Learning with a linear loss function

Applications to sign clustering and robustness in sparse PCA.

CIRM, December 20th, 2023

Guillaume Lecué

joint works with S. Chrétien, M. Cucuringu and L. Neirac

ESSEC, business school.

Oracle:

$$Z^* \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle \mathbb{E}A, Z \rangle$$

$A \in \mathbb{R}^{d \times d}$ is a random matrix and $\mathcal{C} \subset \mathbb{R}^{d \times d}$.

Estimator:

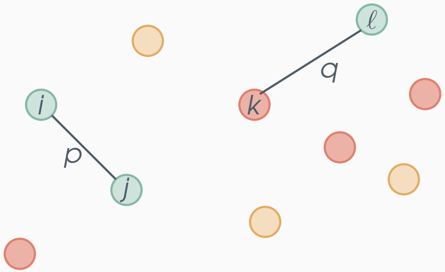
$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle A, Z \rangle$$

We look at $\hat{Z}$ as an ERM = empirical risk minimizer for a **linear loss function** $\ell_Z(A) = -\langle A, Z \rangle$.

Semi-definite program (SDP): optimization problem of the form

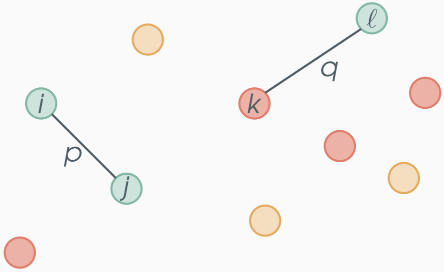
$$\max_{\substack{Z \in \mathbb{R}^{d \times d} \\ Z \succeq 0, \\ \langle B_j, Z \rangle = b_j \text{ for } j=1,\dots,m}} \langle A, Z \rangle.$$

where $A, B_1, \ldots B_m \in \mathbb{R}^{d \times d}$ and $b_1, \ldots, b_m \in \mathbb{R}$ are given.



$0 < q < p < 1$

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- We observe $A \in \mathbb{R}^{d \times d}$, adjacency matrix:
$$A_{ij} \sim \begin{cases} \text{Bern(p) if } i \sim j \\ \text{Bern(q) if } i \nsim j. \end{cases}$$
- We do not observe $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- **Aim:** estimate the membership matrix $\bar{Z} = (1_{\{i \sim j\}})_{1 \le i,j \le d}$



$$0 < q < p < 1$$

- $\lambda := \sum_{k=1}^K |\mathcal{C}_k|^2$
- $\mathcal{C} := \left\{Z \succeq 0 : Z \geq 0, \mathrm{diag}(Z) \preceq I_d, \sum_{i,j=1}^d Z_{ij} \leq \lambda\right\}$,
- Oracle:

$$\bar{Z} = Z^* \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle \mathbb{E}A, Z\rangle.$$

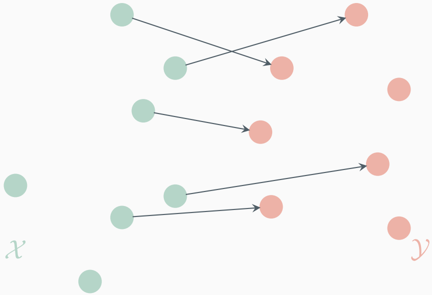
But we only observe A

- Natural estimator:

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle A, Z\rangle$$

(An ERM w.r.t. a linear loss function).

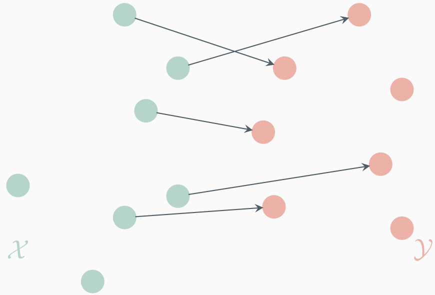
- Data: two clouds of points, $\mathcal{X} = (x_1, \dots, x_d)$ and $\mathcal{Y} = (y_1, \dots, y_d)$.



