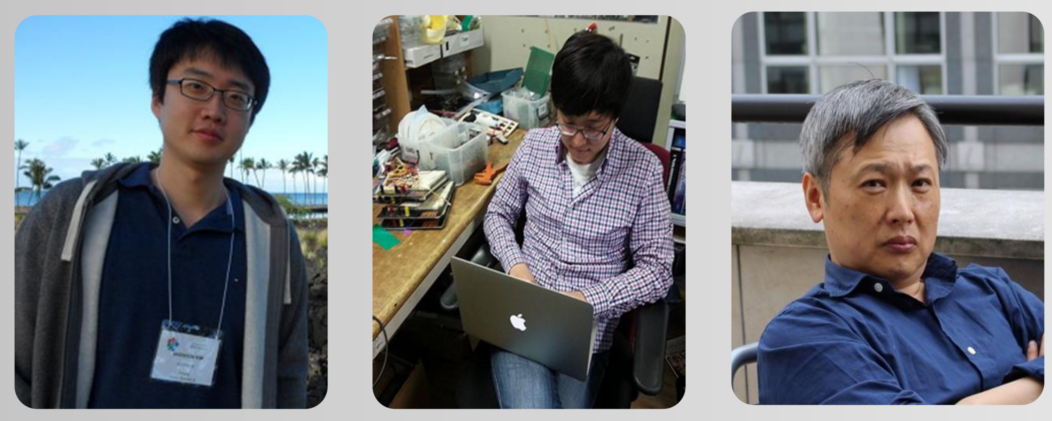
On the Optimal Bit Complexity of Circulant Binary Embedding

The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)

Saehoon Kim, Jungtaek Kim and Seungjin Choi

Machine Learning Group, Department of Computer Science and Engineering,
Pohang University of Science and Technology (POSTECH)



Saehoon Kim Jungtaek Kim Seungjin Choi

Motivation

- *Circulant Binary Embedding works well* with nearly linear time and space complexities
- Theoretical Justifications on circulant binary embedding are not sufficiently studied

Contribution

- **We develop *a non-trivial extension of existing analysis* to achieve the optimal bit complexity of CBE**
- **Our analysis is well matched to the original implementation of CBE and its empirical justification**

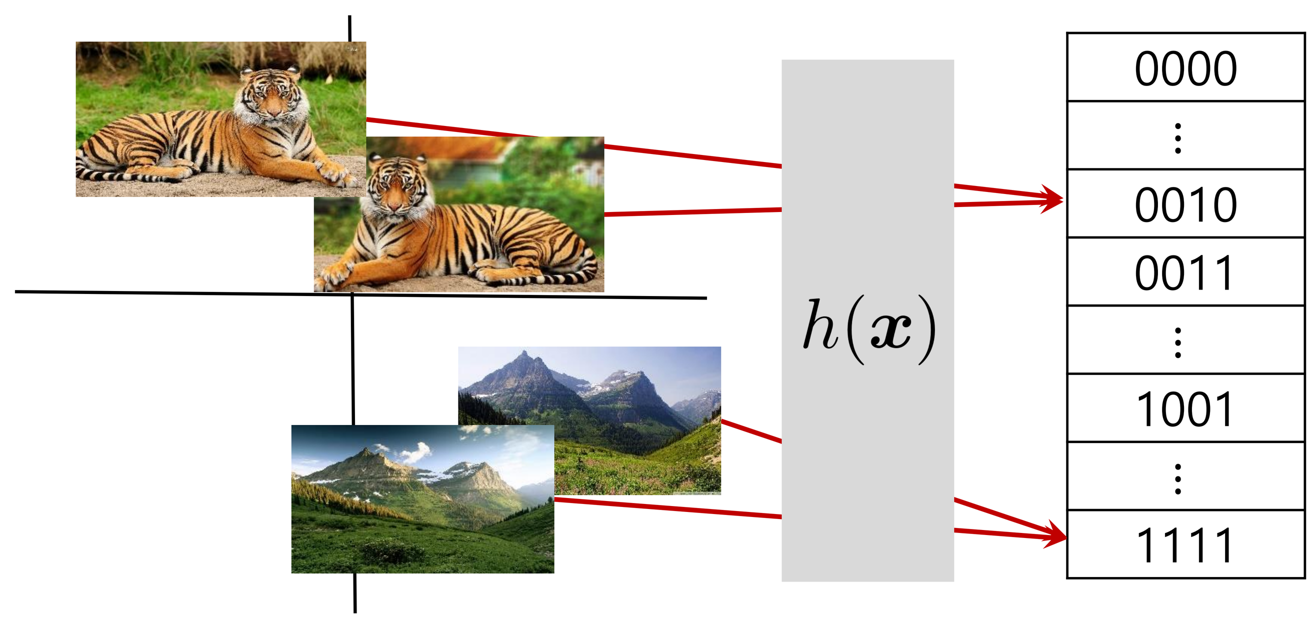
Circulant Binary Embedding

For $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n] \in \mathcal{S}^{(d-1)\times n}$, circulant binary embedding refers to methods for embedding points in $\mathcal{S}^{d-1}$ into vertices in the Hamming cube of dimension k, such that $\forall i, j \in \{1, \dots, n\}$

