Structured and Incoherent Parametric Dictionary Design

Mehrdad Yaghoobi, Laurent Daudet, Mike E. Davies

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

INSPIRE09, September 23, 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: $\widehat{\mathbf{x}} = \arg\min_x \|\mathbf{x}\|_0 \ \text{s.t.}\ \|\mathbf{y} - \mathbf{D}\mathbf{x}\|^2 \leq \xi,$

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

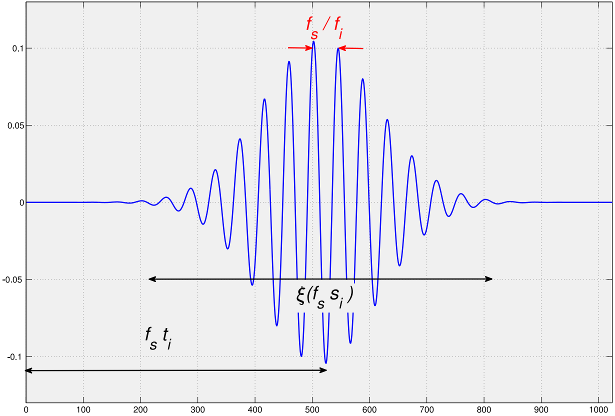
Let Γ be a set of parameters in the admissible set Υ.
$\mathbf{D}_{\Gamma} : \Upsilon \rightarrow \mathbb{R}^{d \times N}$ is defined to be a mapping from the parameter space to the dictionary space.
Given a mapping $\mathbf{D}_{\Gamma}$, a dictionary called Parametric Dictionary, when it is generated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

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



Structured Dictionaries

- Structured Dictionary: A dictionary is structured if the atoms are correlated.
  Examples: shift-invariant, multi-scale, multi-frequency and signature based dictionaries.
- Structured Parametric Dictionary: A parametric dictionary is structured if there exist at least two distinct atoms which depend on a non-empty set of parameters.
  Examples: most of the structured dictionaries have corresponding structured parametric dictionaries. A shift-resilience structure, which let the dictionary be implemented using filter-banks, will be explored here.

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,

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

- *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.

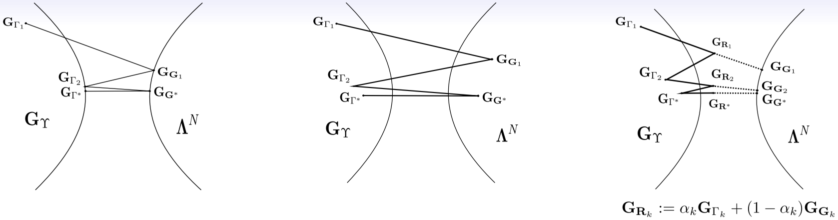
- *Practical Solution:* Convex relaxation of the set of ETF's,

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

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

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

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

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

6: *Parameter update* $\Gamma_k \rightarrow \Gamma_{k+1}$*:* mapping to the dictionary space and optimizing the dictionary in the range space.

Parmeter Update

Parameter Update by Optimization in the Range Space

1: $\mathbf{G}^{\frac{1}{2}} = \Sigma_d^{\frac{1}{2}}\mathbf{U} \;:\; \mathbf{G}_{R_{k+1}} = \mathbf{U}\Sigma\mathbf{U}^T$

2: $\mathbf{A}^* = \mathbf{V}\mathbf{W}^T \;:\; \mathbf{D}_{\Gamma_k}\mathbf{G}^{\frac{T}{2}} = \mathbf{V}\Delta\mathbf{W}^T$

3: $\mathbf{D}_{\Gamma_{k+1}} = \arg\min_{\mathbf{D}\in\mathscr{D}} \|\mathbf{D} - \mathbf{A}^*\mathbf{G}^{\frac{1}{2}}\|_F$

4: Updating Γ_{k+1} with the parameters of $\mathbf{D}_{\Gamma_{k+1}}$

1: A projection of $G_{R_{k+1}} \in \Lambda^N$ onto the space of rank-d symmetric matrices in $\mathbb{R}^{N\times N}$, $\mathcal{S}^+(d,N)$.

