

# Two-way kernel matrix puncturing: towards resource-efficient PCA and spectral clustering

Romain COUILLET, Florent CHATELAIN, Nicolas LE BIHAN

*MIAI LargeDATA Chair, GIPSA-lab, University Grenoble-Alpes, France.*

June 11, 2021

## “One-way” kernel puncturing: Improving over subsampling

► **Setting:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$

$$\mathbf{K} = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

## “One-way” kernel puncturing: Improving over subsampling

► Setting:  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$

→ data subsampling  $\mathbf{K}_{ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$

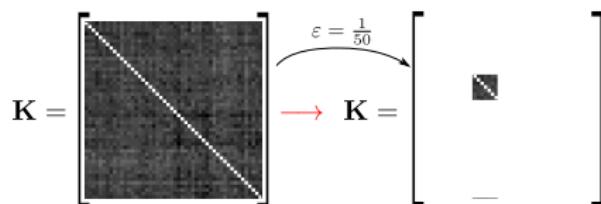
$$\mathbf{K} = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \xrightarrow{\varepsilon = \frac{1}{50}} \mathbf{K} = \begin{bmatrix} \text{■} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \text{—} \end{bmatrix}$$

## “One-way” kernel puncturing: Improving over subsampling

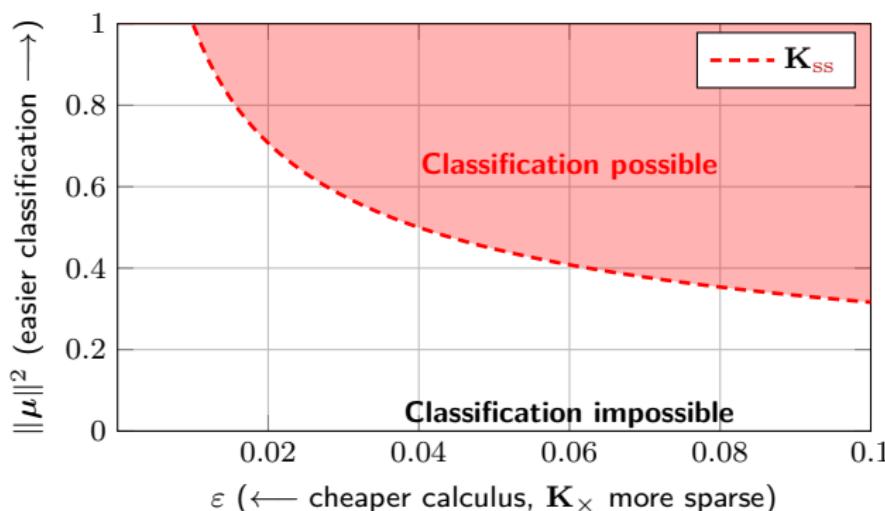
» **Setting:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$

→ **data** subsampling  $\mathbf{K}_{ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$



» **Spectral clustering phase transition:** ( $c_0 = 0.01$ ),



## "One-way" kernel puncturing: Improving over subsampling

» **Setting:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$

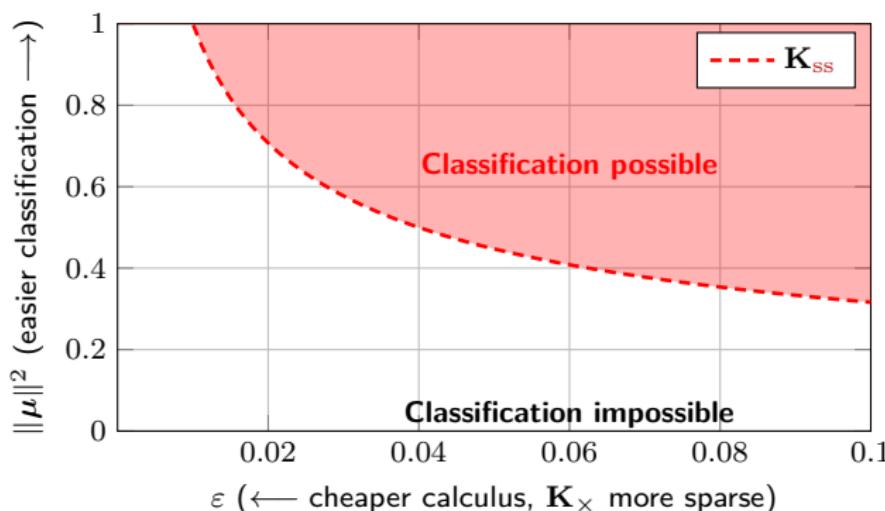
→ **data** subsampling  $\mathbf{K}_{ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$

