Learning ancestral atom of structured dictionary via sparse coding

Bernoulli Society Satellite Meeting

Noboru Murata

Waseda University

2 September, 2013

joint work with Toshimitsu Aritake, Hideitsu Hino

sparse coding

- a methodology for representing observations with a sparse combination of basis vectors (atoms)
- related with various problems:
  - associative memory (Palm, 1980)
  - visual cortex model (Olshausen & Field, 1996)
  - Lasso (least absolute shrinkage and selection operator; Tibshirani, 1996)
  - compressive sensing (Candès & Tao, 2006)
  - image restoration/compression (Elad et al., 2005)

- $\boldsymbol{y} = (y_1, \dots, y_n)^T$: target signal
- $\boldsymbol{d} = (d_1, \dots, d_n)^T$: atom
- $\mathbf{D} = (\boldsymbol{d}_1, \dots, \boldsymbol{d}_m)$: dictionary (redundant: $m > n$)
- $\boldsymbol{x} = (x_1, \dots, x_m)^T$: coefficient vector
- objective:

$$\underset{\boldsymbol{x}}{\text{minimize}} \, \|\boldsymbol{y} - \mathbf{D}\boldsymbol{x}\|_2^2 + \eta \|\boldsymbol{x}\|_*$$

where $\|\cdot\|_*$ is a sparse norm, e.g. ℓ_0 and ℓ_1.

dictionary design

dictionary determines overall quality of reconstruction.

- predefined dictionary: structured
  - wavelets (Daubechies, 1992)
  - curvelets (Candès & Donoho, 2001)
  - contourlets (Do & Vetterli, 2005)
- learned dictionary: unstructured
  - gradient-based method (Olshausen & Field, 1996)
  - Method of Optimal Directions (Engan et al., 1999)
  - K-SVD (Aharon et al., 2006)
- structured dictionary learning: intermediate
  - Image Signature Dictionary (Aharon & Elad, 2008)
  - Double Sparsity (Rubinstein et al., 2010)

- ordered dictionary: $\mathbf{D} = (\boldsymbol{d}_\lambda;\, \lambda \in \Lambda)$
- typical approaches:
  - meta dictionary: $\tilde{\mathbf{D}} = (\tilde{d}_1, \ldots, \tilde{d}_M)$

    $$\boldsymbol{d}_\lambda = \tilde{\mathbf{D}}\, \boldsymbol{\alpha}_\lambda$$

    where $\boldsymbol{\alpha}_\lambda$ is a meta-coefficient vector
    additional constraints are imposed on $\boldsymbol{\alpha}_\lambda$, e.g. sparsity
  - ancestral atom (ancestor): $\boldsymbol{a} = (a_1, \ldots, a_N)^T$

    $$\boldsymbol{d}_\lambda = \mathbf{F}_\lambda\, \boldsymbol{a}$$

    where $\mathbf{F}_\lambda$ is an extraction operator

- structure: designed by a set of extraction operators
- extraction operator: $\mathbf{F}_{p,q}$

$$\boldsymbol{d}_{p,q} = \mathbf{F}_{p,q}\,\boldsymbol{a}, \quad (p : \text{scale or downsample level}, \; q : \text{shift})$$

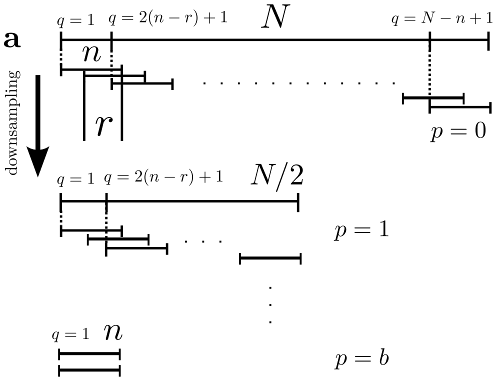
  cut off a piece of ancestor

- generating operator: $\mathbb{D}$

$$\begin{aligned}\mathbb{D}\,\boldsymbol{a} &= (\boldsymbol{d}_{p,q};\,(p,q)\in\Lambda)\\ &= (\mathbf{F}_{p,q}\,\boldsymbol{a};\,(p,q)\in\Lambda)\end{aligned}$$

  a structured collection of $\mathbf{F}_{p,q}$

example of extraction operators



$$[\mathbf{F}_{p,q}]_{ij} = \begin{cases} 1 & j=(i-1)\times 2^p + q, \\ 0 & \text{otherwise,} \end{cases}$$

example of extraction operators

$$\mathbf{F}_{0,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \in \Re^{n \times N}$$

$$\mathbf{F}_{1,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \in \Re^{n \times N}$$