$$\mathbf{G}^{\frac{T}{2}}\mathbf{G}^{\frac{1}{2}} = \arg\min_{\mathbf{G}\in\mathcal{S}^+(d,N)} \|\mathbf{G} - \mathbf{G}_{R_{k+1}}\|_F \;:\; \Sigma_{d\;d\times N}^{\frac{1}{2}} = \mathrm{diag}\{\sigma_i^{\frac{1}{2}}\}_{i\in\mathcal{I}},\ |\mathcal{I}| = d.$$

2: $\mathbf{A}^* = \arg\min_{\mathbf{A}\in\mathbf{R}_*^{d\times d}:\mathbf{A}^T\mathbf{A}=\mathbf{I}_d} \|\mathbf{D}_{\Gamma_k} - \mathbf{A}\mathbf{G}^{\frac{1}{2}}\|_F$.

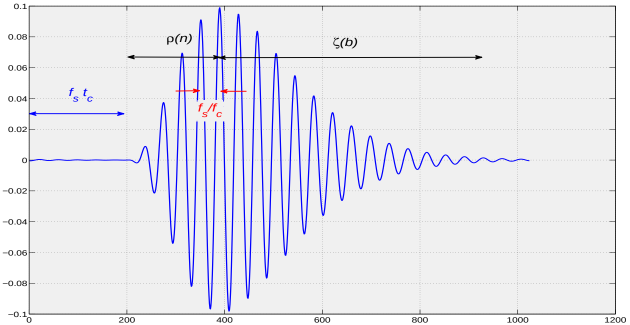
3: Dictionary and parameter update by minimizing the objective with the gradient descent, Newton's or Gauss-Newton's method.

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 in an unstructured PDD.
$t_c = t_0 + l\Delta$ is the generative model in the structured PDD and
$\gamma = [t_0 \ f_c \ n \ b]^T$ is the set of parameters where Δ is fixed.

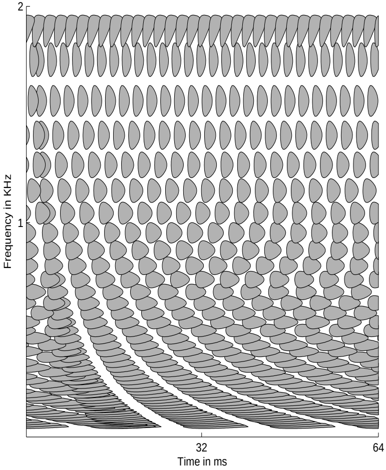


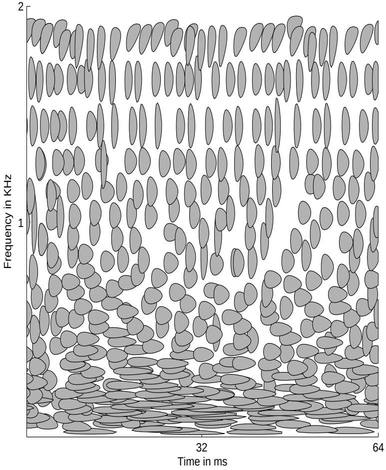
Simulation setting

General Parameters (ETF is plausible with this setting)	
Name	Value
d	256
N	418
α_i	0.5
Q	9.26
b_{min}	24.7 Hz
f_s	4000 Hz

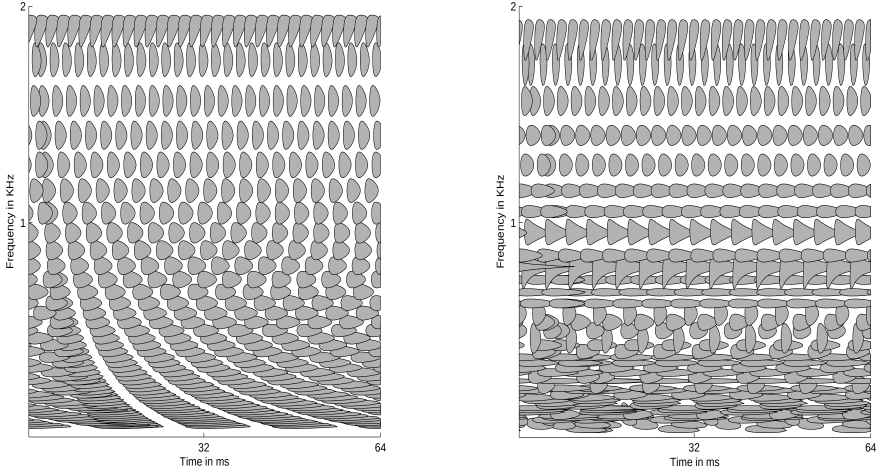
Initial Parameters	
Name	Value
n	4
b	1
f_c^k	$50 + .27kB$
t_0	0

