WWW.KFS.OEAW.AC.AT

ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Fast Matching Pursuit with Multi-Gabor Dictionaries

Zdeněk Průša

Finding the best K-sparse approximation of a signal in a redundant dictionary is an NP-hard problem significant for many applications. Suboptimal greedy matching pursuit (MP) algorithms are generally used for such task. We present here an acceleration technique for matching pursuit-based algorithms acting on a multi-Gabor dictionary; a concatenation of several Gabor-type time-frequency dictionaries; each of which consisting of translations and modulations of a possibly different prototype window and time and frequency shift parameters. The technique is based on pre-computing and thresholding inner products between atoms and on updating the residuum and the approximation error estimate directly in the coefficient domain i.e. without the round-trip to the signal domain. We provide an implementation written in C (compatible with C99 and C++11) and we also provide Matlab and GNU Octave wrappers. For some settings, the implementation is almost 100 times faster than the standard Matching Pursuit Toolkit (MPTK) [1].

Multi-Gabor dictionary

A Gabor dictionary $\mathbf{D}_{(\mathbf{g},a,M)}$ generated from a window $\mathbf{g} \in \mathbb{R}^L, \|\mathbf{g}\|_2 = 1$, time shift a and a number of modulations M is given as

$$\mathbf{D}_{(\mathbf{g},a,M)}(l, m+nM) = \mathbf{d}_{m+nM}(l) = \mathbf{T}_{na}\mathbf{M}_{m/M}\mathbf{g}(l) = \mathbf{g}(l-na)\mathrm{e}^{\mathrm{i}2\pi m(l-na)/M}$$

for $l = 0, \dots, L-1$ and $m = 0, \dots, M-1$ for each $n = 0, \dots, N-1$, where $N = L/a$ is the number of window time shifts and the overall number of atoms is $P = MN$. The expression $(l - na)$ is assumed to be evaluated modulo L according to the circular indexing.

A multi-Gabor dictionary consisting of W Gabor dictionaries is defined as

$$\left[\mathbf{D}_{(\mathbf{g}_1,a_1,M_1)} \middle| \mathbf{D}_{(\mathbf{g}_2,a_2,M_2)} \middle| \dots \mathbf{D}_{(\mathbf{g}_W,a_W,M_W)}\right].$$

The parameters a_w and M_w are restricted such that M_w/a_w is an integer and every pair a_w, a_v and M_w, M_v is divivible by $a_{\min} = \min\{a_w, a_v\}$ and $M_{\min} = \min\{M_w, M_v\}$ respectivelly. Windows $\mathbf{g}_w$ are chosen such that the overall dictionary forms a frame for $\mathbb{C}^L$.

Gram Matrix

A cross-Gram matrix from dictionary v to dictionary w will be denoted as $\mathbf{G}_{w,v} = \mathbf{D}_w^*\mathbf{D}_v$ and the overall Gram matrix has the following block structure

$$\begin{bmatrix} \mathbf{G}_{1,1} & \mathbf{G}_{1,2} & \dots & \mathbf{G}_{1,W} \\ \mathbf{G}_{2,1} & \mathbf{G}_{2,2} & \dots & \mathbf{G}_{2,W} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_{W,1} & \mathbf{G}_{W,2} & \dots & \mathbf{G}_{W,W} \end{bmatrix}.$$

If $a_v = a_w$ and $M_v = M_w$,

the cross-Gram matrix exhibits the following structure

$$\mathbf{G}_{w,v}(m+nM, k+jM) = \mathbf{h}_{w,v}(m-k, n-j)_M \mathrm{e}^{\mathrm{i}2\pi k\frac{a}{M}(n-j)}$$

for $m = 0, \dots, M-1$ for each $n = 0, \dots, N-1$ and $k = 0, \dots, M-1$ for each $j = 0, \dots, N-1$ where

