Parametric Dictionary Design for Sparse Coding

Mehrdad Yaghoobi, Laurent Daudet, Mike E. Davies

IDCom, School of Engineering and Electronics
The University of Edinburgh, UK

SPARS09, April 7, 2009

Outline

Sparse Coding

Generative model

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \nu$$

$\mathbf{y} \in \mathbb{R}^d$, $\mathbf{D} \in \mathbb{R}^{d \times N}$, $\mathbf{x} \in \mathbb{R}^N$ and $\nu \in \mathbb{R}^d$. *Under-determined generative model*

$$\Leftrightarrow d < N$$

$$\overbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}}^{\mathbf{y}} = \overbrace{\begin{bmatrix} d_{1,1} & & d_{1,k} & & d_{1,N} \\ d_{2,1} & & d_{2,k} & & d_{2,N} \\ \vdots & \dots & \vdots & \dots & \vdots \\ d_{d,1} & & d_{d,k} & & d_{d,N} \end{bmatrix}}^{\mathbf{D}} \overbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ \vdots \\ x_N \end{bmatrix}}^{\mathbf{x}} + \overbrace{\begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_d \end{bmatrix}}^{\nu}$$

- sparse coding: $\widehat{\mathbf{x}} = \arg\min_x \|\mathbf{x}\|_0 \text{ s.t. } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|^2 \leq \xi,$
- $\xi = 0$ called sparse representation.
- $\xi > 0$ called sparse approximation.

Dictionary Selection Methods

- *Concatenation of orthonormal bases:* Let $\mathcal{O}$ be the set of all orthonormal dictionaries in $\mathbb{R}^{d\times d}$. $\mathcal{D} = \{\mathbf{D}_i\}_{i\in\mathcal{I}}, \forall i \in \mathcal{I}, \mathbf{D}_i \in \mathcal{O}$ is given. A dictionary $\mathbf{D}$ in $\mathbb{R}^{d\times d|\mathcal{I}|}$ is generated using,

$$\mathbf{D} = [\mathbf{D}_1 \cdots \mathbf{D}_i \cdots \mathbf{D}_{|\mathcal{I}|}].$$

- *Dictionary learning using a set of training samples $\mathscr{L} = \{\mathbf{y}_i\}_{i\in\mathcal{I}}$:* The goal is to find a dictionary such that it provides sparser coding for the given class of signals. A general formulation for dictionary learning is,

$$\widehat{\mathbf{D}} = \arg\min_{\mathbf{D}} \sum_{i\in\mathcal{I}} \min_{\mathbf{x}\in\{\theta:\|\mathbf{y}_i-\mathbf{D}\theta\|^2\leq\xi\}} \|\mathbf{x}\|_0.$$

- *Dictionary design subject to a certain property:* These properties include, but not restrict to, Restricted Isometry Property (RIP), minimum coherence μ and minimum cumulative coherence $\mu_1(m)$.

Incoherent Dictionary

Coherence

The coherence of a dictionary $\mathbf{D}$ is defined to be the maximum correlation of two distinct atoms and can be found using,

$$\mu_{\mathbf{D}} = \max_{i,j:j\neq i}\{|\langle \mathbf{d}_i, \mathbf{d}_j\rangle|\}.$$

Incoherent dictionary: A dictionary is incoherent when $\mu_{\mathbf{D}}$ is "small".

Equiangular Tight Frame (ETF)

A column normalized dictionary $\mathbf{D}_G$ is called ETF, when there is a $\gamma \ : \ 0 < \gamma < \pi/2$, such that $|\langle \mathbf{d}_i, \mathbf{d}_j\rangle| = \cos(\gamma) \ : \ \forall i, j \ i \neq j$.
If there exists an ETF in $\mathbb{R}^{d\times N}$, it is the solution of $\arg\min_{\mathbf{D}\in\mathscr{D}}\{\mu_{\mathbf{D}}\}$, and for any $\mathbf{D} \in \mathbb{R}^{d\times N}$, $\mu_{\mathbf{D}} \geq \mu_G := \sqrt{\frac{N-d}{d(N-1)}}$.

Parametric Dictionary

Parametric Dictionary

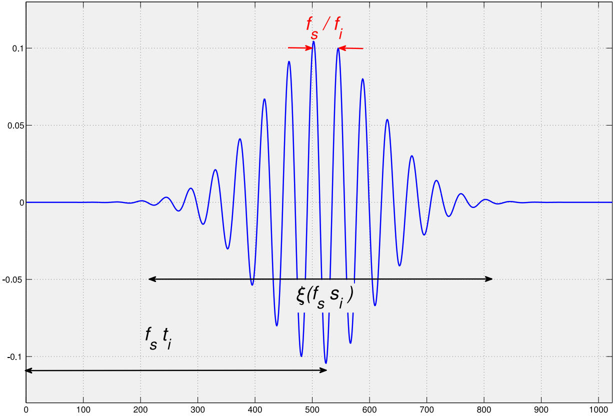
Let Γ be a set of parameters in the admissible set Υ.
$\mathbf{D}_\Gamma : \Upsilon \rightarrow \mathbb{R}^{d \times N}$ is defined to be a mapping from the parameter space to the dictionary space.
Given a mapping $\mathbf{D}_\Gamma$, a dictionary called Parametric Dictionary, when it is generated using a $\Gamma \in \Upsilon$.

Special Case: A special mapping $\mathbf{D}_\Gamma$ is when it is column-separable. In this framework each atom $\mathbf{d}_i$ is generated by a set of parameters γ_i. This type of mapping has often been used in sparse coding.

An example: Multi-scale Gabor Parametric Dictionary

The parameters in a real value multi-scale Gabor atom are scale, phase and the time-frequency shift and the generative function is Gaussian.

$$\mathbf{g}_{\gamma_i}(t) = a_i e^{\left(\frac{t-t_i}{s_i}\right)^2} . \cos(2\pi f_i (t - t_i) + \phi_i)$$



Parametric Dictionary Design

Parametric Dictionary Design

Let $\mathbf{D}_\Gamma$, $\boldsymbol{\Upsilon}$ and a certain dictionary property be given. Parametric Dictionary Design is how to find a $\Gamma^* \in \Upsilon$ such that $\mathbf{D}_{\Gamma^*}$ (approximately) has such a property.

- *Objective:* Finding a dictionary close to being ETF.
- This search is easier to be done in the space of N by N real matrices, by finding the Gram matrix, $\mathbf{G}_\mathbf{D} := \mathbf{D}^T\mathbf{D}$, associated with the optimal parametric dictionary.
- *Closeness measure:* The infinity norm $\|.\|_\infty$ in $\mathbb{R}^{N\times N}$, which is the maximum absolute value of the elements.

Parametric Dictionary Design: Formulation

- Let Θ_d^N be the set of Gram matrices of the ETF's in $\mathbb{R}^{d\times N}$. An incoherent PDD is found using the following optimization problem,

$$\arg \min_{\Gamma\in\Upsilon, \mathbf{G}_G\in\Theta_d^N} \|\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma - \mathbf{G}_G\|_\infty$$

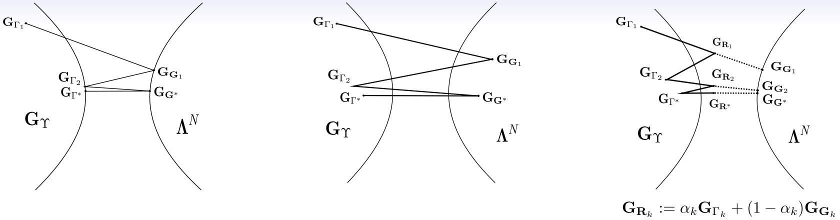
- *Difficulties:* There is no ETF for some selection of (d, N), the set of ETF's in $\mathbb{R}^{d\times N}$ is non-convex and it is difficult to minimize the ℓ_∞ problem.
- *Practical Solution:* Convex relaxation of the set of ETF's,

$$\Lambda^N = \{\mathbf{G}\in\mathbb{R}^{N\times N} : \mathbf{G}=\mathbf{G}^T, \operatorname{diag}\mathbf{G}=1, \max_{i\neq j}|g_{i,j}|\leq \mu_G\},$$

and using ℓ_2-norm as the closeness measure,

$$\arg \min_{\Gamma\in\Upsilon, \mathbf{G}\in\Lambda^N} \|\mathbf{D}_\Gamma^T \mathbf{D}_\Gamma - \mathbf{G}\|_F^2.$$

Optimization Methods



(a) (b) (c)

$$\min_{\mathbf{G}_\Gamma \in \mathbf{G}_\Upsilon, \mathbf{G}_\mathbf{G} \in \Lambda^N} \|\mathbf{G}_\Gamma - \mathbf{G}_\mathbf{G}\|_F^2$$

$$\mathbf{G}_\Upsilon = \{\mathbf{G}_\Gamma : \mathbf{G}_\Gamma = \mathbf{D}_\Gamma^T \mathbf{D}_\Gamma, \Gamma \in \Upsilon\}$$

(a) Alternating Projection: Needs to know the projection operators.

(b) Alternating Minimization: Reduces the objective at each parameter update.

(c) Relaxed Alternating Minimization : When the solution in one of the sets, here Υ, is needed to be found.