Simulations: Wigner-Ville plot of the unstructured designed dictionary



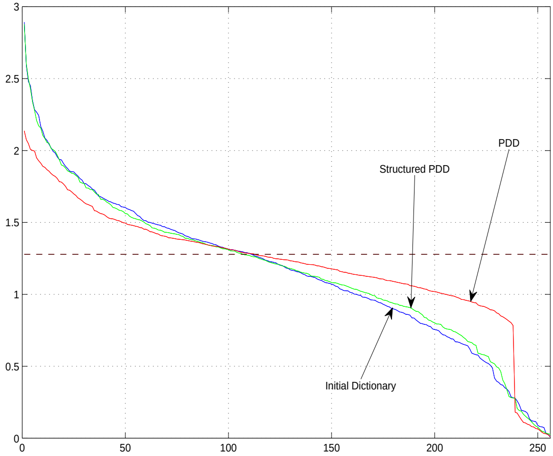


Simulations: Wigner-Ville plot of the structured designed dictionary



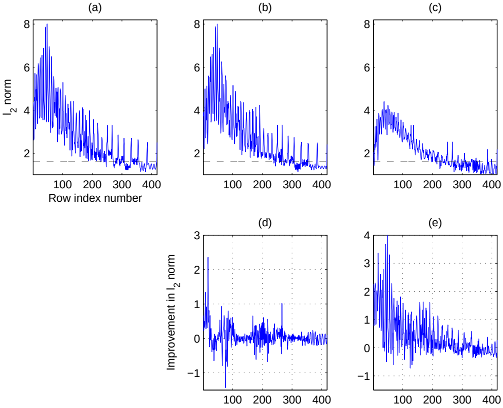
Simulations: Getting Close to Being Tight Frame

Eigenvalues of the Gram matrices.



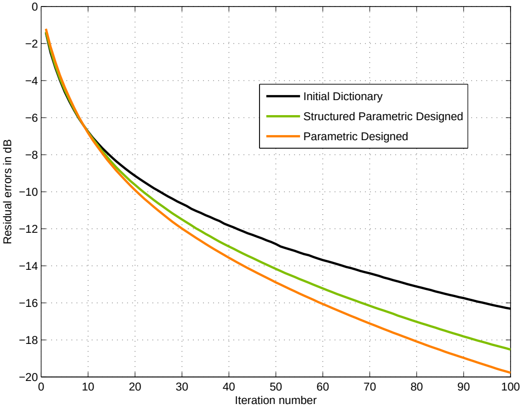
Simulations: Getting Close to Being ETF

The column ℓ_2-norm plots of the Gram matrices $(\sum_j |\langle \mathbf{d}_i, \mathbf{d}_j \rangle|^2)$ of the original (a), structured designed (b) and unstructured (c) dictionaries. The improvement in ℓ_2 norms of the structured designed (d) and unstructured (e) dictionaries.



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

- The PDD was extended to find a structured parametric dictionary.
- As the PDD problem is a difficult optimization problem, a new method for the parameter update step was introduced which let the update be found using conventional optimization algorithms, *e.g.* the gradient descent and Newton's methods.
- The structured and unstructured parametric dictionary design methods was compared with some simulations.
- Simulations on the sparse approximations using MP showed that the structured parametric dictionary improves the error decay rate even though the improvements observed in the algorithm evaluation are small.

Future Work

- Other dictionary structures should be investigated.
- The parametric dictionary model can be applied to the dictionary learning problem.