- projection-based algorithm
- related spaces:
  $\mathcal{D}$: space of dictionary
  $\mathcal{S}$: space of structured dictionary
  $\mathcal{A}$: space of ancestor
- related maps:

dictionary learning:	$\mathcal{D} \to \mathcal{D}$
structured dictionary:	$\mathcal{A} \to \mathcal{S} \subset \mathcal{D}$

- introduce a fiber bundle structure to $\mathcal{D}$ by defining projection from $\mathcal{D}$ to $\mathcal{S}$

- condition:

$$\mathbf{G} = \sum_{(p,q)\in\Lambda} \mathbf{F}_{p,q}^T \mathbf{F}_{p,q} : \mathcal{A} \to \mathcal{A} \quad \text{is bijective}$$

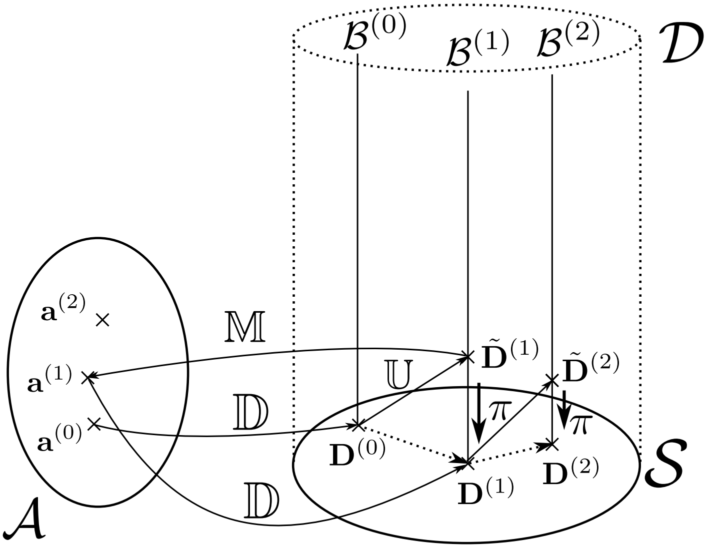
  where $\mathbf{F}_{p,q}^T$ is the adjoint operator of $\mathbf{F}_{p,q}$

- mean operator: $\mathbb{M} : \mathcal{D} \to \mathcal{A}$

$$\begin{aligned} \mathbb{M}\,\mathbf{D} &= \mathbb{M}\,(\boldsymbol{d}_{p,q}; (p,q) \in \Lambda) \\ &= \mathbf{G}^{-1} \sum_{p,q} \mathbf{F}_{p,q}^T\, \boldsymbol{d}_{p,q} \end{aligned}$$

- important relations:

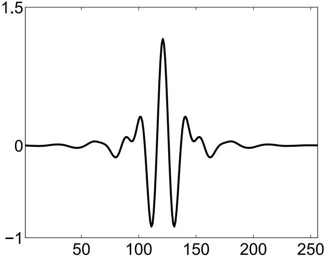
$$\begin{aligned} \boldsymbol{\pi} &= \mathbb{D} \circ \mathbb{M} : \mathcal{D} \to \mathcal{S} \quad \text{(projection)} \\ \mathrm{Id} &= \mathbb{M} \circ \mathbb{D} : \mathcal{A} \to \mathcal{A} \quad \text{(identity)} \end{aligned}$$



