Fast greedy algorithms for dictionary selection with generalized sparsity constraints

Kaito Fujii & Tasuku Soma (University of Tokyo)

The 32nd Annual Conference on Neural Information Processing Systems

Our contributions

- We propose a fast greedy algorithm for dictionary selection, named **Replacement OMP**.
- We propose a novel class of sparsity constraints, ***p*-replacement sparsity families**.
- We empirically show that in a smaller time, Replacement OMP returns a solution competitive with ones obtained by dictionary learning methods.
- We extend the algorithms to the online setting.

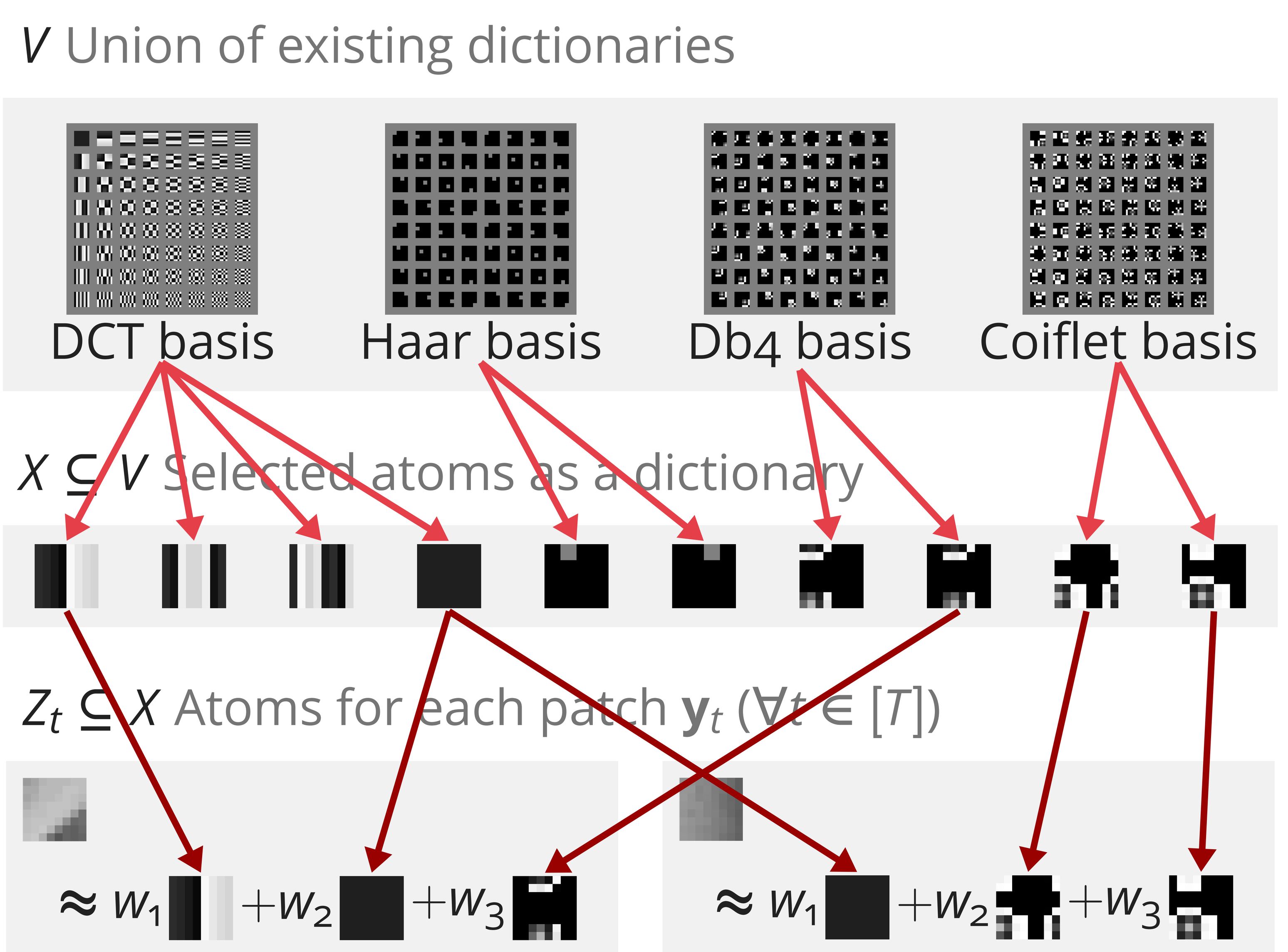
Dictionary selection

sparsity constraint e.g. $\mathcal{I} = \{(Z_1, \cdots, Z_T) : |Z_t| \le s\}$

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

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

$$f_t(Z_t) \triangleq \max_{\mathbf{w}_t:\, \text{supp}(\mathbf{w}_t)\subseteq Z_t} u_t(\mathbf{w}_t) \qquad \text{e.g. } u_t(\mathbf{w}_t) = \|\mathbf{y}_t\|^2 - \|\mathbf{y}_t - \mathbf{A}\mathbf{w}_t\|^2$$



