Fast greedy algorithms for dictionary selection with generalized sparsity constraints

Kaito Fujii & **Tasuku Soma** (UTokyo)

Neural Information Processing Systems 2018, spotlight presentation

Dec. 7, 2018

Dictionary

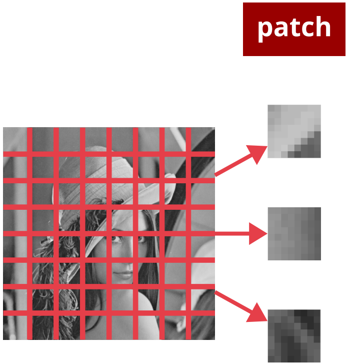
If real-world signals consist of a few patterns,

a "good" dictionary gives sparse representations of each signal

Dictionary

If real-world signals consist of a few patterns,

a "good" dictionary gives sparse representations of each signal

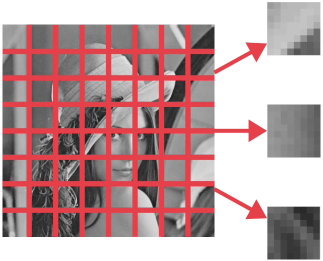


Dictionary

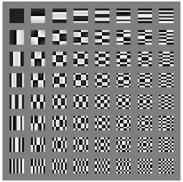
If real-world signals consist of a few patterns,

a "good" dictionary gives sparse representations of each signal

patch



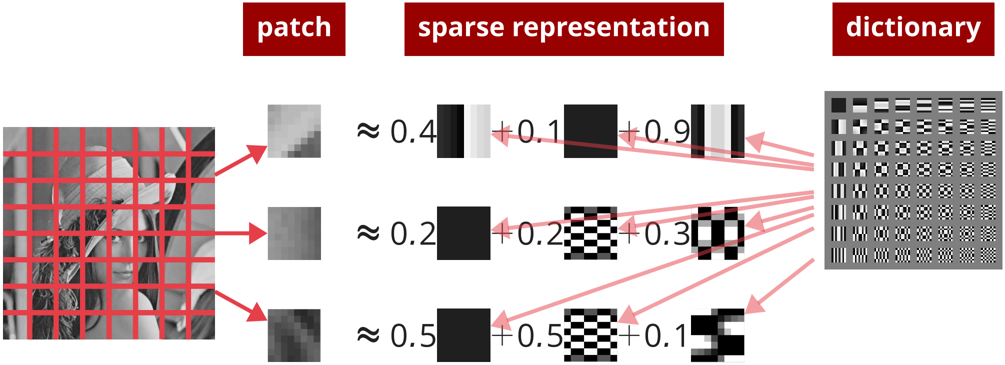
dictionary



Dictionary

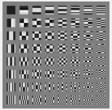
If real-world signals consist of a few patterns,

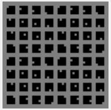
a "good" dictionary gives sparse representations of each signal

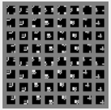


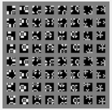
Dictionary selection [Krause–Cevher'10]

Union of existing dictionaries


DCT basis


Haar basis


Db4 basis


Coiflet basis

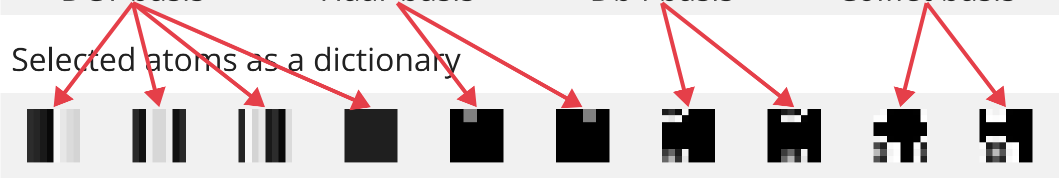
Selected atoms as a dictionary

Atoms for each patch $\mathbf{y}_t$ $(\forall t \in [T])$

Dictionary selection [Krause–Cevher'10]

