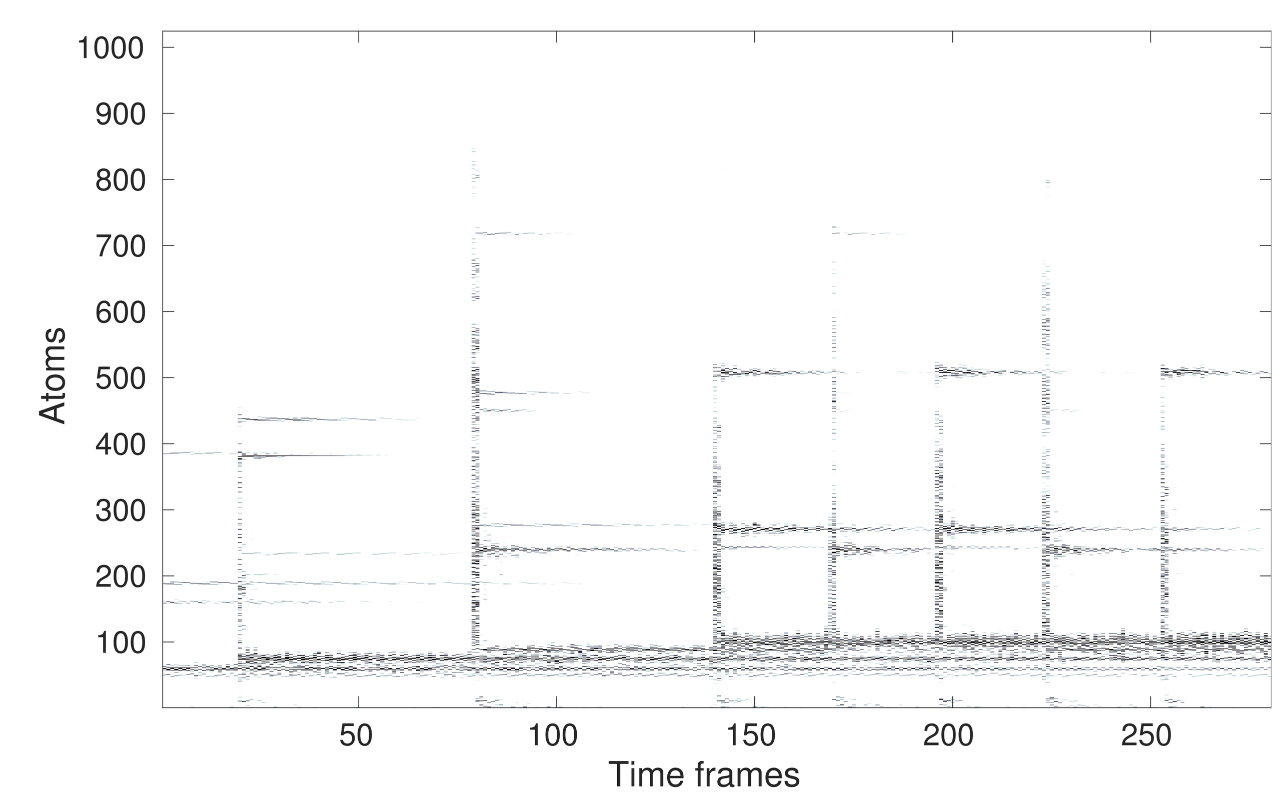

Figure: Sparse coding of the same signal, using a global sparsity prior with $K_{\text{tot}} =$ 8960 atoms in total, using the proposed Multivariate IHT algorithm. SNR = 31.8 dB. The proposed approach manages to recover tonal components as well as transients.

Advantages of proposed approach:

- Frames are processed simultaneously, which allows for more flexible sparsity patterns
- Sparity is enforced **on average**, which means that more dense frames can “borrow" coefficients from sparser frames