Replacement Greedy & Replacement OMP

Assumption

u_t is m_{2s}-strongly concave on $\Omega_{2s} = \{(\mathbf{x}, \mathbf{y}) : \|\mathbf{x} - \mathbf{y}\|_0 \le 2s\}$ and $M_{s,2}$-smooth on $\Omega_{s,2} = \{(\mathbf{x}, \mathbf{y}) : \|\mathbf{x}\|_0 \le s, \|\mathbf{y}\|_0 \le s, \|\mathbf{x} - \mathbf{y}\|_0 \le 2\}$

1 Initialize $X := \emptyset$, $Z_t := \emptyset$ $(\forall t \in [T])$

2 Greedy atom selection (Repeat k times)

Pick $a^* \in V$ that maximizes

$$\begin{cases} \max_{(Z'_1,\cdots,Z'_T)\in\mathcal{F}_{a^*}} \sum_{t=1}^{T} \{f_t(Z'_t) - f_t(Z_t)\} & \text{(Replacement Greedy)} \\ \max_{(Z'_1,\cdots,Z'_T)\in\mathcal{F}_{a^*}} \left\{ \frac{1}{M_{s,2}} \sum_{t=1}^{T} \left\| \nabla u_t\left(\mathbf{w}_t^{(Z_t)}\right)_{Z'_t \setminus Z_t} \right\|^2 - M_{s,2} \sum_{t=1}^{T} \left\| \left(\mathbf{w}_t^{(Z_t)}\right)_{Z_t \setminus Z'_t} \right\|^2 \right\} & \text{(Replacement OMP)} \end{cases}$$

and let $X \leftarrow X + a^*$ and $Z_t \leftarrow Z'_t$ $(\forall t \in [T])$

$$\mathcal{F}_a(Z_1, \cdots, Z_T) \triangleq \left\{(Z'_1, \cdots, Z'_T) \in \mathcal{I} \mid Z'_t \subseteq Z_T + a,\ |Z_t \setminus Z'_t| \le 1\ (\forall t \in [T])\right\}$$

feasible replacements

$\mathbf{w}_t^{(Z_t)} \in \operatorname{argmax}_{\mathbf{w}:\, \text{supp}(\mathbf{w})\subseteq Z_t} u_t(\mathbf{w})$ optimal vector for u_t with support Z_t

Theoretical results for standard dictionary selection

algorithm	approximation ratio	running time
SDS_{MA} [Krause–Cevher'10]	$\frac{m_1 m_s}{M_1 M_s}(1 - 1/e)$	$O((k + d)nT)$
SDS_{OMP} [Krause–Cevher'10]	$O(1/k)$	$O((s + k)sdknT)$
Replacement Greedy	$\frac{m_{2s}^2}{M_{s,2}^2}\left(1 - \exp\left(-\frac{M_{s,2}}{m_{2s}}\right)\right)$	$O(s^2 dknT)$
Replacement OMP	$\frac{m_{2s}^2}{M_{s,2}^2}\left(1 - \exp\left(-\frac{M_{s,2}}{m_{2s}}\right)\right)$	$O((n + ds)kT)$

- Replacement Greedy was originally proposed for two-stage submodular maxization in [Stan+'17]
- We extend SDS_{MA}, Replacement Greedy, and Replacement OMP to the online setting

p-Replacement sparsity families

A generalized class of sparsity constraints including average/block sparsity [Cevher–Krause'11]

Theorem

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

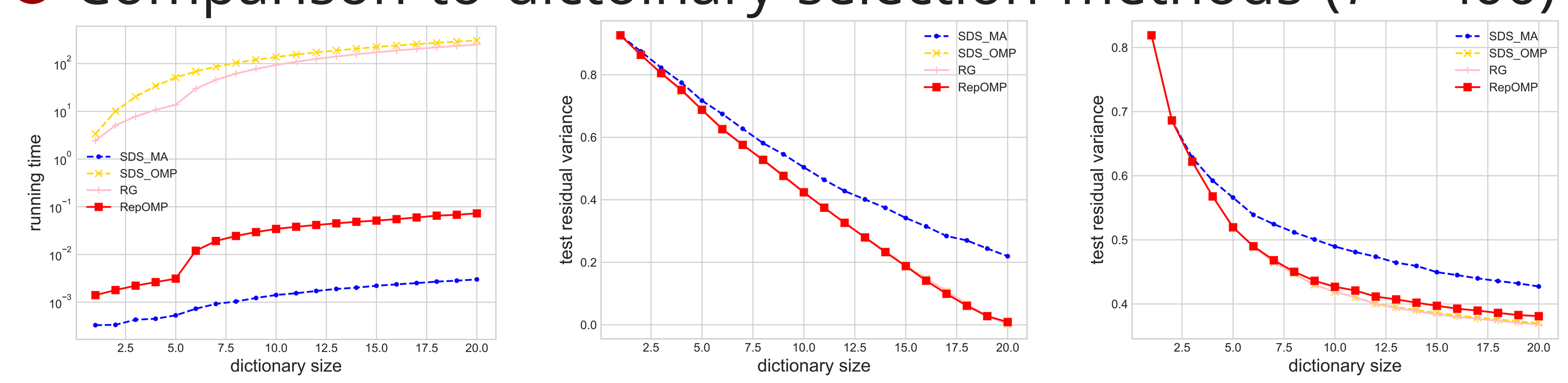
$\forall (Z_1, \ldots, Z_T), (Z_1^*, \ldots, Z_T^*) \in \mathcal{I}$,

$\exists (Z_1^{p'}, \ldots, Z_T^{p'}) \in \bigcup_{a\in V} \mathcal{F}_a(Z_1, \ldots, Z_T)$ $(p' \in [p])$ s.t.