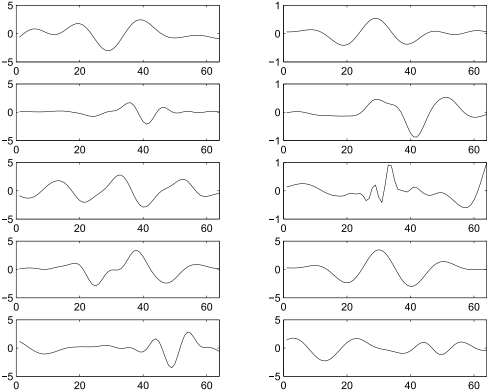
$\mathcal{B}^{(0)}$
$\mathcal{B}^{(1)}$
$\mathcal{B}^{(2)}$
$\mathcal{D}$
$\mathbf{a}^{(2)}$
$\mathbf{a}^{(1)}$
$\mathbf{a}^{(0)}$
$\mathbb{M}$
$\mathbb{U}$
$\mathbb{D}$
$\mathbb{D}$
$\tilde{\mathbf{D}}^{(1)}$
$\tilde{\mathbf{D}}^{(2)}$
π
π
$\mathbf{D}^{(0)}$
$\mathbf{D}^{(1)}$
$\mathbf{D}^{(2)}$
$\mathcal{S}$
$\mathcal{A}$

procedure ANCESTOR LEARNING($\boldsymbol{a}^{(0)}$, $\mathbb{D}$, $\mathbb{M}$, $\mathbb{U}$, $\varepsilon > 0$)
    repeat
        $\mathbf{D}^{(t)} \leftarrow \mathbb{D}\,\boldsymbol{a}^{(t)}$ ▷ generate dictionary
        $\tilde{\mathbf{D}}^{(t+1)} \leftarrow \mathbb{U}\,\mathbf{D}^{(t)}$ ▷ update dictionary
        $\boldsymbol{a}^{(t+1)} \leftarrow \mathbb{M}\,\tilde{\mathbf{D}}^{(t+1)}$ ▷ update ancestor
    until $\|\boldsymbol{a}^{(t+1)} - \boldsymbol{a}^{(t)}\| < \varepsilon$
    return $\boldsymbol{a}$
end procedure

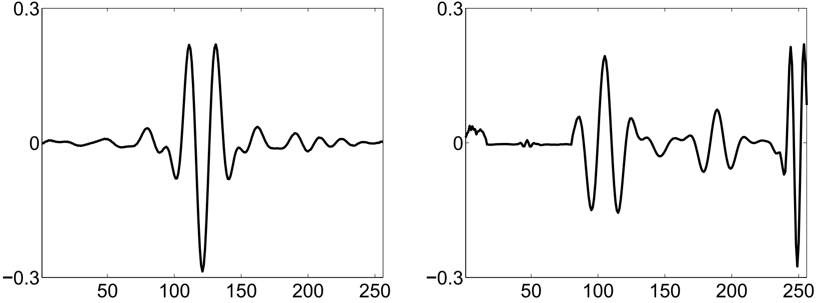
- $\mathbb{U}$: OMP + K-SVD (+ renormalization)
  - OMP: greedy algorithm for ℓ_0 norm SC
  - K-SVD: k-means algorithm for dictionary learning
- compare with ISD (Aharon & Elad, 2008)
  - gradient-based algorithm for estimating ancestor
  - including only shift operation in original version



ancestor (ground truth)