- Data: two clouds of points, $\mathcal{X} = (x_1, \dots, x_d)$ and $\mathcal{Y} = (y_1, \dots, y_d)$.
- Aim: match each point of $\mathcal{X}$ with a point of $\mathcal{Y}$ reaching the Wasserstein distance

$$\tau^* \in \min_{\tau \in \mathfrak{S}_d} \sum_{i=1}^{d} \|x_i - y_{\tau(i)}\|^2 := W_2^2(\mathcal{X}, \mathcal{Y})$$

(quadratic assignment problem)



- Data: two clouds of points, $\mathcal{X} = (x_1, \dots, x_d)$ and $\mathcal{Y} = (y_1, \dots, y_d)$.
- Aim: match each point of $\mathcal{X}$ with a point of $\mathcal{Y}$ reaching the Wasserstein distance

$$\tau^* \in \min_{\tau \in \mathfrak{S}_d} \sum_{i=1}^{d} \|x_i - y_{\tau(i)}\|^2 := W_2^2(\mathcal{X}, \mathcal{Y})$$

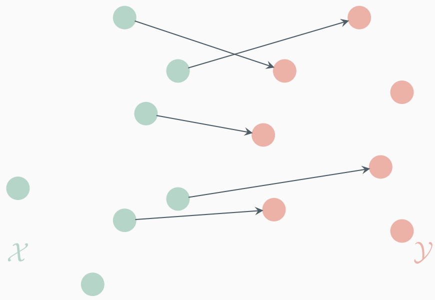
(quadratic assignment problem)

- Equivalent matrix formulation:

Define $Z^* := \left(1_{\{j=\tau^*(i)\}}\right)_{1 \le i,j \le d}$

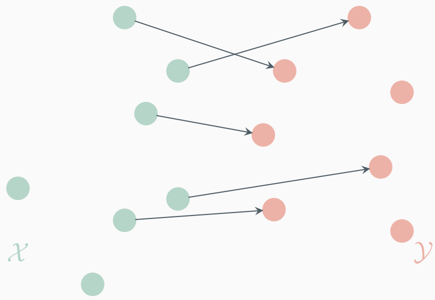
$$\to \quad Z^* \in \operatorname*{argmin}_{Z \in \mathcal{C}} \sum_{i,j} \|x_i - y_j\|_2^2 Z_{ij},$$

with $\mathcal{C} := \{d \times d \text{ bistochastic matrices}\}$.



- Oracle:

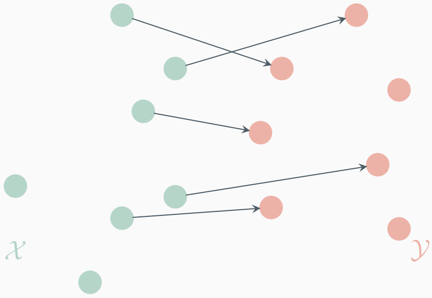
$$Z^* \in \operatorname*{argmin}_{Z \in \mathcal{C}} \sum_{i,j} \|x_i - y_j\|_2^2 Z_{ij},$$



- Oracle:

$$Z^* \in \operatorname*{argmin}_{Z \in \mathcal{C}} \sum_{i,j} \|x_i - y_j\|_2^2 Z_{ij},$$

- We observe: $X_i = x_i + G_i$ and $Y_i = y_i + G_i'$, G_i, G_i': i.i.d. centered noises.



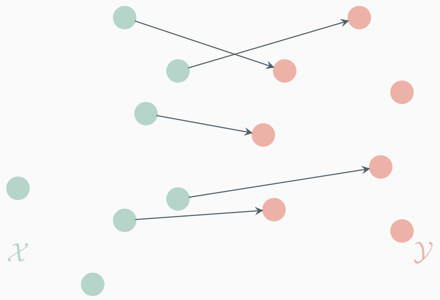
- Oracle:

$$Z^* \in \operatorname*{argmin}_{Z\in\mathcal{C}} \sum_{i,j} \|x_i - y_j\|_2^2 Z_{ij},$$

- We observe: $X_i = x_i + G_i$ and $Y_i = y_i + G_i'$, G_i, G_i': i.i.d. centered noises.

- Quadratic assigment problem:

$$\hat{Z} \in \operatorname*{argmin}_{Z\in\mathcal{C}} \langle A, Z\rangle \text{ where } A = \left(\|X_i - Y_j\|_2^2\right)_{1\leq i,j\leq N}$$



- Oracle:

$$Z^* \in \operatorname*{argmin}_{Z\in\mathcal{C}} \sum_{i,j} \|x_i - y_j\|_2^2 Z_{ij},$$

- We observe: $X_i = x_i + G_i$ and $Y_i = y_i + G'_i$, G_i, G'_i: i.i.d. centered noises.