→ **kernel** puncturing

$\mathbf{K}_\varepsilon \equiv \mathbf{K} \odot \mathbf{B}$  with  $B_{ij} \sim \text{Bern}(\varepsilon)$  i.i.d.

$$\mathbf{K} = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \quad \xrightarrow{\varepsilon = \frac{1}{50}} \quad \mathbf{K}_\varepsilon = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

» **Spectral clustering phase transition:** ( $c_0 = 0.01$ ),



## "One-way" kernel puncturing: Improving over subsampling

» **Setting:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$

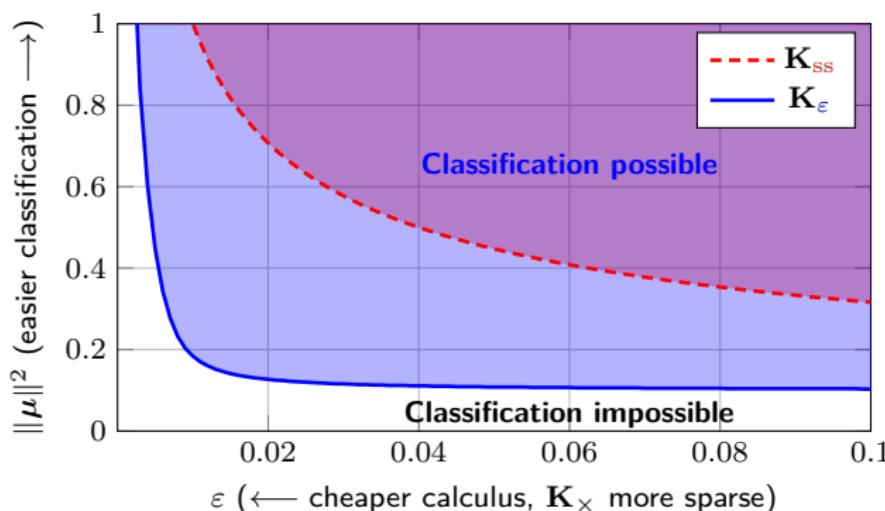
→ **data** subsampling  $\mathbf{K}_{ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$

→ **kernel** puncturing

$\mathbf{K}_\varepsilon \equiv \mathbf{K} \odot \mathbf{B}$  with  $B_{ij} \sim \text{Bern}(\varepsilon)$  i.i.d.

$$\mathbf{K} = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \quad \xrightarrow{\varepsilon = \frac{1}{50}} \quad \mathbf{K}_\varepsilon = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

» **Spectral clustering phase transition:** ( $c_0 = 0.01$ ),



## "One-way" kernel puncturing: Improving over subsampling

» **Setting:**  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ ,  $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\mu, \mathbf{I}_p)$ ,  $p, n \rightarrow \infty$ ,  $p/n \rightarrow c_0$

→ kernel matrix  $\mathbf{K} = \frac{1}{p} \mathbf{X}^T \mathbf{X} \in \mathbb{R}^{n \times n}$

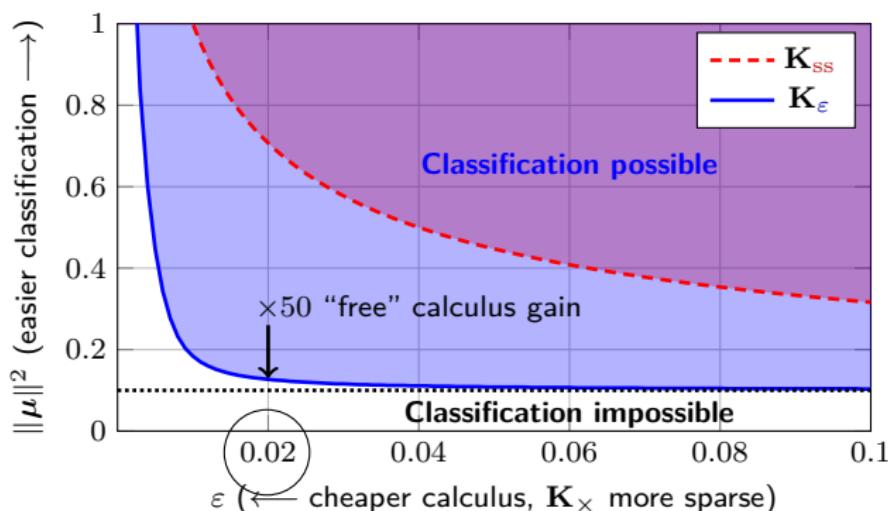
→ **data** subsampling  $\mathbf{K}_{ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$