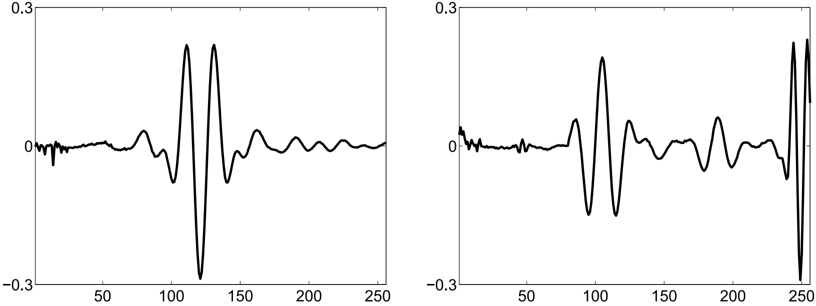
subset of observations



proposed method

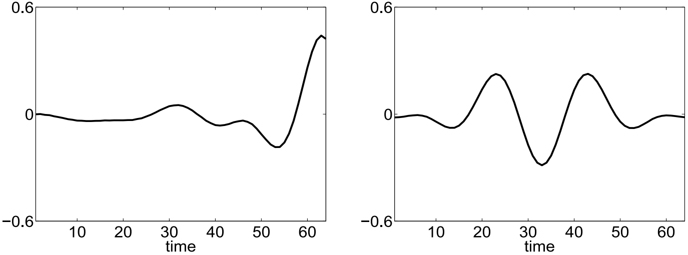
ISD

noiseless case

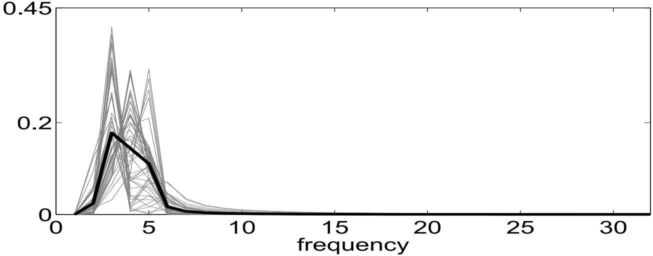


proposed method ISD

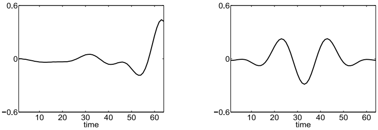
noisy case



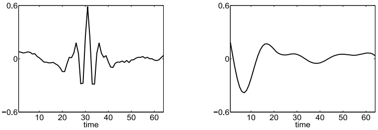
(a-0) examples of atom in level 0 (proposed)



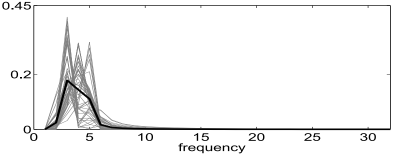
(c-0) spectrum of level 0 (proposed)



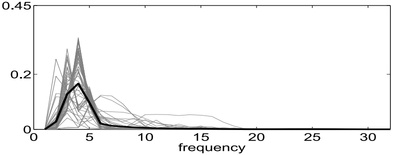
(a-0) examples of atom in level 0 (proposed)



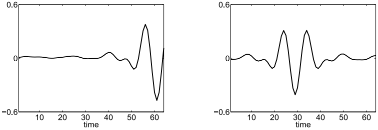
(b-0) examples of atom in level 0 (ISD)



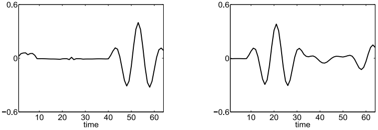
(c-0) spectrum of level 0 (proposed)



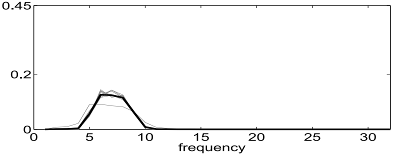
(d-0) spectrum of level 0 (ISD)



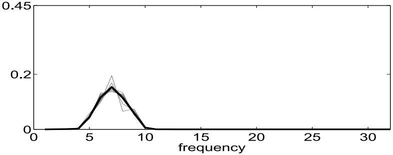
(a-1) examples of atom in level 1 (proposed)



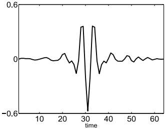
(b-1) examples of atom in level 1 (ISD)



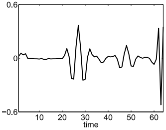
(c-1) spectrum of level 1 (proposed)



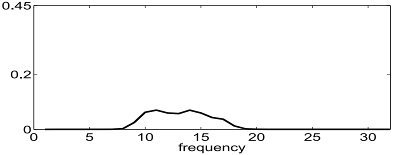
(d-1) spectrum of level 1 (ISD)



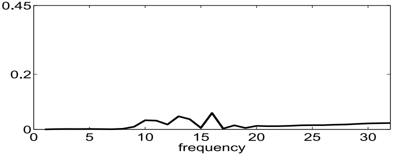
(a-2) examples of atom in level 2 (proposed)