PDD Algorithm

Parametric Dictionary Design

1: **initialization:** $k = 1$, $\mathbf{D}_{\Gamma_1} \in \mathscr{D}$, $\{\alpha_i\}_{1 \le i \le K} : 0 < \alpha_i \le 1$
2: **while** $k \le K$ **do**
3: $\mathbf{G}_{\Gamma_k} = \mathbf{D}_{\Gamma_k}^T \mathbf{D}_{\Gamma_k}$
4: $\mathbf{G}_{P_{k+1}} = \arg\min_{\mathbf{G} \in \Lambda^N} \|\mathbf{G}_{\Gamma_k} - \mathbf{G}\|_F$
5: $\mathbf{G}_{R_{k+1}} = \alpha_k \mathbf{G}_{P_{k+1}} + (1 - \alpha_k) \mathbf{G}_{\Gamma_k}$
6: $\mathbf{D}_{\Gamma_{k+1}} \in \mathbf{D}_{\Gamma_k} \cup \{\forall \mathbf{D} \in \mathscr{D} : \|\mathbf{D}^T\mathbf{D} - \mathbf{G}_{R_{k+1}}\|_F < \|\mathbf{G}_{\Gamma_k} - \mathbf{G}_{R_{k+1}}\|_F\}$
7: $k = k + 1$
8: **end while**

4: *Projection of* $\mathbf{G}_{\Gamma_k}$ *onto* Λ^N*:*

$$\mathcal{P}_{\Lambda^N}(\mathbf{G}_\mathbf{D} = \mathbf{D}^T\mathbf{D}) = \{g_{P\,i,j}\}$$

$$g_{P\,i,j} = \begin{cases} \mathrm{sign}(g_{\mathbf{D}\,i,j})\mu_G & i \neq j \\ 1 & \textit{otherwise} \end{cases},$$

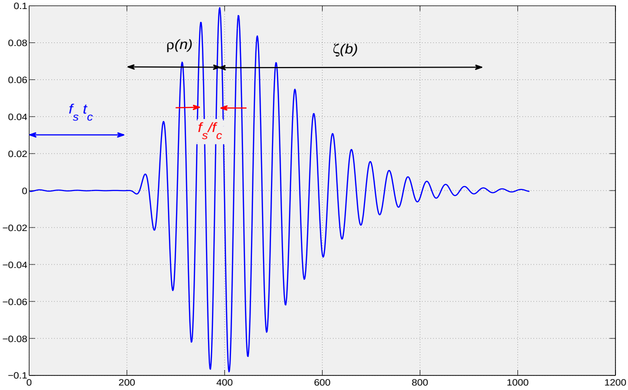
6: *Parameter update* $\Gamma_k \to \Gamma_{k+1}$*:* using a gradient descent method.

Case Study: Gammatone Parametric Dictionary

Gammatone Atom:

$$g(t-t_c) = a.(t-t_c)^{n-1} e^{-2\pi bB(t-t_c)} \cos(2\pi f_c(t-t_c))$$

$B = f_c/Q + b_{min}$, where Q is a constant and b_{min} is the minimum bandwidth.
$\gamma = [t_c \ f_c \ n \ b]^T$ is the set of parameters which could be found using PDD.



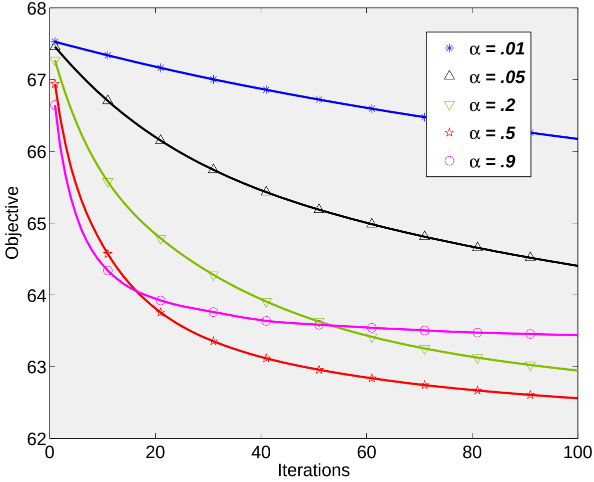
Simulation setting

General Parameters (ETF is plausible with this setting)	
Name	Value
d	1024
N	2048
Q	9.26449
b_{min}	24.7 Hz

Initial Parameters	
Name	Value
f_s	12000 Hz
n	4
b	0.5
f_c^k	$(f_s/2 + Qb_{min})e^{-0.45k/Q} - Qb_{min}$
t_c^l	$t_p + 0.75(l-1)\ t_p$