Experiments

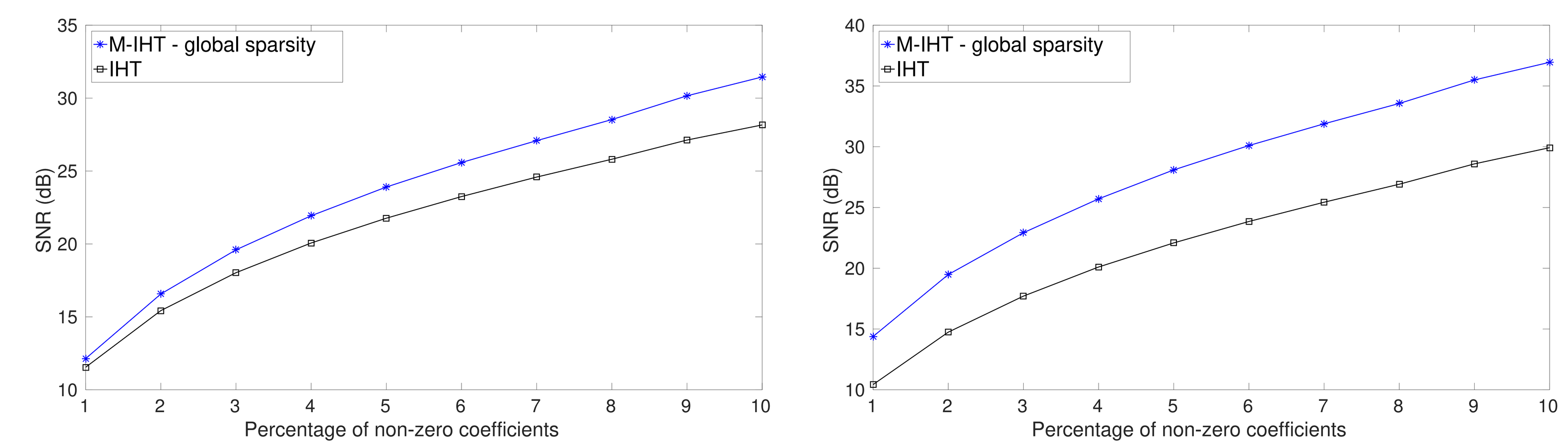

Figure: Audio coding experiment. SNR of reconstructed signal, as a function of the percentage of non-zero coefficients. Left: music signals. Right: speech signals.

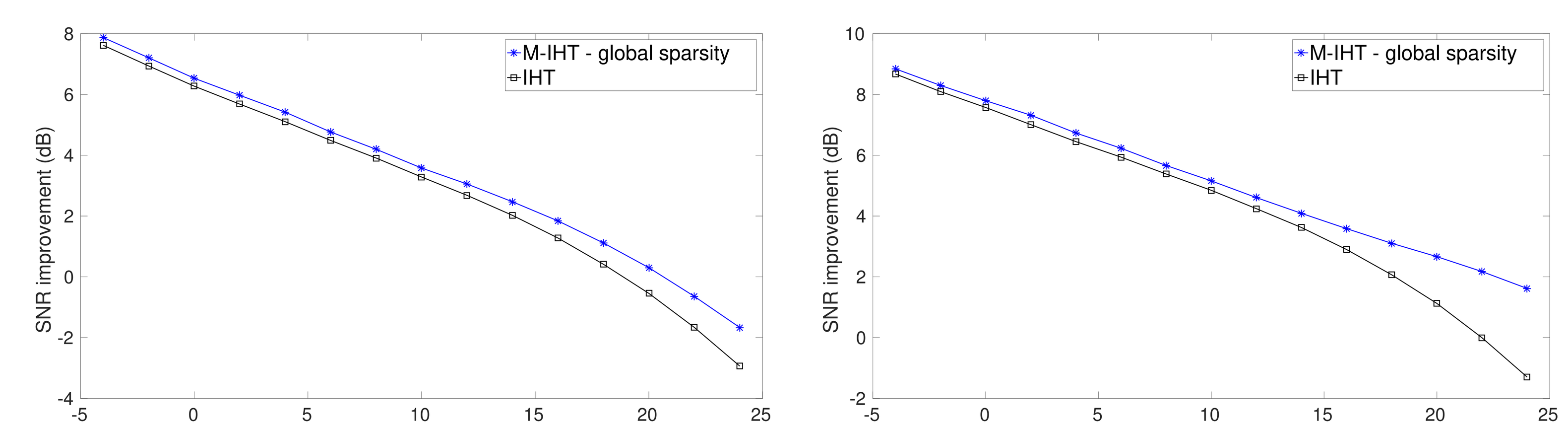

Figure: Audio denoising experiment. SNR improvement of the reconstructed signal, as a function of the input SNR. Left: music signals. Right: speech signals.

	IHT	Multivariate IHT with global sparsity
CPU time (s)	20.8	2.0

Table: Average CPU time to process 1s of signal (in Matlab).

Conclusion

- Multivariate IHT allows to process several frames simultaneously, thus more flexible sparsity patterns.
- A *global* sparsity prior performs better than a frame-wise sparsity prior, allowing both very sparse and very dense frames to be efficiently represented simultaneously.
- Multivariate IHT is much faster (in Matlab) than the classic IHT.

MacSeNet
Machine Sensing Training Network

The research leading to these results has received funding from the European Union's H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet