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

Romain COUILLET, Florent CHATELAIN, Nicolas LE BIHAN

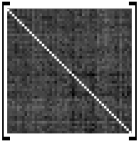
MIAI LargeDATA Chair, GIPSA-lab, University Grenoble–Alpes, France.

June 11, 2021

"One-way" kernel puncturing: Improving over subsampling

- **Setting**: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, \mathbf{I}_p)$, $p, n \to \infty$, $p/n \to c_0$
  - $\to$ kernel matrix $\mathbf{K} = \frac{1}{p}\mathbf{X}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n}$

$\mathbf{K} =$

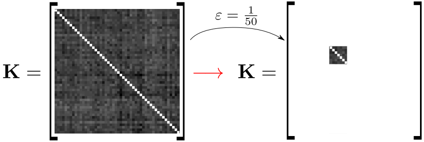


"One-way" kernel puncturing: Improving over subsampling

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

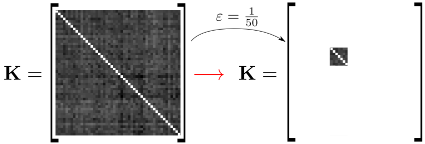
→ kernel matrix $\mathbf{K} = \frac{1}{p}\mathbf{X}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n}$

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

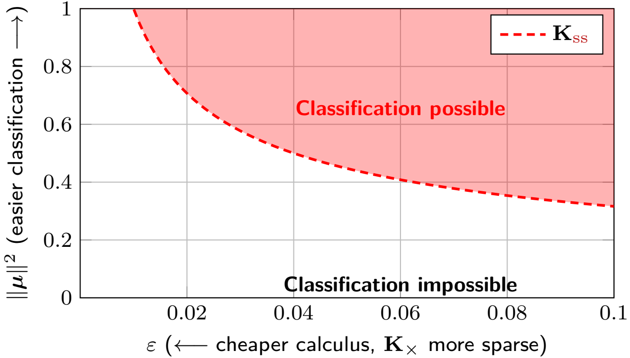


"One-way" kernel puncturing: Improving over subsampling

- **Setting**: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, \mathbf{I}_p)$, $p, n \to \infty$, $p/n \to c_0$
  - $\to$ kernel matrix $\mathbf{K} = \frac{1}{p}\mathbf{X}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n}$
  - $\to$ **data** subsampling $\mathbf{K}_{\rm ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$



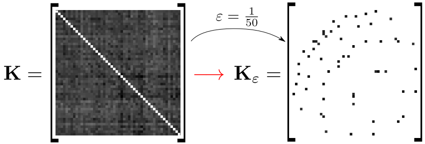
- **Spectral clustering phase transition**: $(c_0 = 0.01)$,



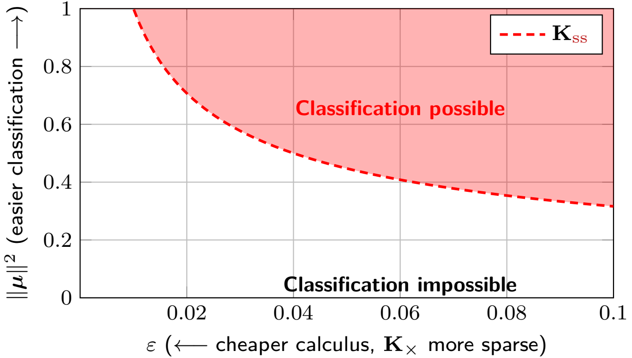
"One-way" kernel puncturing: Improving over subsampling

- **Setting**: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, \mathbf{I}_p)$, $p, n \to \infty$, $p/n \to c_0$
  - → kernel matrix $\mathbf{K} = \frac{1}{p}\mathbf{X}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n}$
  - → **data** subsampling $\mathbf{K}_{\mathrm{ss}} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$
  - → **kernel** puncturing

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



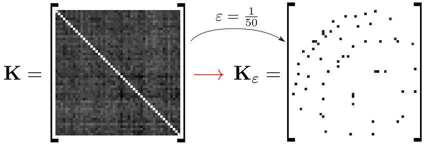
- **Spectral clustering phase transition**: ($c_0 = 0.01$),



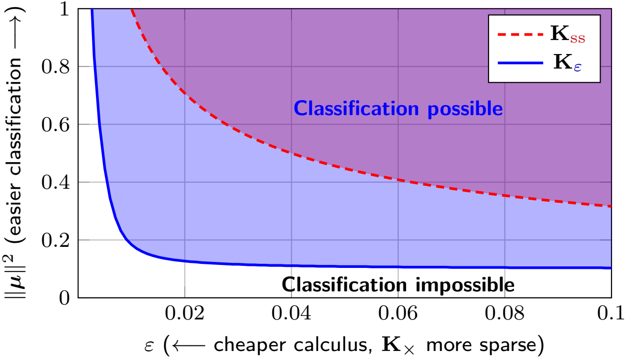
"One-way" kernel puncturing: Improving over subsampling

- **Setting**: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, \mathbf{I}_p)$, $p, n \to \infty$, $p/n \to c_0$
  - → kernel matrix $\mathbf{K} = \frac{1}{p}\mathbf{X}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n}$
  - → **data** subsampling $\mathbf{K}_{\rm ss} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$
  - → **kernel** puncturing

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



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



“One-way” kernel puncturing: Improving over subsampling

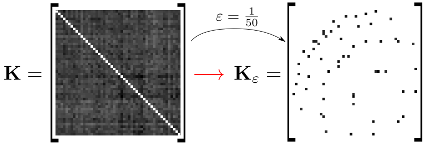
➽ **Setting**: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbb{R}^p \ni \mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, \mathbf{I}_p)$, $p, n \to \infty$, $p/n \to c_0$

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

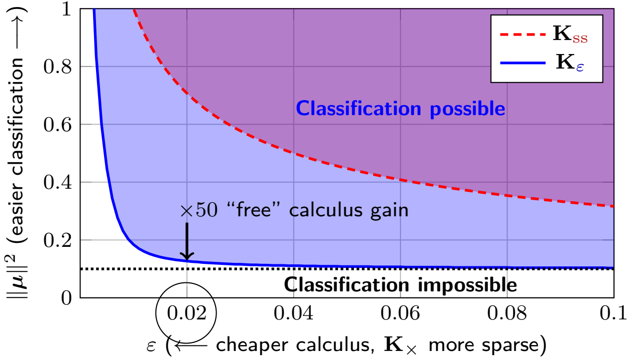
→ **data** subsampling $\mathbf{K}_{\mathrm{ss}} \in \mathbb{R}^{n\varepsilon \times n\varepsilon}$

→ **kernel** puncturing

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

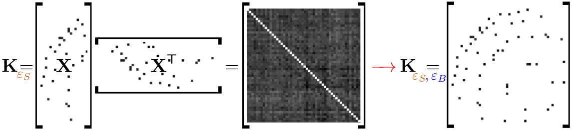


➽ **Spectral clustering phase transition**: $(c_0 = 0.01)$,



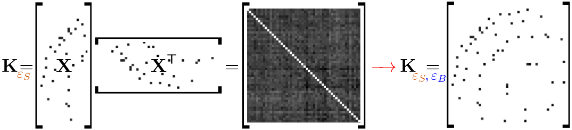
➽ **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim\mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$



➣ **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim \mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$

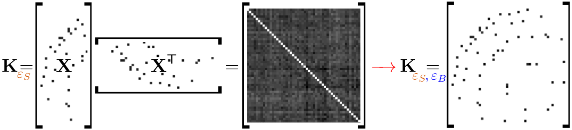


➣ **Asymptotic clustering error** $\mathbb{P}_e$: as $n,p\to\infty$, $p/n\to c_0$, $\varepsilon_S,\varepsilon_B \ll 1$,

"Two-way" kernel puncturing: Trading off storage and complexity

➽ **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim \mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$

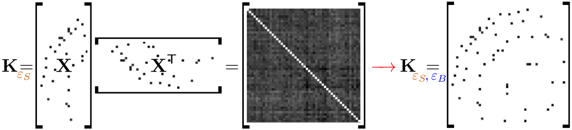


➽ **Asymptotic clustering error** $\mathbb{P}_e$: as $n, p \to \infty$, $p/n \to c_0$, $\varepsilon_S, \varepsilon_B \ll 1$,

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

- **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim\mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$



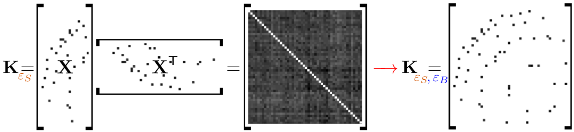
- **Asymptotic clustering error** $\mathbb{P}_e$: as $n,p\to\infty$, $p/n\to c_0$, $\varepsilon_S,\varepsilon_B \ll 1$,

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

"Two-way" kernel puncturing: Trading off storage and complexity

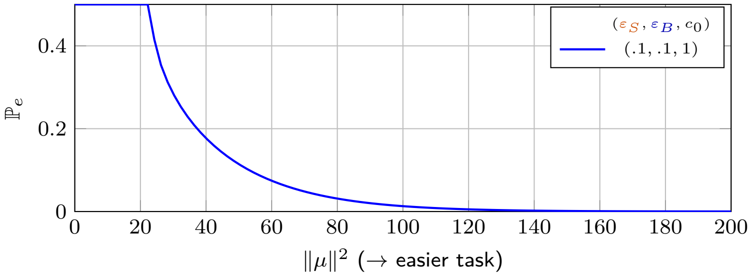
➽ **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim\mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$



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

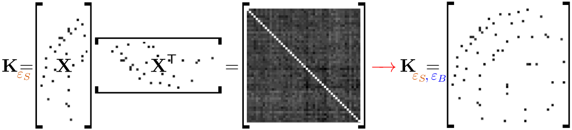
$$\mathbb{P}_e = \mathcal{Q}\left(\sqrt{\zeta/(1-\zeta)}\right) + O(n^{-1}), \qquad \zeta = 1 - \frac{c_0}{\|\boldsymbol{\mu}\|^4\varepsilon_S^2\varepsilon_B}.$$



"Two-way" kernel puncturing: Trading off storage and complexity

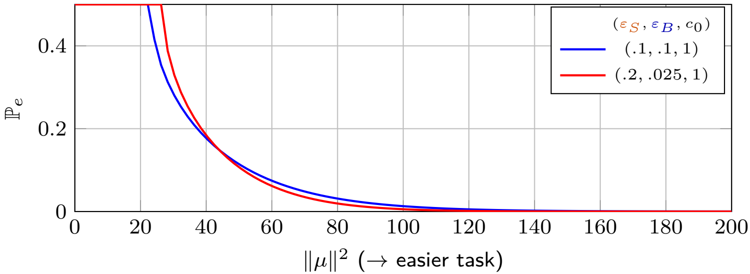
➸ **Puncturing in sample vs. in kernel:** (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim \mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$



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

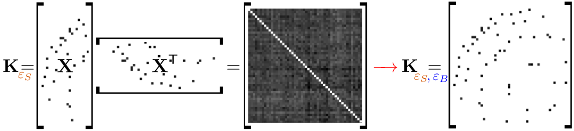
$$\mathbb{P}_e = \mathcal{Q}\left(\sqrt{\zeta/(1-\zeta)}\right) + O(n^{-1}), \qquad \zeta = 1 - \frac{c_0}{\|\boldsymbol{\mu}\|^4\varepsilon_S^2\varepsilon_B}.$$



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

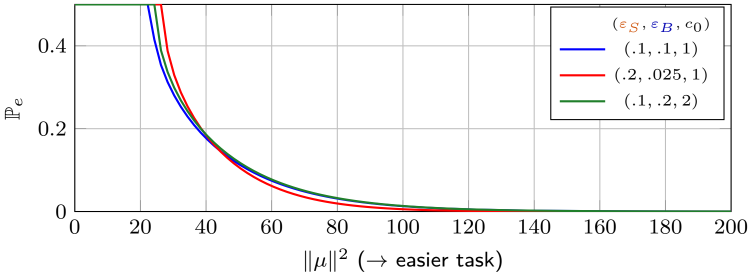
➻ **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim \mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$



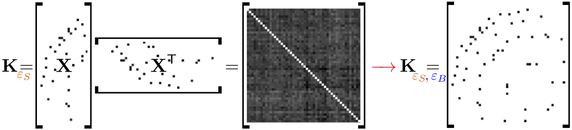
➻ **Asymptotic clustering error** $\mathbb{P}_e$: as $n,p\to\infty$, $p/n\to c_0$, $\varepsilon_S,\varepsilon_B \ll 1$,

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



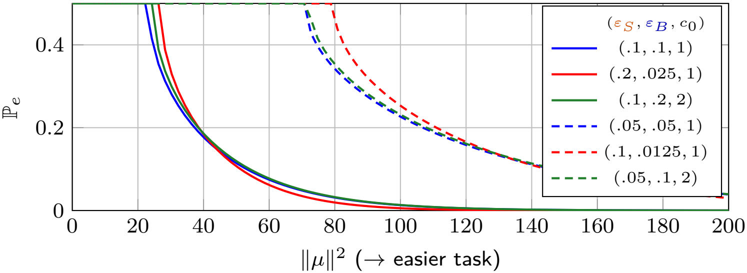
- **Puncturing in sample vs. in kernel**: (reminder $\mathbf{x}_i \sim \mathcal{N}(\pm\boldsymbol{\mu}, I_p)$)

$$\mathbf{K}_{\varepsilon_S,\varepsilon_B} = \frac{1}{p}\left\{(\mathbf{X}\odot\mathbf{S})^\mathsf{T}(\mathbf{X}\odot\mathbf{S})\right\}\odot\mathbf{B} \qquad (\mathbf{S}_{ij}\sim \mathrm{Bern}(\varepsilon_S),\ \mathbf{B}_{ij}\sim\mathrm{Bern}(\varepsilon_B))$$