- each atom in $Z_t^* \setminus Z_t$ appears at least once in $(Z_t^{p'} \setminus Z_t)_{p'=1}^{p}$
- each atom in $Z_t \setminus Z_t^*$ appears at most once in $(Z_t \setminus Z_t^{p'})_{p'=1}^{p}$

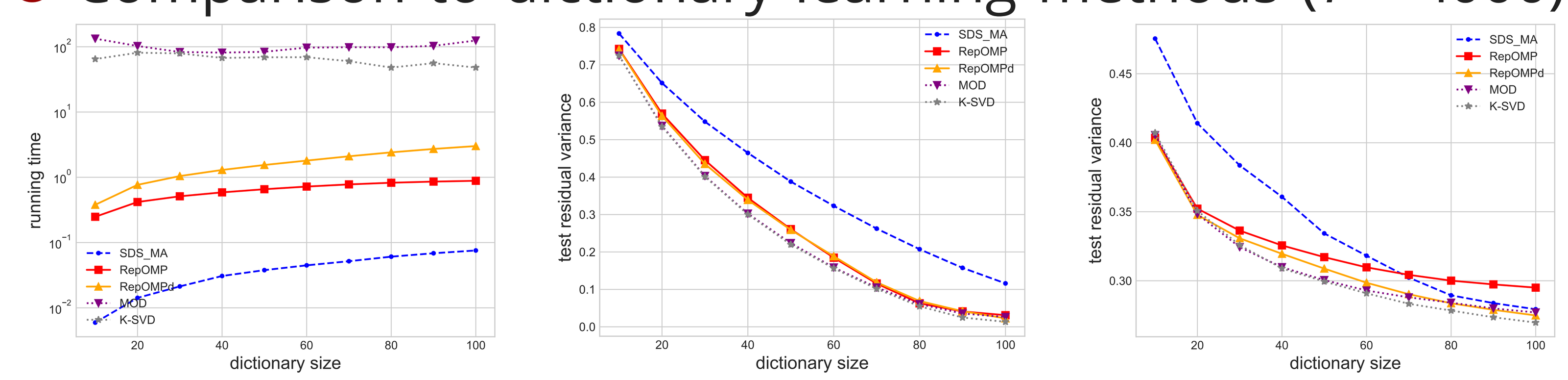
Experimental results

- Comparison to dictoinary selection methods ($T = 100$)



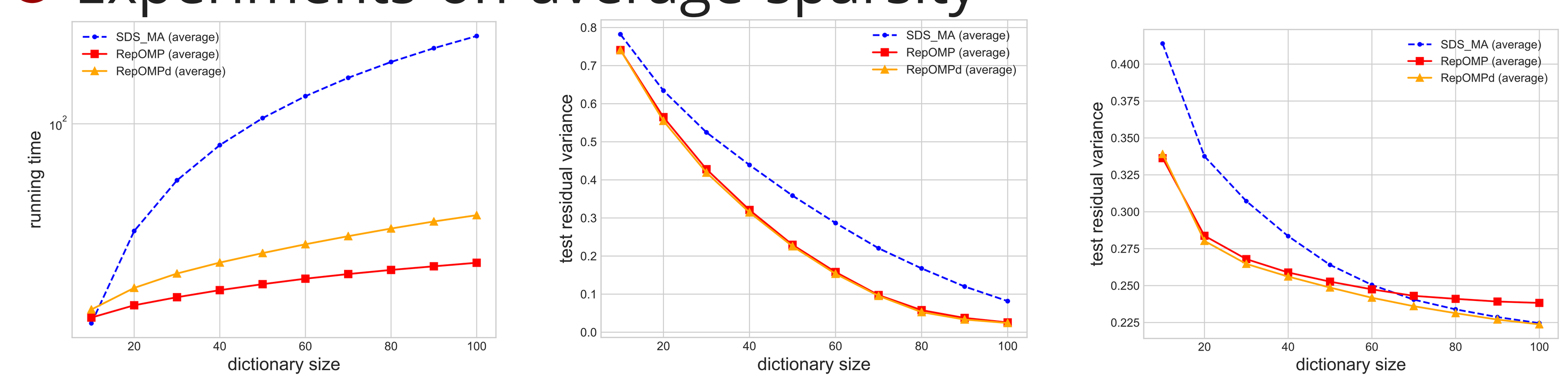
time, synthetic — residual, synthetic — residual, real

- Comparison to dictionary learning methods ($T = 1000$)



time, synthetic — residual, synthetic — residual, real

- Experiments on average sparsity



time, synthetic — residual, synthetic — residual, real