(b-2) examples of atom in level 2 (ISD)



(c-2) spectrum of level 2 (proposed)

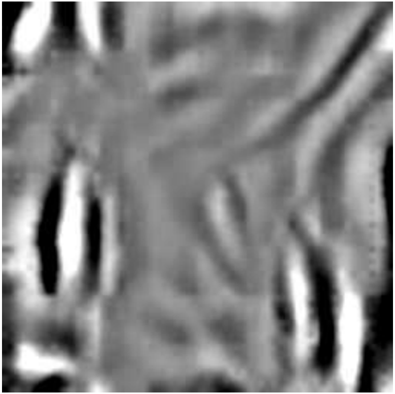


(d-2) spectrum of level 2 (ISD)

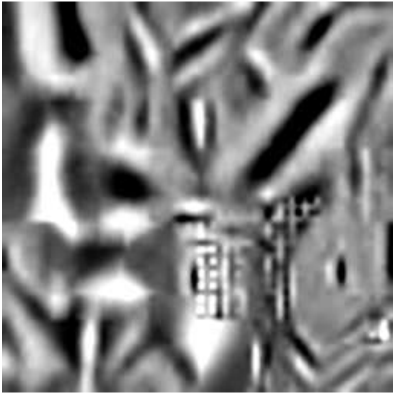
images (2D atoms)

training image

test image (peppers)

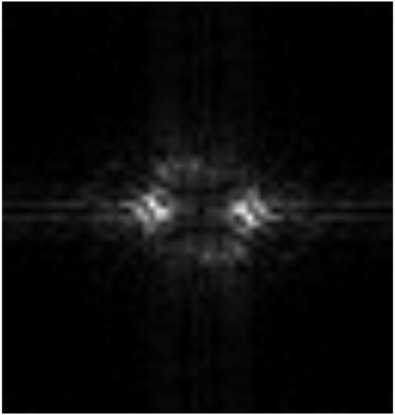


proposed method

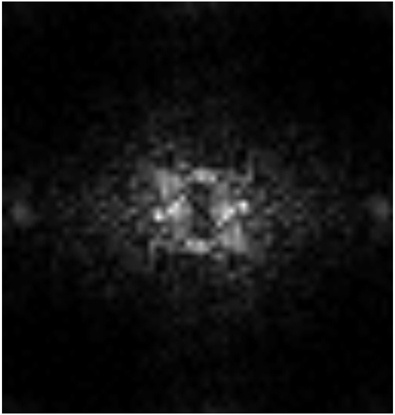


ISD

estimated ancestor

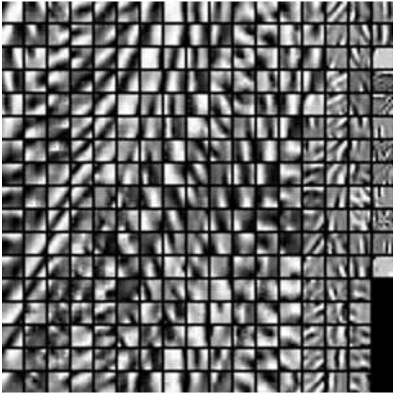


proposed method

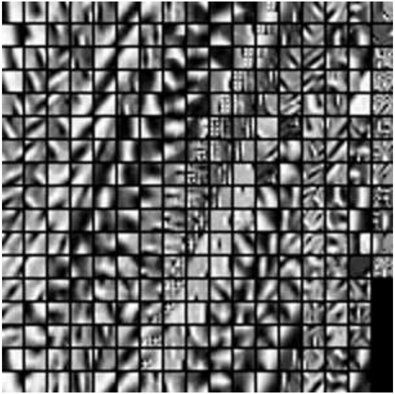


ISD

spectrum of estimated ancestor



proposed method



ISD

learned dictionary

concluding remarks

we proposed

- a dictionary generation scheme from an ancestor
- a condition of structured dictionary identifiabilty
- a projection-based algorithm to learn the ancestor

possible application would be

- image analysis and compression
- frequency analysis of signals