- Quadratic assigment problem:

$$\hat{Z} \in \operatorname*{argmin}_{Z\in\mathcal{C}} \langle A, Z\rangle \text{ where } A = \left(\|X_i - Y_j\|_2^2\right)_{1\le i,j\le N}$$

We show that in the noiseless case:

$$Z^* \in \operatorname*{argmin}_{Z\in\mathcal{C}} \langle \mathbb{E}A, Z\rangle$$

We look at all the previous estimators as ERMs for a linear loss function.

Other problems

Angular synchronization	$A = \left(e^{\iota \delta_{ij}}\right)_{1\leq i,j,\leq d}$	Bandeira, Boumal, Singer,'16
Variable clustering	$A = (1/N)\sum_{i=1}^N X_i X_i^\top$	Bunea, Giraud, Luo, Royer, Verzelen'18
Max-Cut	$A = \left(s_{ij}A^*_{ij}\right)_{1\leq i,j,\leq d}$	Hong, Lee, Wei,'21
Phase recovery	$A = (XX^\top) \circ (I_N - BB^+)$	Waldspurger, D'Aspremont, Mallat'13
Single index model	$A = (1/N)\sum_{i=1}^N Y_i T(X_i)$	Yang, Balasubramanian, Liu'18
Distance metric learning	$A = \sum_{i,j=1}^N (X_i - X_j)(X_i - X_j)^\top$	Xing, Jordan, Russels'02

2 | General excess risk and estimation bounds with applications in signed clustering and robust sparse PCA.

- H: Hilbert space
- X: random vector in H, distributed according to P
- $Pg := \mathbb{E}[g(X)]$ and $\|g\|_{L_p} = (P|g|^p)^{\frac{1}{p}}$,
- $\mathcal{C}$: subset of H
- $\ell_Z : X \in H \to -\langle X, Z\rangle$, the linear loss function of $Z \in \mathcal{C}$.

Definition | Star-shaped

We say that a set $\mathcal{C}$ is star-shaped in Z^* when for all $Z \in \mathcal{C}$, the segment $[Z, Z^*]$ is in $\mathcal{C}$.

In practice, $\mathcal{C}$ will always be star-shaped in Z^*, since convex sets are star-shaped in any of their elements.

- We are interested in the oracle

$$Z^* \in \operatorname*{argmin}_{Z \in \mathcal{C}} P\ell_Z.$$

- We observe $\mathcal{D} := \{X_1, \ldots, X_N\}$, a sample distributed according to P.

Learning with a linear loss function.

From the observations $\mathcal{D}$, we construct several estimators $\hat{Z}$ for the oracle Z^* and provide statistical guarantees on:

- the excess risk $P\mathcal{L}_{\hat{Z}} = P(\ell_{\hat{Z}} - \ell_{Z^*})$
- the error rate $\|\hat{Z} - Z^*\|_{L_2} := \sqrt{\mathbb{E}\left[\langle X, \hat{Z} - Z^* \rangle^2\right]}$

Here, we consider the following natural estimator for the oracle $Z^* \in \operatorname{argmin}_{Z \in \mathcal{C}} P\ell_Z$:

Definition | ERM estimator

$$\hat{Z}^{\mathrm{ERM}} \in \operatorname*{argmin}_{Z \in \mathcal{C}} P_N \ell_Z$$

where $P_N \ell_Z := \frac{1}{N}\sum_{i=1}^N \ell_Z(X_i) = \frac{1}{N}\sum_{i=1}^N -\langle X_i, Z\rangle = \langle -\bar{X}_N, Z\rangle$.

Definition | Local complexity fixed point

Consider $G : H \to \mathbb{R}$ and $0 < \delta < 1$. The fixed point complexity parameter with respect to the local curvature function G at deviation $1 - \delta$ is

$$r_G^*(\delta) = \inf\left(r > 0 : \mathbb{P}\left[\sup_{Z \in \mathcal{C} : G(Z - Z^*) \leq r} (P_N - P)\mathcal{L}_Z \leq \frac{r}{2}\right] \geq 1 - \delta\right)$$

Definition | Local complexity fixed point

Consider $G : H \to \mathbb{R}$ and $0 < \delta < 1$. The fixed point complexity parameter with respect to the local curvature function G at deviation $1 - \delta$ is

$$r_G^*(\delta) = \inf\left(r > 0 : \mathbb{P}\left[\sup_{Z \in \mathcal{C}: G(Z - Z^*) \leq r} (P_N - P)\mathcal{L}_Z \leq \frac{r}{2}\right] \geq 1 - \delta\right)$$

local curvature assumption

For all $Z \in \mathcal{C}$, if $P\mathcal{L}_Z \leq r_G^*(\delta)$ then $P\mathcal{L}_Z \geq G(Z^* - Z)$. For example, $G(Z) = P\mathcal{L}_Z$, $G(Z) = \|Z\|_{L_2}^2$ or $G(Z) = \|Z\|_1$.

Definition | Local complexity fixed point

Consider $G : H \to \mathbb{R}$ and $0 < \delta < 1$. The fixed point complexity parameter with respect to the local curvature function G at deviation $1 - \delta$ is

$$r_G^*(\delta) = \inf\left(r > 0 : \mathbb{P}\left[\sup_{Z \in \mathcal{C} : G(Z - Z^*) \leq r} (P_N - P)\mathcal{L}_Z \leq \frac{r}{2}\right] \geq 1 - \delta\right)$$

local curvature assumption

For all $Z \in \mathcal{C}$, if $P\mathcal{L}_Z \leq r_G^*(\delta)$ then $P\mathcal{L}_Z \geq G(Z^* - Z)$. For example, $G(Z) = P\mathcal{L}_Z$, $G(Z) = \|Z\|_{L_2}^2$ or $G(Z) = \|Z\|_1$.

Theorem | Excess risk for the ERM estimator with G-localization

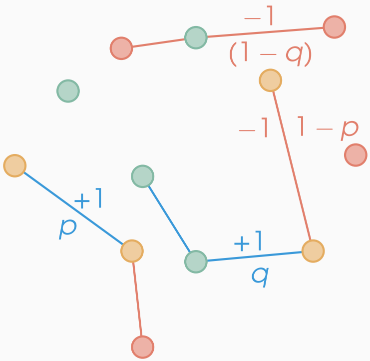
We assume that $\mathcal{C}$ is star-shaped in Z^* and that the local curvature Assumption holds for some $0 < \delta < 1$. Then, with probability at least $1 - \delta$, $G(Z^* - \hat{Z}^{ERM}) \leq P\mathcal{L}_{\hat{Z}^{ERM}} \leq r_G^*(\delta)$.

Application | The ERM estimator for signed clustering

Signed stochastic block model (SSBM)

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \dots, \mathcal{C}_K$

[1] This SSBM is the one introduced by Cucuringu et al.'19 and adapts the one of Fei and Chen'19.



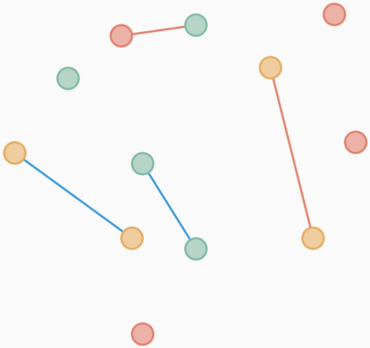
$0 \leq q < 1/2 < p \leq 1$

Signed stochastic block model (SSBM)

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \dots, \mathcal{C}_K$
- We observe $A \in \mathcal{S}_d(\mathbb{R})$ s.t.: $A_{ij} = 1$,

$$2\mathrm{B}_{ij} - 1 \text{ , with } \mathrm{B}_{ij} \sim \begin{cases} \mathrm{Bern}(p) \text{ if } i \sim j \\ \mathrm{Bern}(q) \text{ if } i \not\sim j \end{cases}$$

[1] This SSBM is the one introduced by Cucuringu et al.'19 and adapts the one of Fei and Chen'19.



$$0 \le q < 1/2 < p \le 1$$

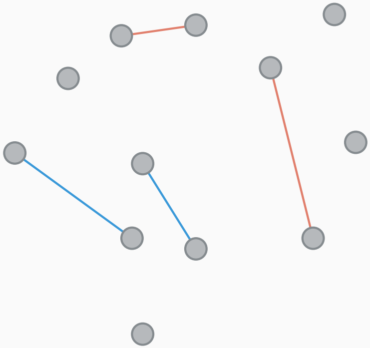
Signed stochastic block model (SSBM)

