Bachelor Project -
Random Subdictionaries of Convolutional Dictionaries

Supervisor: Karin Schnass

A useful concept that allows for efficient signal processing is that the signals in question are sparse in a dictionary. A dictionary Φ is defined as a collection of K unit norm vectors $\phi_k \in \mathbb{R}^d$ called atoms. The atoms are stacked as columns in a matrix, which by abuse of notation is also referred to as the dictionary, that is $\Phi = (\phi_1, \ldots, \phi_K) \in \mathbb{R}^{d\times K}$. A signal $y \in \mathbb{R}^d$ is then called sparse in a dictionary Φ if up to a small approximation error or noise η it can be represented as linear combination of a small (sparse) number of dictionary atoms,

$$y = \sum_{k\in I} \phi_k x[k] + \eta = \Phi_I x_I + \eta \quad \text{or} \quad y = \Phi x + \eta \quad \text{with} \quad \|x\|_0 = |I| = S, \tag{1}$$

where $\|\cdot\|_0$ counts the non zero components of a vector or matrix. The index set I storing the non zero entries is called the support with the understanding that for the sparsity level $S = |I|$ we have $S \ll d \leq K$ and that $\|\eta\|_2 \ll \|y\|_2$ or even better $\eta = 0$.
While having an S-sparse approximations is useful (e.g. store S values and addresses instead of d values), looking through $\binom{K}{S}$ possible index sets to find this best S-sparse approximation is certainly not practical. Therefore researchers have developed suboptimal but faster approximation routines, such as thresholding, (Orthogonal) Matching Pursuit and the Basis Pursuit Principle, together with conditions when these routines will find a best approximation. For instance for an incoherent dictionary, meaning

$$\max_{i\neq j} |\langle \phi_i, \phi_j\rangle| = \mu \ll 1, \tag{2}$$

OMP/BP will recover any best S-sparse approximation as long as $S \lesssim \mu^{-1}$, [3]. As long as $S \log K \lesssim \mu^{-2}$ it will recover a best S-sparse approximation with high probability, [4, 2].
Unfortunately these results do not cover a very important class of dictionaries called shift-invariant or convolutional dictionaries. Let T_n define the shift-operator, which acts on a vector in $v \in \mathbb{R}^d$ as

$$(T_n v)[i] = v[(i-n)_{\bmod d}].$$

A convolutional dictionary has the property that for any atom ϕ_i also all of its shifted versions $T_n\phi_i$ are included in the dictionary, meaning that the dictionary actually consists of several mother atoms m_k and all their shifted versions

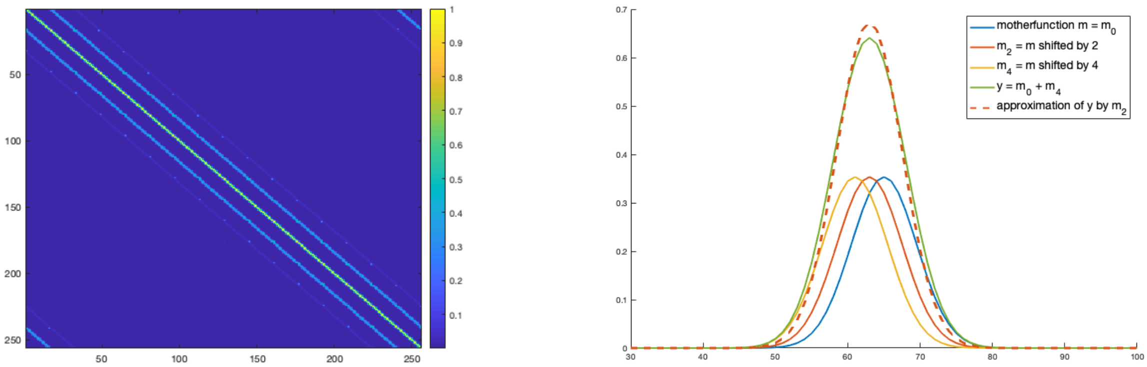
$$\Phi = (m_1, T_1 m_1, \ldots T_{d-1} m_1, m_2, \ldots T_{d-1} m_2, \ldots m_K, T_1 m_K, \ldots T_{d-1} m_K) \in \mathbb{R}^{d\times dK}.$$

The advantage of such dictionaries is that they are very easy to store and to calculate with. For instance inner products between a signal and all shifted versions of one mother function can be efficiently calculated using the FFT. The main theoretical disadvantage is that they are usually very coherent $\mu \approx 1$, and so we lose the main tool for any theoretical analysis. On the other hand the coherence structure - meaning the structure of the Gram-matrix $\Phi^\star\Phi$ - of a shift invariant dictionary is very particular, see Figure 1(a)[1].

$$\langle T_n m_k, T_\ell m_j\rangle = \langle T_{n-\ell} m_k, m_j\rangle = f(n-\ell, k, j)$$

[1]For ease of notation we will from now on assume that addition and subtraction of indices is modulo d.

So a subdictionary of a convolutional dictionary generated by localised and incoherent motherfunctions is incoherent if the selected atoms are sufficiently separated. Assuming a signal model where the signals are only linear combination of such sufficiently separated atoms, also makes sense. After all, a signal that is the sum of two very close bump functions is also well approximated by just one rescaled bump function, compare Figure 1(b).



(a) (b)

Figure 1: (a) Gram matrix of shift invariant dictionary, (b) sum of two bum functions, approximated by a bumb function

The goal of this project is to empirically find the probability that a subdictionary Φ_I of a convolutional dictionary is well-conditioned, if the support I is chosen uniformly at random from all well-separated supports of size S. For this bachelor project we will focus on convolutional frames where the mother functions themselves are related via

$$m_k[j] = (M_k m)[j] = e^{2\pi i k j/d} \cdot m[j]$$

so the inner product structure is of the form

$$\langle T_n m_k, T_\ell m_j \rangle = \langle T_{n-\ell} m_k, m_j \rangle = f(n-\ell, k-j).$$

Tasks:

- Read [3] and familiarise yourself with the cumulative coherence function and estimates for the conditioning of random subdictionaries based on the Gerszgorin disk theorem.
- Read [4] and familiarise yourself with the probability estimates for the conditioning of a random subdictionary.
- Familiarise yourself with the concept of a (discrete) Gabor frame, [1], and the provided MATLAB toolbox.
- For a given separation level θ, implement an algorithm that selects a random index set $I = \{(n_1, m_1), \dots (n_S, m_S)\} \subset [d] \times [d]$ that satisfies

$$\|(n_i - n_j, m_i - m_j)\|_2 \geq \theta \quad \forall i, j \in \{1 \dots S\}, i \neq j.$$

- For a variety of Gabor frames, separation levels θ and sparsity levels S approximate empirically the probability distribution of the (cumulative) coherence and conditioning of a random θ-separated subdictionary.

References

[1] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Birkhäuser, 2001.

[2] K. Schnass. Average performance of Orthogonal Matching Pursuit (OMP) for sparse approximation. *IEEE Signal Processing Letters*, 25(12):1865–1869, 2018.

[3] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, October 2004.

[4] J.A. Tropp. On the conditioning of random subdictionaries. *Applied and Computational Harmonic Analysis*, 25(1):1–24, 2008.