$$\mathrm{HammingDist}\left(h^C(\boldsymbol{x}_i), h^C(\boldsymbol{x}_j)\right) = \frac{\theta_{\boldsymbol{x}_i,\boldsymbol{x}_j}}{\pi},$$

where $h^C(\boldsymbol{x}_i) = \mathrm{sgn}\left(\boldsymbol{G}_c^\top \boldsymbol{D}\boldsymbol{x}_i\right)$, $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with a Rademacher sequence and $\boldsymbol{G}_c \in \mathbb{R}^{d\times d}$ is a circulant matrix.

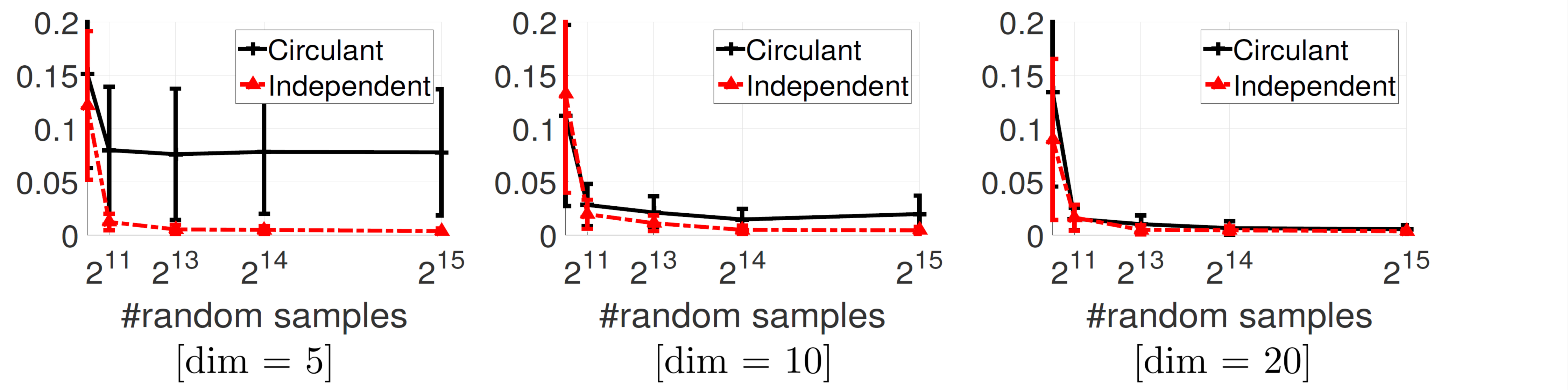
Application



Circulant Matrix

$$\boldsymbol{G}_c = \begin{pmatrix} g_1 & g_d & \cdots & g_3 & g_2 \\ g_2 & g_1 & \cdots & g_4 & g_3 \\ \vdots & g_2 & g_1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & g_d \\ g_d & g_{d-1} & \cdots & g_2 & g_1 \end{pmatrix}$$

When Does It Work?



Bit Complexity for Performance Comparison

Definition 1. *Given $\epsilon \in (0,1)$ and any finite set of d-dimensional vectors, $\mathcal{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, a mapping $h : \mathcal{S}^{d-1} \to \{0,1\}^k$ is said to be an ϵ-distortion* binary embedding *if*

$$\left| d_H(h(\boldsymbol{x}_i), h(\boldsymbol{x}_j)) - \frac{\theta_{\boldsymbol{x}_i,\boldsymbol{x}_j}}{\pi} \right| \le \epsilon,$$

for $\forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}$.

Theorem 1. *Given $\epsilon \in (0,1)$ and any finite data set $\mathcal{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \subset \mathcal{S}^{d-1}$, with probability at least $1 - \exp(-c\epsilon^2 k)$, $k = \mathcal{O}\left(\frac{1}{\epsilon^2}\log n\right)$ implies that we have $h : \mathcal{S}^{d-1} \to \{0,1\}^k$ such that for all $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}$*

$$\left| d_H(h(\boldsymbol{x}_i), h(\boldsymbol{x}_j)) - \frac{\theta_{\boldsymbol{x}_i,\boldsymbol{x}_j}}{\pi} \right| \le \epsilon,$$

where $c > 0$ is a constant.

Table 1: Comparison of existing analyses (unstructured BE and CBE)

Methods	Bit Complexity	Conditions
Unstructured BE	$\mathcal{O}\left(\epsilon^{-2}\log n\right)$	-
Our analysis	$\mathcal{O}\left(\epsilon^{-2}\log n\right)$	*small infinity norm*
Arxiv'16 (Near-optimal)	$\mathcal{O}\left(\epsilon^{-3}\log n\right)$	*small infinity norm*
Arxiv'15 (Near-optimal)	$\mathcal{O}\left(\epsilon^{-2}\log^2 n\right)$	*small infinity norm*

Our Main Analysis

Condition 1. *Suppose that we have $\mathcal{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \subset \mathcal{S}^{d-1}$. Letting $\rho \triangleq \sup_{1\le i \le n} \|\boldsymbol{x}_i\|_\infty$, there exist nonnegative constants such that*

- $c_2 \epsilon k \rho \log d < 1$.
- $c_3 \rho k < \epsilon$.
- $c_4 k^3 \rho^2 \epsilon^2 < 1$,

where k is #bits, n is #data points, and d is the data dimension.