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \dots, \mathcal{C}_K$
- We observe $A \in \mathcal{S}_d(\mathbb{R})$ s.t.: $A_{ii} = 1$,

$$A_{ij} := s_{ij}(2\mathrm{B}_{ij} - 1), \text{ with } \mathrm{B}_{ij} \sim \begin{cases} \mathrm{Bern}(p) \text{ if } i \sim j \\ \mathrm{Bern}(q) \text{ if } i \not\sim j \end{cases}$$

and $s_{ij} \sim \mathrm{Bern}(\delta)$

[1] This SSBM is the one introduced by Cucuringu et al.'19 and adapts the one of Fei and Chen'19.

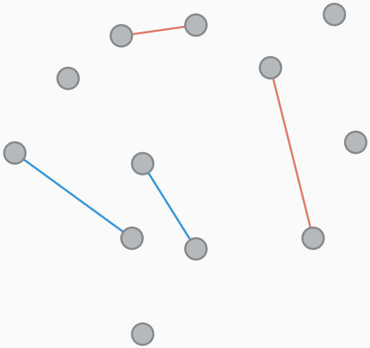


$$0 \leq q < 1/2 < p \leq 1$$

Signed stochastic block model (SSBM)

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- We observe $A \in \mathcal{S}_d(\mathbb{R})$ s.t.: $A_{ii} = 1$,
  $$A_{ij} := s_{ij}(2\mathrm{B}_{ij} - 1), \text{ with } \mathrm{B}_{ij} \sim \begin{cases} \mathrm{Bern}(p) \text{ if } i \sim j \\ \mathrm{Bern}(q) \text{ if } i \not\sim j \end{cases}$$
  and $s_{ij} \sim \mathrm{Bern}(\delta)$
- $\{\mathrm{B}_{ij}, s_{ij},\ 1 \leq i < j \leq d\}$ are independent.

[1] This SSBM is the one introduced by Cucuringu et al.'19 and adapts the one of Fei and Chen'19.



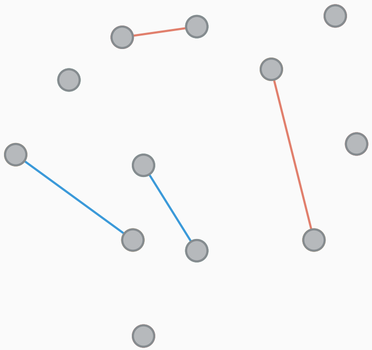
$0 \leq q < 1/2 < p \leq 1$

Signed stochastic block model (SSBM) [1]

- $G = (V, E)$, graph with d nodes
- V can be partitioned into K communities $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- We observe $A \in \mathcal{S}_d(\mathbb{R})$ s.t.: $A_{ii} = 1$,
  $A_{ij} := s_{ij}(2\mathrm{B}_{ij} - 1)$, with $\mathrm{B}_{ij} \sim \begin{cases} \mathrm{Bern}(p) \text{ if } i \sim j \\ \mathrm{Bern}(q) \text{ if } i \not\sim j \end{cases}$
  and $s_{ij} \sim \mathrm{Bern}(\delta)$
- $\{\mathrm{B}_{ij}, s_{ij},\ 1 \leq i < j \leq d\}$ are independent.

Aim: recover $\bar{Z} := \left(1_{\{i \sim j\}}\right)_{1 \leq i,j \leq d}$ the membership matrix.

[1] This SSBM is the one introduced by Cucuringu et al.'19 and adapts the one of Fei and Chen'19.



Property

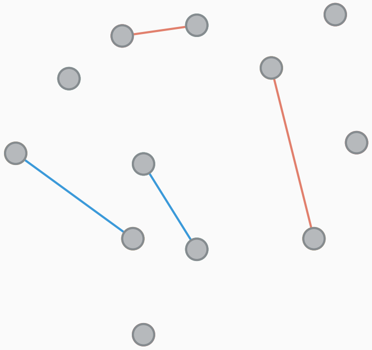
We show that $\bar{Z} = Z^*$, where Z^* is the oracle

$$Z^* \in \operatorname*{argmax}_{Z \in \mathcal{C}} \langle \mathbb{E}A - \alpha J, Z \rangle$$

$\alpha = \delta(p + q - 1)$, $J = (1)_{d \times d}$, $\mathcal{C} = \{Z \succeq 0 : Z_{ij} \in [0, 1], Z_{ii} = 1\}$.