$$\mathbf{h}_{w,v} = \mathbf{D}_w^*\mathbf{g}_v \in \mathbb{C}^{MN}.$$

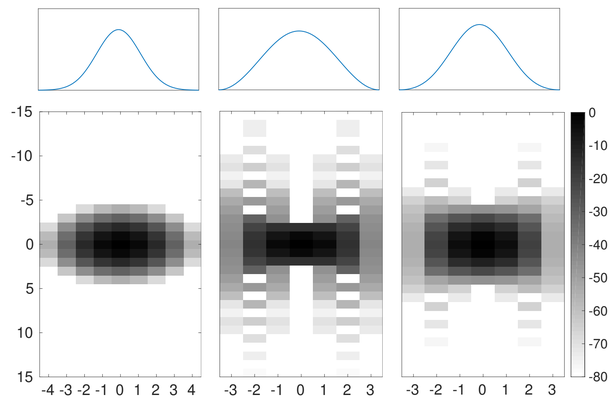


Figure: Examples of abs. values of truncated kernels. (Left) Gaussian, (middle) Hann and (right) Blackman windows.

If $a_v \neq a_w$ or $M_v \neq M_w$,

the cross-Gram matrix can be constructed by sub-sampling a kernel computed for the common lattice $a_{\min} = \min\{a_v, a_w\}$, $M_{\max} = \max\{M_v, M_w\}$ as

$$\mathbf{h}_{w,v} = \mathbf{D}^*_{(\mathbf{g}_w,a_{\min},M_{\max})}\mathbf{g}_v \in \mathbb{C}^{M_{\max}L/a_{\min}}$$

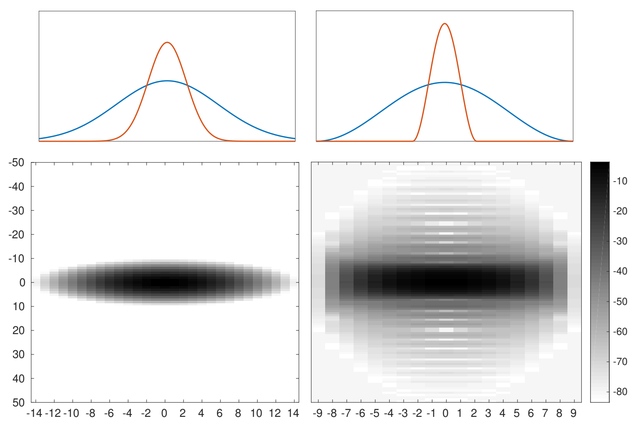


Figure: Examples of abs. values of truncated cross-kernels. (Left) Gaussian and (right) Hann windows.

$$\mathbf{c}(m,n)_M = \mathbf{c}(m+nM)$$

Matching Pursuit

The best K-sparse approximation of a signal $\mathbf{x} \in \mathbb{C}^L$ in an overcomplete dictionary of P normalized atoms (vectors) $\mathbf{D} = [\mathbf{d}_0|\mathbf{d}_1|\dots|\mathbf{d}_{P-1}] \in \mathbb{C}^{L\times P}$, $\|\mathbf{d}_p\|_2 = 1$ can be formally written as the solution to

$$\min\|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2 \quad \text{subject to} \quad \|\mathbf{c}\|_0 \leq K.$$

The matching pursuit (MP) algorithm [2] finds a *suboptimal* solution by an iterative greedy atom selection.

Input: *Input signal* $\mathbf{x}$, *dictionary* $\mathbf{D}$
Output: *Solution vector* $\mathbf{c}$
Initialization: $\mathbf{c} = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{x}$, $E_0 = \|\mathbf{r}_0\|_2^2$, $k = 0$
while *Stopping criterion not met* ***do***
1. ***Selection:*** $p_{max} = \operatorname{argmin}_p \|\mathbf{r}_k - \langle\mathbf{r}_k, \mathbf{d}_p\rangle\mathbf{d}_p\|_2^2 = \operatorname{argmax}_p |\langle\mathbf{r}_k, \mathbf{d}_p\rangle|$
   $c_{max} = \langle\mathbf{r}_k, \mathbf{d}_{p_{max}}\rangle$
2. ***Update:***
   2.1 ***Solution:*** $\mathbf{c}(p_{max}) = \mathbf{c}(p_{max}) + c_{max}$
   2.2 ***Error:*** $E_{k+1} \leftarrow E_k - |c_{max}|^2$
   2.3 ***Residual:*** $\mathbf{r}_{k+1} = \mathbf{r}_k - c_{max}\mathbf{d}_{p_{max}}$
   $k \leftarrow k+1$
end

Alternative residual update step 2.3 in the coefficient domain:

$$\mathbf{D}^*\mathbf{r}_{k+1} = \mathbf{D}^*\mathbf{r}_k - c_{\max}\mathbf{G}(:, p_{\max}).$$

Faster Approximate Residual Update

Single dictionary: Employing the kernel, the residual update in the coefficient domain reduces to a subtraction of a truncated, modulated and weighted kernel

$$\mathbf{h}^{(k,j)}(m,n)_M = \mathbf{h}(m,n)_M \mathrm{e}^{\mathrm{i}2\pi k\frac{a}{M}n}$$

from the neighborhood of the time-frequency position k, j ($p_{\max} = k + jM$) which can be written as

$$\mathbf{D}^*\mathbf{r}_{k+1}(\mathcal{M},\mathcal{N})_M = \mathbf{D}^*\mathbf{r}_k(\mathcal{M},\mathcal{N})_M - c_{\max}\mathbf{h}^{(k,j)}$$

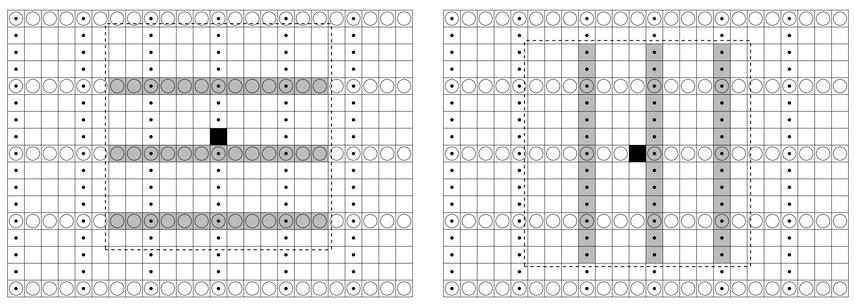
assuming $\mathbf{h}$ has already been truncated and $\mathcal{M}$ and $\mathcal{N}$ denote the index set of the appropriate neighborhood.

Multiple dictionaries: In addition, subsampling of the appropriate cross-kernel $\mathbf{h}_{w,v}$ during the update is introduced

$$\mathbf{D}_w^*\mathbf{r}_{k+1}(\mathcal{M},\mathcal{N})_M = \mathbf{D}_w^*\mathbf{r}_k(\mathcal{M},\mathcal{N})_M - c_{\max}\mathbf{h}_{w,v}^{(k,j)}(\mathcal{K},\mathcal{I}),$$

where $\mathcal{K}, \mathcal{I}$ are indices in the kernel with steps $M_{\max}/M_w$ and $a_w/a_{\min}$ respectivelly.
Example: $a_1 = 4a_{\min}, M_1 = 8a_{\min}$ and $a_2 = a_{\min}, M_2 = 2a_{\min}$. Both cross-kernels $\mathbf{h}_{1,2}$ and $\mathbf{h}_{2,1}$ are computed with $a_{\min} = a_2$ and $M_{\max} = M_1$.

• a_1, M_1 ◯ a_2, M_2 □ $a_{\min}, M_{\max}$ ■ Selected ■ Updated



Real Atoms from a Complex Dictionary

For each pair of conjugated atoms arranged into a matrix $\mathbf{P} = [\mathbf{d}|\overline{\mathbf{d}}] \in \mathbb{C}^{L\times 2}$, the projection can be written as $\mathbf{P}(\mathbf{P}^*\mathbf{P})^{-1}\mathbf{P}^*\mathbf{r}_k$ and the energy of a potential new residual becomes $\|\mathbf{r}_k - \mathbf{P}(\mathbf{P}^*\mathbf{P})^{-1}\mathbf{P}^*\mathbf{r}_k\|_2^2$. Due to the orthogonality of the projection, the pair which decreases the residual error the most is the one with highest $\|\mathbf{P}(\mathbf{P}^*\mathbf{P})^{-1}\mathbf{P}^*\mathbf{r}_k\|_2$. The inverted expression can be written as

$$(\mathbf{P}^*\mathbf{P})^{-1} = \begin{bmatrix} 1 & \langle\overline{\mathbf{d}}, \mathbf{d}\rangle \\ \langle\mathbf{d}, \overline{\mathbf{d}}\rangle & 1 \end{bmatrix}^{-1} = \frac{1}{1 - \left|\langle\mathbf{d}, \overline{\mathbf{d}}\rangle\right|^2}\begin{bmatrix} 1 & -\langle\mathbf{d}, \overline{\mathbf{d}}\rangle \\ -\overline{\langle\mathbf{d}, \overline{\mathbf{d}}\rangle} & 1 \end{bmatrix}$$

and the *dual* pair of conjugated atoms is obtained as

$$\widetilde{\mathbf{P}} = [\widetilde{\mathbf{d}} \mid \widetilde{\overline{\mathbf{d}}}] = \mathbf{P}(\mathbf{P}^*\mathbf{P})^{-1} = \frac{1}{1 - \left|\langle\mathbf{d}, \overline{\mathbf{d}}\rangle\right|^2}\left[\mathbf{d} - \overline{\langle\mathbf{d}, \overline{\mathbf{d}}\rangle}\overline{\mathbf{d}} \;\middle|\; \overline{\mathbf{d}} - \langle\mathbf{d}, \overline{\mathbf{d}}\rangle\mathbf{d}\right].$$

The energy of the projection can be written as

$$\begin{aligned}\|\mathbf{P}(\mathbf{P}^*\mathbf{P})^{-1}\mathbf{P}^*\mathbf{r}_k\|_2^2 &= \|\langle\mathbf{r}_k, \widetilde{\mathbf{d}}\rangle\mathbf{d} + \langle\mathbf{r}_k, \widetilde{\overline{\mathbf{d}}}\rangle\overline{\mathbf{d}}\|_2^2 \\ &= 2\left|\langle\mathbf{r}_k, \widetilde{\mathbf{d}}\rangle\right|^2\left(1 + \operatorname{Re}(\mathrm{e}^{\mathrm{i}2\phi}\langle\mathbf{d}, \overline{\mathbf{d}}\rangle)\right),\end{aligned}$$

where $\phi = \arg(\langle\mathbf{r}_k, \widetilde{\mathbf{d}}\rangle)$. The inner product of the residual with the first of the dual atoms $\widetilde{\mathbf{d}}$ can be expanded into

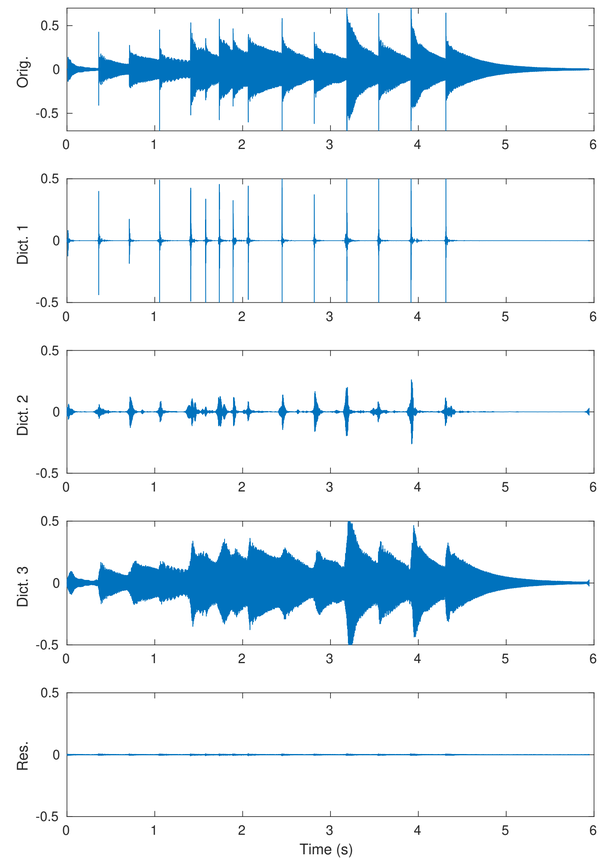
$$\widetilde{c} = \langle\mathbf{r}_k, \widetilde{\mathbf{d}}\rangle = \left(\langle\mathbf{r}_k, \mathbf{d}\rangle - \langle\mathbf{d}, \overline{\mathbf{d}}\rangle\overline{\langle\mathbf{r}_k, \mathbf{d}\rangle}\right) / \left(1 - \left|\langle\mathbf{d}, \overline{\mathbf{d}}\rangle\right|^2\right),$$