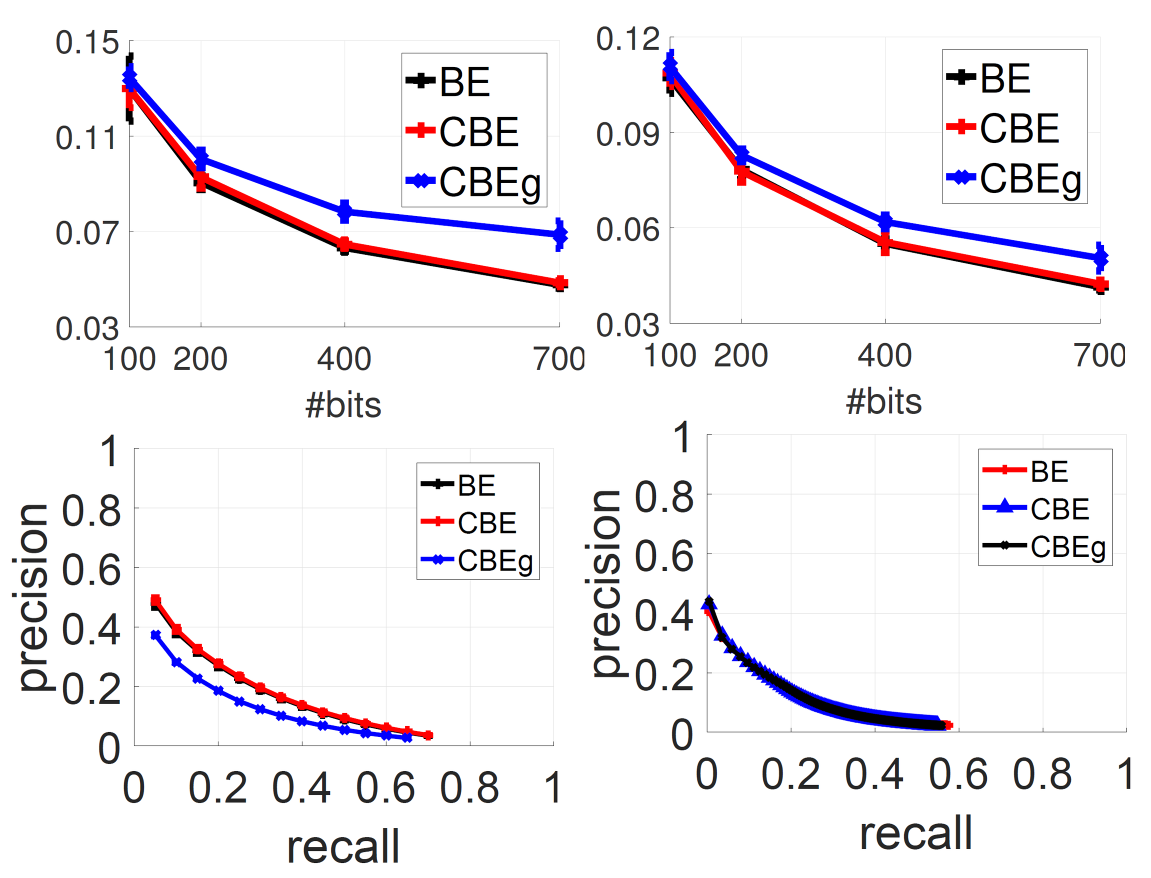
Theorem 2. *Given $\epsilon \in (0,1)$ and any finite dataset $\mathcal{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} \subset \mathcal{S}^{d-1}$, under Condition 1, with prob. at least $1 - \exp(-c_5\epsilon^2 k)$, $k = \mathcal{O}\left(\epsilon^{-2}\log n\right)$ implies that CBE guarantees ϵ-distortion binary embedding such that for all $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{D}$*

$$\left| d_H(h^C(\boldsymbol{x}_i), h^C(\boldsymbol{x}_j)) - \frac{\theta_{\boldsymbol{x}_i,\boldsymbol{x}_j}}{\pi} \right| \le \epsilon,$$

where $c_5 > 0$ is a constant.

Detailed proofs available in the paper

Experiments on Several Datasets



Angle preservation (MNIST and CIFAR-10)

NN search (GIST1M and Flickr45K)

References

- Yu, F. X. et al, Circulant binary embedding, ICML'14
- Yu, F. X. et al, On binary embedding using circulant matrices, Arxiv'15
- Oymak, S. Near-optimal sample complexity bounds for circulant binary embedding, Arxiv'16