→ **kernel** puncturing

$\mathbf{K}_\varepsilon \equiv \mathbf{K} \odot \mathbf{B}$  with  $B_{ij} \sim \text{Bern}(\varepsilon)$  i.i.d.

$$\mathbf{K} = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \xrightarrow{\varepsilon = \frac{1}{50}} \mathbf{K}_\varepsilon = \begin{bmatrix} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

» **Spectral clustering phase transition:** ( $c_0 = 0.01$ ),



## “Two-way” kernel puncturing: Trading off storage and complexity

► Puncturing in sample vs. in kernel:

(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$

$$\mathbf{K}_{\varepsilon_S} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{X} & \mathbf{X}^\top & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \ddots & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \rightarrow \mathbf{K}_{\varepsilon_S, \varepsilon_B} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{X} & \mathbf{X}^\top & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$

$$\mathbf{K}_{\varepsilon_S} \begin{bmatrix} \vdots & \vdots \\ \mathbf{X} & \vdots \\ \vdots & \vdots \end{bmatrix} \xrightarrow{\mathbf{X}^\top \mathbf{X} = \mathbf{I}} = \begin{bmatrix} \ddots & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \xrightarrow{\mathbf{B}} \mathbf{K}_{\varepsilon_S, \varepsilon_B} \begin{bmatrix} \vdots & \vdots \\ \mathbf{B} & \vdots \\ \vdots & \vdots \end{bmatrix}$$

- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$

$$\mathbf{K}_{\varepsilon_S} = \begin{bmatrix} \ddots & & \\ & \ddots & \\ & & \ddots \end{bmatrix} \xrightarrow{\mathbf{X}^T} = \begin{bmatrix} \ddots & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \xrightarrow{\longrightarrow} \mathbf{K}_{\varepsilon_S, \varepsilon_B} = \begin{bmatrix} \ddots & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}),$$

## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$

The diagram illustrates the puncturing process. On the left, a matrix  $\mathbf{K}_{\varepsilon_S}$  is shown as a sparse matrix  $\mathbf{X}$  with a red border. A red bracket indicates the columns  $\mathbf{X}^\top$ . An arrow points to the right, where a matrix  $\mathbf{K}_{\varepsilon_S, \varepsilon_B}$  is shown as a sparse matrix with a red border, representing the punctured kernel.

- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

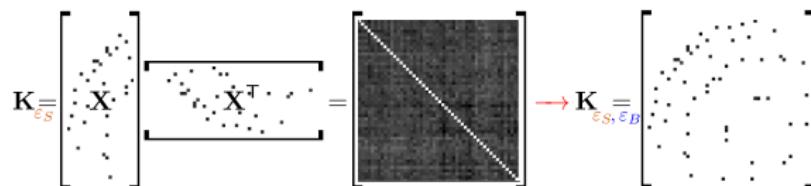
$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}), \quad \zeta = 1 - \frac{c_0}{\|\mu\|^4 \varepsilon_S^2 \varepsilon_B}.$$

## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

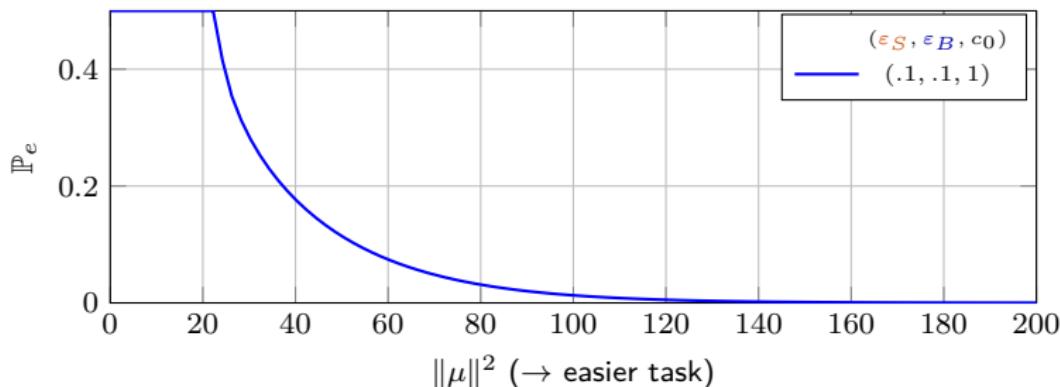
(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$



- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}), \quad \zeta = 1 - \frac{c_0}{\|\mu\|^4 \varepsilon_S^2 \varepsilon_B}.$$

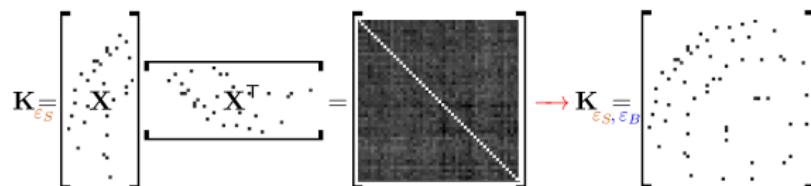


## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

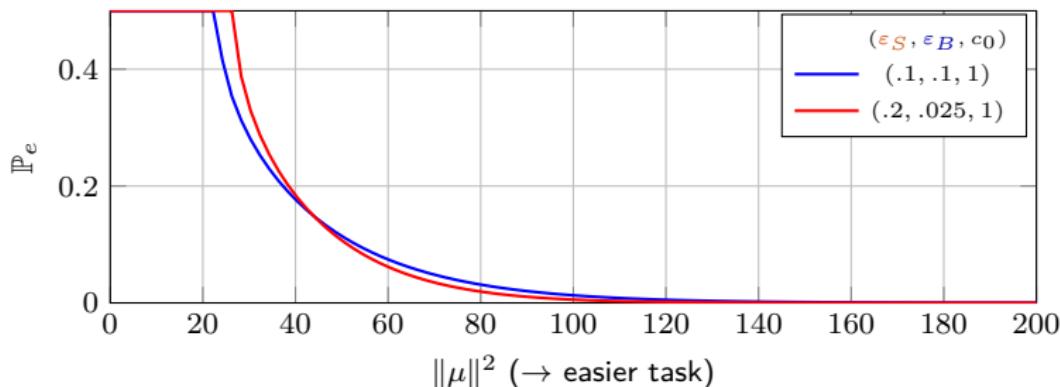
(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$



- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}), \quad \zeta = 1 - \frac{c_0}{\|\mu\|^4 \varepsilon_S^2 \varepsilon_B}.$$

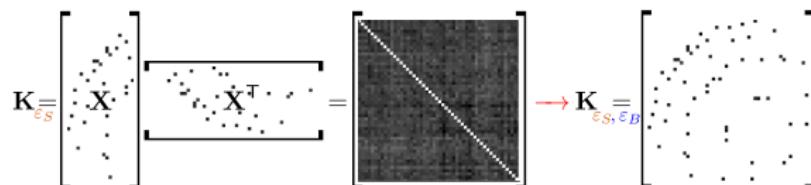


## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

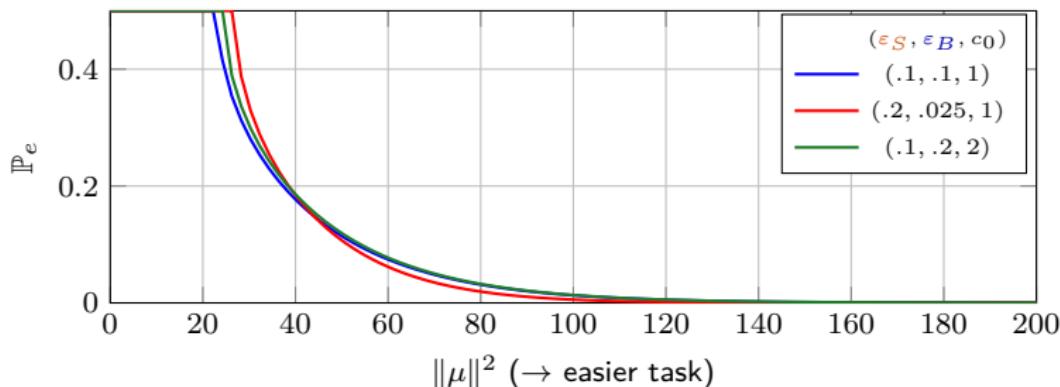
(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\mu, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$



- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}), \quad \zeta = 1 - \frac{c_0}{\|\mu\|^4 \varepsilon_S^2 \varepsilon_B}.$$

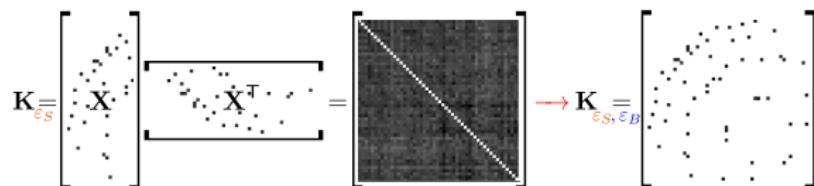


## “Two-way” kernel puncturing: Trading off storage and complexity

- Puncturing in sample vs. in kernel:

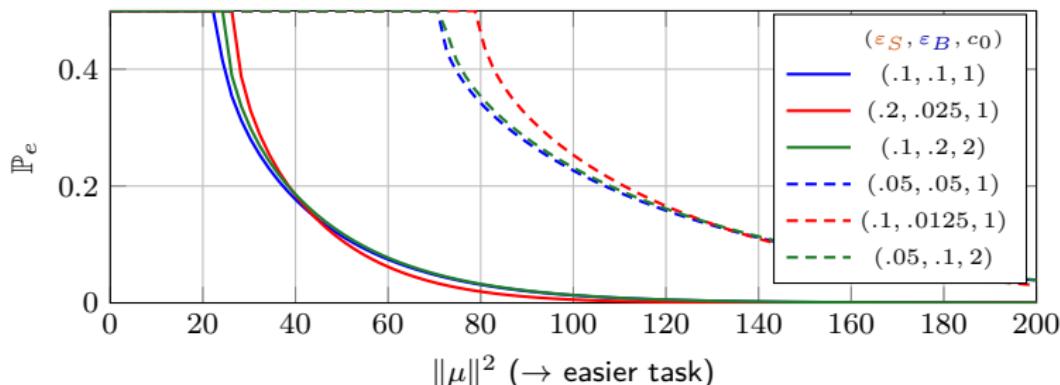
(reminder  $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$ )

$$\mathbf{K}_{\varepsilon_S, \varepsilon_B} = \frac{1}{p} \left\{ (\mathbf{X} \odot \mathbf{S})^\top (\mathbf{X} \odot \mathbf{S}) \right\} \odot \mathbf{B} \quad (\mathbf{S}_{ij} \sim \text{Bern}(\varepsilon_S), \mathbf{B}_{ij} \sim \text{Bern}(\varepsilon_B))$$



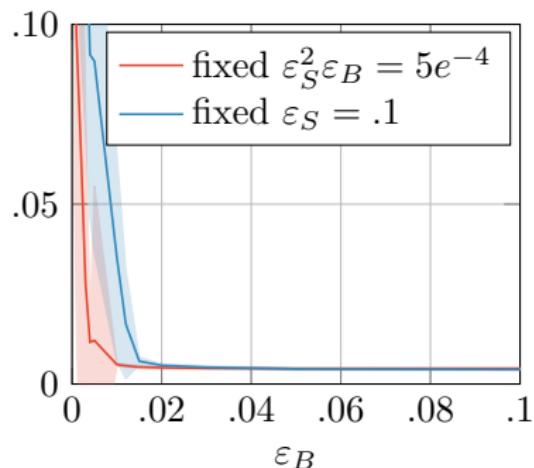
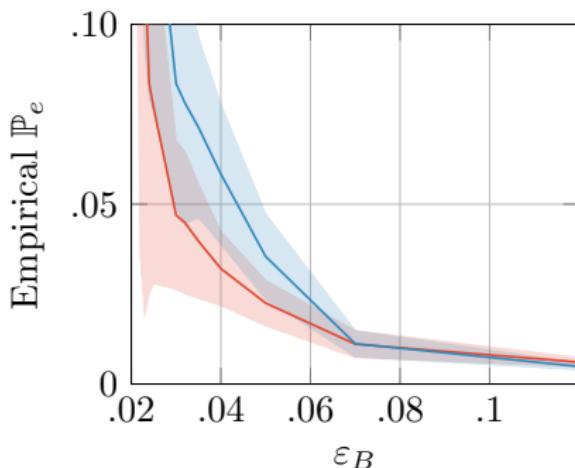
- Asymptotic clustering error  $\mathbb{P}_e$ : as  $n, p \rightarrow \infty$ ,  $p/n \rightarrow c_0$ ,  $\varepsilon_S, \varepsilon_B \ll 1$ ,

$$\mathbb{P}_e = \mathcal{Q} \left( \sqrt{\zeta / (1 - \zeta)} \right) + O(n^{-1}), \quad \zeta = 1 - \frac{c_0}{\|\boldsymbol{\mu}\|^4 \varepsilon_S^2 \varepsilon_B}.$$



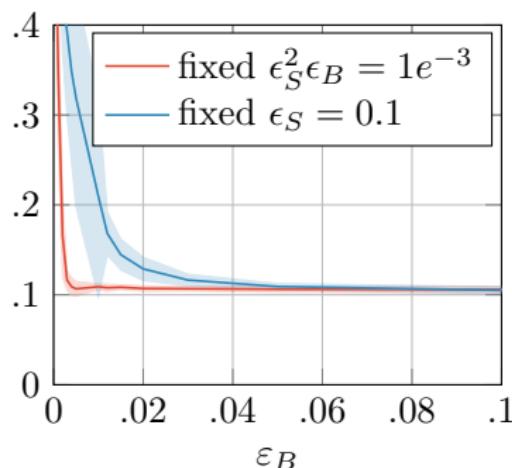
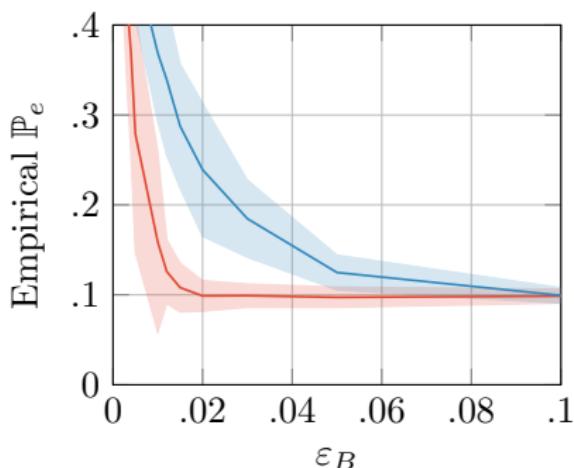
## “Two-way” kernel puncturing: From theory to practice!

► BigGAN images ('tabby' vs 'collie'): VGG features,  $p = 4\,096$ ,  $n = 2\,500$  (**left**),  $n = 10\,000$  (**right**)



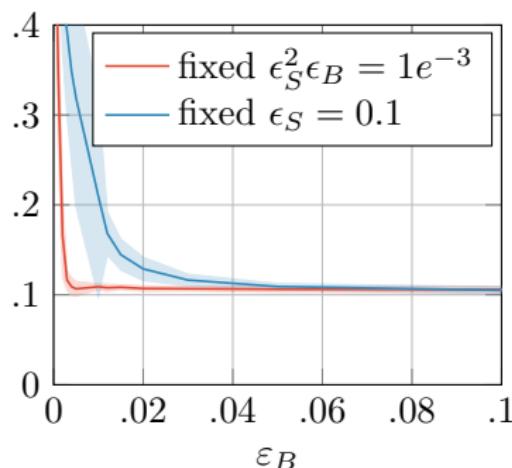
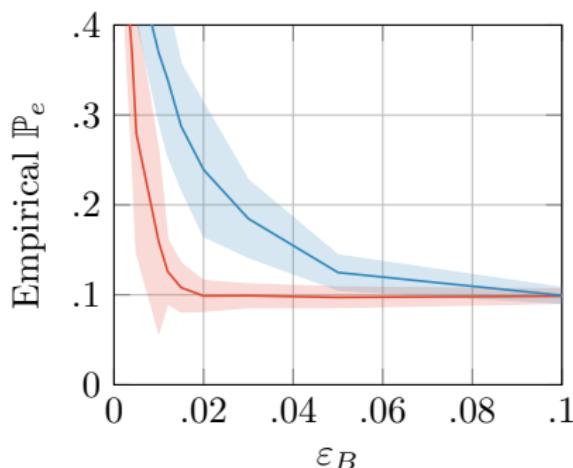
## "Two-way" kernel puncturing: From theory to practice!

► MNIST-fashion ('trouser' vs 'pullover'):  $p = 784$ ,  $n = 512$  (**left**),  $n = 2048$  (**right**)



## "Two-way" kernel puncturing: From theory to practice!

► MNIST-fashion ('trouser' vs 'pullover'):  $p = 784$ ,  $n = 512$  (**left**),  $n = 2048$  (**right**)



Thank You.