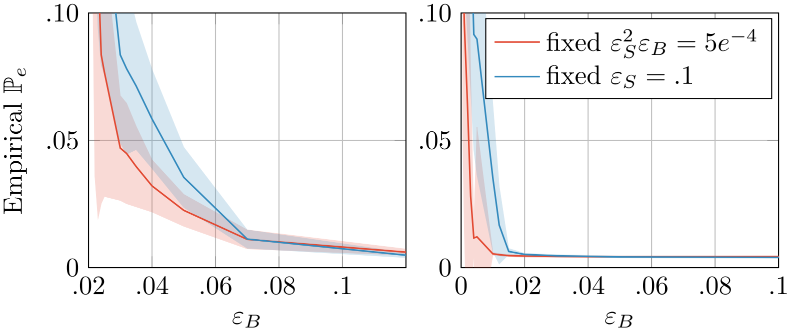


- **Asymptotic clustering error** $\mathbb{P}_e$: as $n,p\to\infty$, $p/n\to c_0$, $\varepsilon_S,\varepsilon_B \ll 1$,

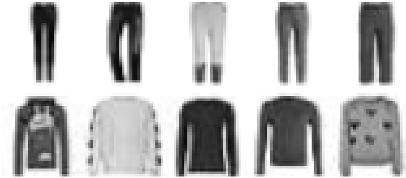
$$\mathbb{P}_e = \mathcal{Q}\left(\sqrt{\zeta/(1-\zeta)}\right) + O(n^{-1}), \qquad \zeta = 1 - \frac{c_0}{\|\boldsymbol{\mu}\|^4\varepsilon_S^2\varepsilon_B}.$$

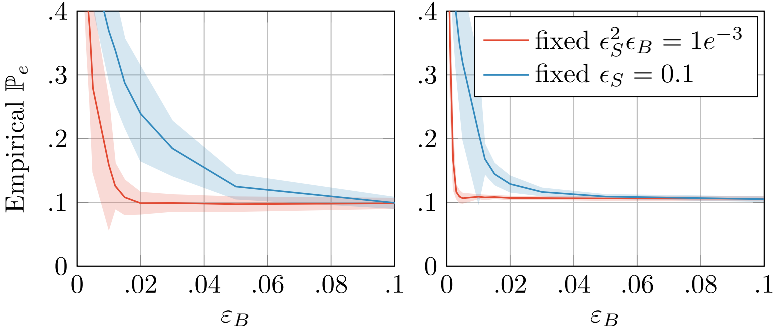


➽ **BigGAN images ('tabby' vs 'collie'):** VGG features, $p = 4\,096$, $n = 2\,500$ (**left**), $n = 10\,000$ (**right**)



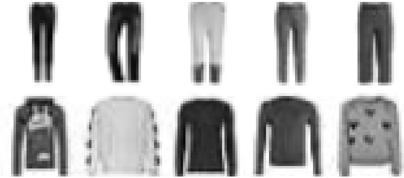
➽ **MNIST-fashion ('trouser' vs 'pullover')**: $p = 784$, $n = 512$ (**left**), $n = 2048$ (**right**)

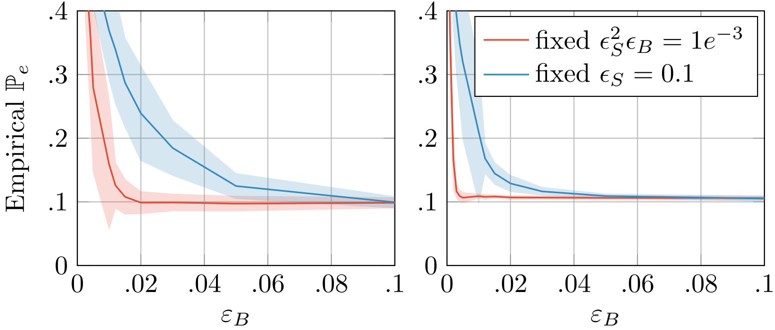




“Two-way” kernel puncturing: From theory to practice!

➽ **MNIST-fashion (‘`trouser`’ vs ‘`pullover`’)**: $p = 784$, $n = 512$ (**left**), $n = 2048$ (**right**)





Thank You.