and the inner product with the conjugated one is just conjugated.

Matlab Example

```
[f,fs] = gspi;
[c,frec,info] = multidgtrealmp(f,...
{'blackman',64,512,...
 'blackman',256,2048,...
 'blackman',1024,8192});
frecd = info.synthetize(c);
xvals = (0:numel(f)-1)/fs;
plot(xvals,f);ylabel('Orig.');
plot(xvals,frecd(:,1));ylabel('Dict. 1');
plot(xvals,frecd(:,2));ylabel('Dict. 2');
plot(xvals,frecd(:,3));ylabel('Dict. 3');
plot(xvals,f-frec);ylabel('Res.');
atomsTotal = sum(cellfun(@numel,c));
atomsSelected = ...
sum(cellfun(@(e)numel(find(abs(e)>0)),c));
error=20*log10(norm(f-frec)/norm(f));
```

Selects `atomsSelected = 8094` from `atomsTotal = 3151104` with approximation error -40 dB.



(available in LTFAT at `http://ltfat.github.io`)

Comparison with MPTK [1]

Timing in seconds for performing 200 thousands iterations on a 137 seconds long piano recording sampled at 44.1 kHz totaling $6 \cdot 10^6$ samples using single Gabor dictionary with various redundancies (above) and five Gabor dictionaries (below).

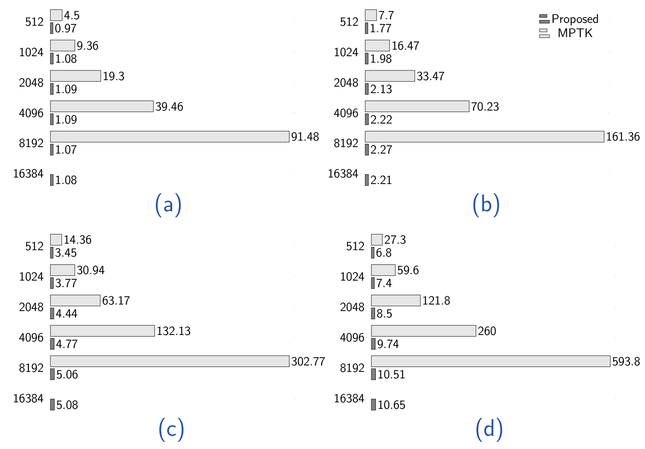


Figure: Single Gabor dictionary for different M and various redundancies: (a) $a = M/4$ (b) $a = M/8$ (c) $a = M/16$ (d) $a = M/32$.

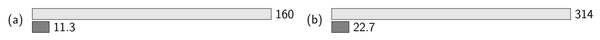


Figure: Five Gabor dictionaries with $M_w = 512, \dots, 8192$, (a) $a_w = M_w/4$, (b) $a_w = M_w/8$

References

[1] S. Krstulović and R. Gribonval, "MPTK: Matching pursuit made tractable," in *Proc. Int. Conf. on Acoustics Speech and Signal Processing ICASSP 2006*, vol. 3, May 2006, pp. III–496–III–499.

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Tran. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

This work was supported by the Austrian Science Fund: Y 551–N13 and I 3067–N30.

ISF-INSTITUT FÜR SCHALLFORSCHUNG