ERM estimator for signed clustering

$$\hat{Z} \in \operatorname*{argmax}_{Z \in \mathcal{C}} \langle A - \alpha J, Z \rangle$$

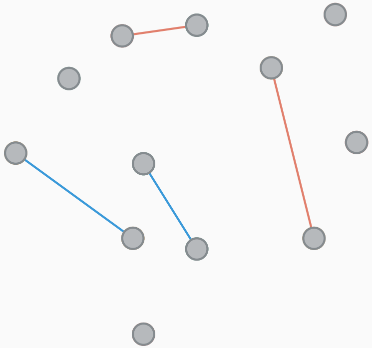


Curvature of the excess risk

For $\theta = \delta(p - q)$, we have for all $Z \in \mathcal{C}$,

$$\langle \mathbb{E}A - \alpha J, Z^* - Z \rangle = \theta \|Z^* - Z\|_1.$$

$\rightarrow$ We compute a complexity fixed point for the localization function $G : Z \rightarrow \|Z\|_1$.



Curvature of the excess risk

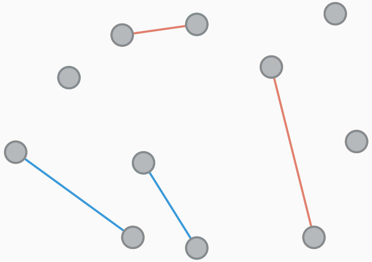
For $\theta = \delta(p - q)$, we have for all $Z \in \mathcal{C}$,

$$\langle \mathbb{E}A - \alpha J, Z^* - Z \rangle = \theta \|Z^* - Z\|_1.$$

$\rightarrow$ We compute a complexity fixed point for the localization function $G : Z \rightarrow \|Z\|_1$.

Complexity fixed point parameter

$$r_G^*(\Delta) = \inf\left(r > 0 : \mathbb{P}\left[\sup_{Z \in \mathcal{C} : \|Z - Z^*\|_1 \leq r} \langle A - \mathbb{E}[A], Z - Z^* \rangle \leq \frac{r}{2} \right] \geq 1 - \Delta \right)$$



Assumptions

(i) $|\mathcal{C}_k| \sim d/k$ for any $k \in [K]$

(ii) $d\nu\delta \geq \log d$

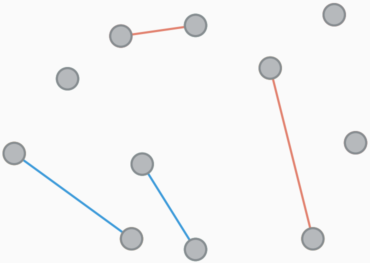
(iii) $sd \gtrsim K^2\nu$

(iv) $\frac{K}{d}\log(2eKd) \leq \max\left(\frac{\theta^2}{\rho}, \frac{9\rho}{32}\right)$

Theorem | Exact recovery in signed clustering

Let $\Delta := \exp(-\delta\nu d) + 3/(2eKd)$. Then, $r_G^*(\Delta) = 0$, that is, $\hat{Z} = Z^*$ with probability at least $1 - \Delta$.

for $\nu = \max(2p-1, 1-2q)$, $\rho = \delta\max(1-\delta(2p-1)^2, 1-\delta(2q-1)^2)$, $s = \delta(p-q)^2$ and $\theta = \delta(p-q)$



Assumptions

(i) $|\mathcal{C}_k| \sim d/k$ for any $k \in [K]$

(ii) $d\nu\delta \geq \log d$

(iii) $sd \gtrsim K^2\nu$

(iv) $\frac{K}{d}\log(2eKd) \leq \max\left(\frac{\theta^2}{\rho}, \frac{9\rho}{32}\right)$

Theorem | Exponential rate in signed clustering

Let $\Delta := \exp(-\delta\nu d) + 3/(2eKd)$. Then, with probability at least $1-\Delta$,

$$\|Z^* - \hat{Z}\|_1 \leq \frac{c_0 d^2}{\theta}\exp\left(-\frac{c_1 sd}{K}\right).$$