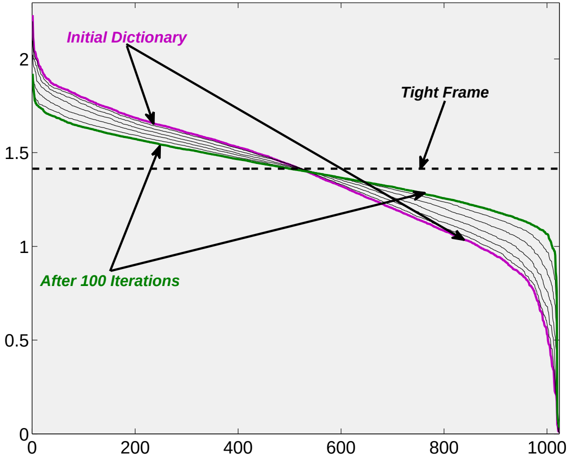
Simulations: Role of α

The objective functions for different $\{\alpha_k\}_{\forall k, \alpha_k = \alpha}$, for a constant α.



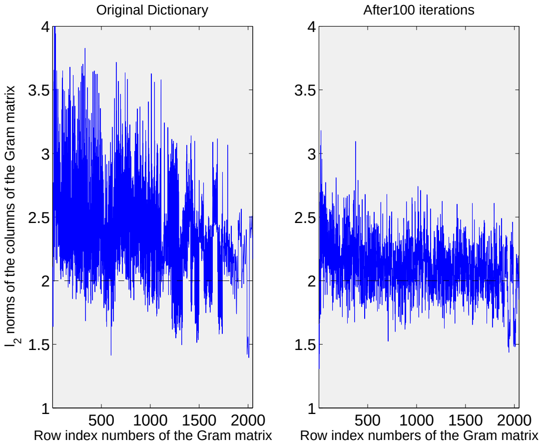
Simulations: Getting Close to Being Tight Frame

Singular values of the dictionaries at some selected iterations.



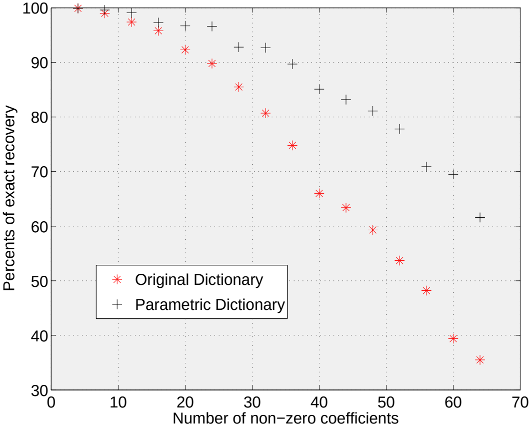
Simulations: Getting Close to Being ETF

The column ℓ_2-norm plots of the Gram matrices $(\sum_j |\langle \mathbf{d}_i, \mathbf{d}_j \rangle|^2)$ of the original (left) and designed (right) dictionaries. The column ℓ_2-norm of $\mathbf{G} \in \Lambda^N$ $(1 + (N-1)\cos^2(\gamma))$ is plotted for reference.



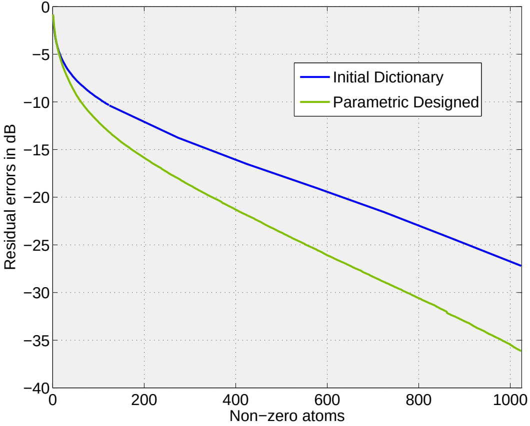
Simulations: Exact Recovery

Exact support recovery of the sparse signals.



Simulations: Residual Error Decay Rate of MP

The residual error plots using matching pursuit for sparse approximation of the audio signal.



Conclusion and Future Work

Conclusion

- A new framework for dictionary design, under a minimum coherence constraint, was presented .
- A difficult optimization problem needs to be solved $\Rightarrow$ a relaxed optimization problem was introduced.
- A practical algorithm was presented to *approximately* solve the relaxed problem.
- Simulations demonstrate that the proposed algorithm not only finds a dictionary which is close to being ETF, but also this dictionary practically shows superior performance in a real application.

Future Work

- Structured Parametric Dictionary Design, to find a fast dictionary.
- Finding a more efficient method for the parameter update step.
- Alternate design criterion.

Thanks for your attention.

Any question ?