Union of existing dictionaries

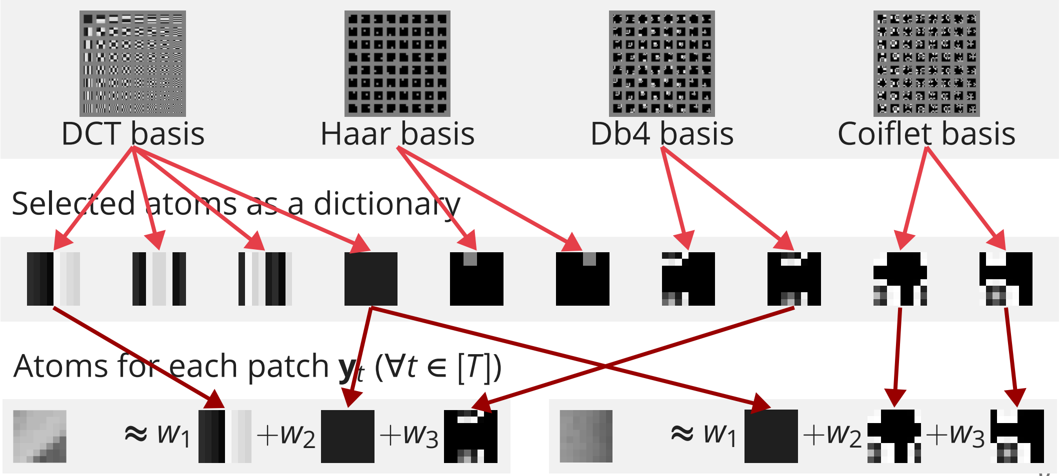
Selected atoms as a dictionary



Atoms for each patch $\mathbf{y}_t$ ($\forall t \in [T]$)

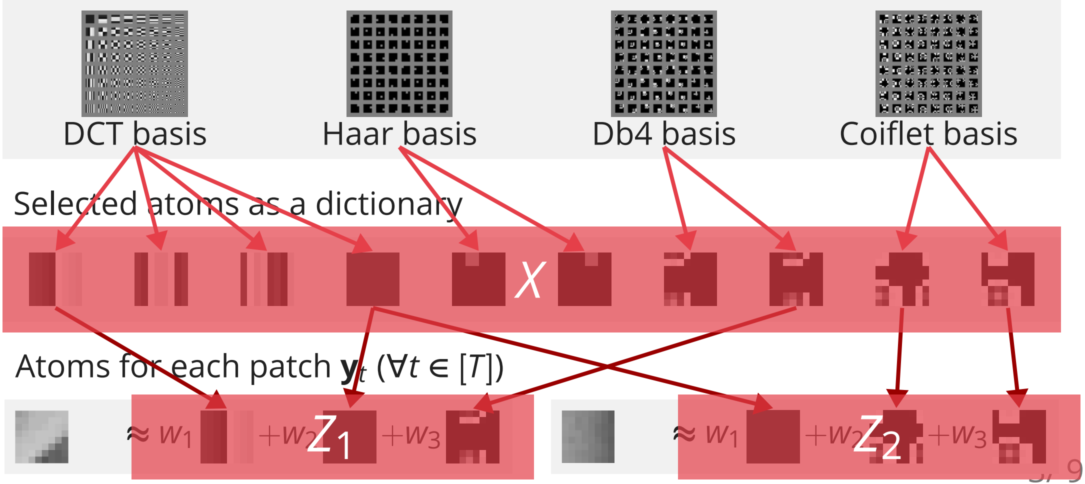
Dictionary selection [Krause–Cevher'10]

Union of existing dictionaries



Dictionary selection [Krause–Cevher'10]

Union of existing dictionaries



Dictionary selection with sparsity constraints

$$\underset{X \subseteq V}{\text{Maximize}} \quad \max_{(Z_1,\cdots,Z_T) \in \mathcal{I}:\, Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \leq k$$

1st maximization:

selecting a set X of atoms as a dictionary

Dictionary selection with sparsity constraints

$$\text{Maximize}_{X \subseteq V} \max_{(Z_1,\cdots,Z_T)\in\mathcal{I}:\, Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \le k$$

2nd maximization:

selecting a set $Z_t \subseteq X$ of atoms

for a sparse representation of each patch

under sparsity constraint $\mathcal{I}$

Dictionary selection with sparsity constraints

$$\underset{X \subseteq V}{\text{Maximize}} \quad \max_{(Z_1,\cdots,Z_T)\in\mathcal{I}:\, Z_t\subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \le k$$

sparsity constraint

set function representing the quality of Z_t for patch $\mathbf{y}_t$

Dictionary selection with sparsity constraints

$$\underset{X \subseteq V}{\text{Maximize}} \quad \max_{(Z_1, \cdots, Z_T) \in \mathcal{I}:\, Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \leq k$$

sparsity constraint

set function representing the quality of Z_t for patch $\mathbf{y}_t$

Our contributions

1. **Replacement OMP**:

   A fast greedy algorithm with approximation ratio guarantees

Dictionary selection with sparsity constraints

$$\underset{X \subseteq V}{\text{Maximize}} \quad \max_{(Z_1, \cdots, Z_T) \in \mathcal{I}:\, Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \leq k$$

sparsity constraint

set function representing the quality of Z_t for patch $\mathbf{y}_t$

Our contributions

1. **Replacement OMP**:

   A fast greedy algorithm with approximation ratio guarantees

2. ***p*-Replacement sparsity families**:

   A novel class of sparsity constraints generalizing existing ones

1 Replacement OMP

Replacement Greedy for two-stage submodular maximization [Stan+'17]

❶ Replacement OMP

Replacement Greedy for two-stage submodular maximization [Stan+'17]

↓ 1st result application to dictionary selection

Replacement Greedy $O(s^2dknT)$ running time

1 Replacement OMP

Replacement Greedy for two-stage submodular maximization [Stan+'17]

1st result application to dictionary selection

Replacement Greedy $O(s^2 dknT)$ running time

2nd result $O(s^2 d)$ acceleration with the concept of OMP

Replacement OMP $O((n + ds)kT)$ running time

① Replacement OMP

algorithm	approximation ratio	running time	empirical performance
SDS_{MA} [Krause–Cevher'10]	✓	✓	
SDS_{OMP} [Krause–Cevher'10]			✓
Replacement Greedy	✓		✓
Replacement OMP	✓	✓	✓

❷ *p*-Replacement sparsity families

average sparsity
[Cevher–Krause'11]

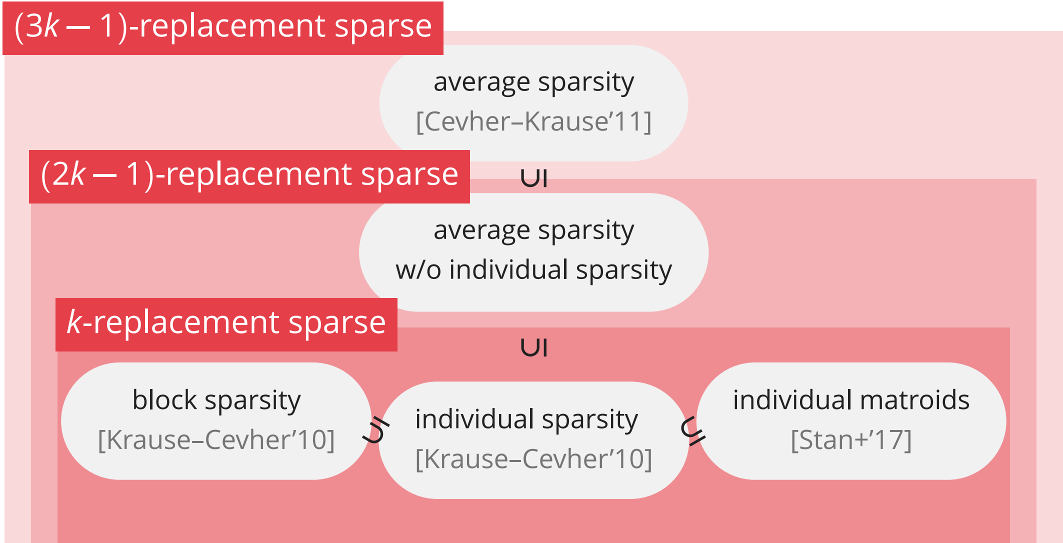
average sparsity
w/o individual sparsity

block sparsity
[Krause–Cevher'10]

individual sparsity
[Krause–Cevher'10]

individual matroids
[Stan+'17]

❷ *p*-Replacement sparsity families



2 p-Replacement sparsity families

We extend Replacement OMP to p-replacement sparsity families

Theorem

Replacement OMP achieves $\frac{m_{2s}^2}{M_{s,2}^2}\left(1-\exp\left(-\frac{k}{p}\frac{M_{s,2}}{m_{2s}}\right)\right)$-approximation

if $\mathcal{I}$ is p-replacement sparse

Assumption

$$f_t(Z_t) \triangleq \max_{\mathbf{w}_t:\, \mathrm{supp}(\mathbf{w}_t)\subseteq Z_t} u_t(\mathbf{w}_t)$$

where u_t is m_{2s}-strongly concave on $\Omega_{2s} = \{(\mathbf{x},\mathbf{y}) \colon \|\mathbf{x}-\mathbf{y}\|_0 \leq 2s\}$

and $M_{s,2}$-smooth on $\Omega_{s,2} = \{(\mathbf{x},\mathbf{y}) \colon \|\mathbf{x}\|_0 \leq s, \|\mathbf{y}\|_0 \leq s, \|\mathbf{x}-\mathbf{y}\|_0 \leq 2\}$

Overview

1. **Replacement OMP**: A fast algorithm for dictionary selection
2. ***p*-Replacement sparsity families**: A class of sparsity constraints

Other contributions

- Empirical comparison with dictionary learning methods
- Extensions to *online dictionary selection*

Poster #78 at Room 210 & 230 AB, Thu 10:45–12:45