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.
If real-world signals consist of a few patterns, a "good" dictionary gives sparse representations of each signal.
If real-world signals consist of a few patterns, a “good” dictionary gives sparse representations of each signal.
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 $y_t$ ($\forall t \in [T]$)
Dictionary selection [Krause–Cevher’10]

### Union of existing dictionaries

<table>
<thead>
<tr>
<th>Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCT basis</td>
</tr>
<tr>
<td>Haar basis</td>
</tr>
<tr>
<td>Db4 basis</td>
</tr>
<tr>
<td>Coiflet basis</td>
</tr>
</tbody>
</table>

**Selected atoms as a dictionary**

**Atoms for each patch** $y_t (\forall t \in [T])$
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 $y_t$ ($\forall t \in [T]$)
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 $y_t$ ($\forall t \in [T]$)

$\approx w_1 + w_2 + w_3$
Dictionary selection with sparsity constraints

Maximize \( \prod_{x \in V} \) \( \max_{(Z_1, \ldots, Z_T) \in \mathcal{I}: Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \) subject to \( |X| \leq k \)

**1st maximization:**
selecting a set \( X \) of atoms as a dictionary

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
Dictionary selection with sparsity constraints

Maximize \( \max_{X \subseteq V} \sum_{t=1}^{T} f_t(Z_t) \) subject to \(|X| \leq k\)

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

Maximize \( \min_{X \subseteq V} \sum_{t=1}^{T} f_t(Z_t) \) 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
Dictionary selection with sparsity constraints

Maximize \( \sum_{t=1}^{T} f_t(Z_t) \) subject to \( |X| \leq k \)

sparsity constraint

set function representing the quality of \( Z_t \) for patch \( y_t \)

Our contributions

1. Replacement OMP:
   A fast greedy algorithm with approximation ratio guarantees
Dictionary selection with sparsity constraints

Maximize \( X \subseteq V \)
\[
\max_{(Z_1, \ldots, Z_T) \in \mathcal{I}: Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t)
\]
subject to \( |X| \leq k \)

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

1st result: application to dictionary selection

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

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

Replacement OMP \(O((n + ds)kT)\) running time
# Replacement OMP

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Approximation Ratio</th>
<th>Running Time</th>
<th>Empirical Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SDS_{MA}$ [Krause–Cevher’10]</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>$SDS_{OMP}$ [Krause–Cevher’10]</td>
<td></td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Replacement Greedy</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Replacement OMP</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
$p$-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]
$p$-Replacement sparsity families

(3$k - 1$)-replacement sparse

(2$k - 1$)-replacement sparse

$k$-replacement sparse

- 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]
We extend Replacement OMP to $p$-replacement sparsity families

**Theorem**

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

**Assumption**

\[
 f_t(Z_t) \triangleq \max_{w_t : \text{supp}(w_t) \subseteq Z_t} u_t(w_t)
\]

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

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