Similar to Fei and Chen 17':

- exact recovery in the regime $K^2 + K\log(d) \lesssim d$
- exponential rate with exponent $\simeq -d/K$ otherwise

When Z^* has some structure (e.g. sparsity), we consider the following type of estimator:

Definition | regularized ERM estimator

$$\hat{Z}^{\mathrm{RERM}} \in \underset{Z \in \mathcal{C}}{\mathrm{argmin}} \left(P_N \ell_Z + \lambda \|Z\| \right)$$

$\lambda > 0$ the regularization parameter and $\|\cdot\|$ a norm.

Definition | Complexity fixed point parameter

For $A > 0$, $\rho > 0$ and $\delta \in (0,1)$, the complexity fixed point for the structural learning problem is

$$r^*(\rho) = \inf\left(r > 0 : \mathbb{P}\left(\sup_{Z \in \mathcal{C}: \|Z - Z^*\| \leq \rho, \|Z - Z^*\|_2^2 \leq r} |(P - P_N)\mathcal{L}_Z| \leq \frac{r}{3A}\right) \geq 1 - \delta\right)$$

when the local excess risk curvature is well described by $G : Z \to \|Z\|_2^2$.

Definition | Complexity fixed point parameter

For $A > 0$, $\rho > 0$ and $\delta \in (0, 1)$, the complexity fixed point for the structural learning problem is

$$r^*(\rho) = \inf\left(r > 0 : \mathbb{P}\left(\sup_{Z \in \mathcal{C}: \|Z - Z^*\| \leq \rho, \|Z - Z^*\|_2^2 \leq r} |(P - P_N)\mathcal{L}_Z| \leq \frac{r}{3A}\right) \geq 1 - \delta\right)$$

when the local excess risk curvature is well described by $G : Z \to \|Z\|_2^2$.

Assumption | Local curvature assumption

There exists $A > 0$, $\rho^* > 0$ and $\delta \in (0, 1)$ such that, for any $Z \in \mathcal{C}$, if $\|Z - Z^*\|_2 = r^*(\rho^*)$ and $\|Z - Z^*\| \leq \rho^*$, then $AP\mathcal{L}_Z \geq \|Z - Z^*\|_2^2$.

Definition | Complexity fixed point parameter

For $A > 0$, $\rho > 0$ and $\delta \in (0, 1)$, the complexity fixed point for the structural learning problem is

$$r^*(\rho) = \inf\left(r > 0 : \mathbb{P}\left(\sup_{Z \in \mathcal{C}: \|Z - Z^*\| \leq \rho, \|Z - Z^*\|_2^2 \leq r} |(P - P_N)\mathcal{L}_Z| \leq \frac{r}{3A}\right) \geq 1 - \delta\right)$$

when the local excess risk curvature is well described by $G : Z \to \|Z\|_2^2$.

Assumption | Local curvature assumption

There exists $A > 0$, $\rho^* > 0$ and $\delta \in (0, 1)$ such that, for any $Z \in \mathcal{C}$, if $\|Z - Z^*\|_2 = r^*(\rho^*)$ and $\|Z - Z^*\| \leq \rho^*$, then $AP\mathcal{L}_Z \geq \|Z - Z^*\|_2^2$.

ρ^* is chosen to take advantage of the structure inducing property of the regularization norm:
$\rightarrow$ sparsity equation.

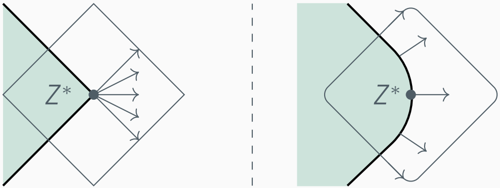
Definition - Subdifferential of $\|.\|$ at a point Z:

$$\partial\|.\|(Z) := \{\Phi \in H : \|Z+h\| - \|Z\| \geq \langle \Phi, h\rangle \text{ for all } h \in H\}.$$

Property - For S^* and B^* the dual sphere and dual ball, we have

$$\partial\|\cdot\|(Z) = \begin{cases} \{\Phi \in S^* : \langle \Phi, Z\rangle = \|Z\|\} & \text{if } Z \neq 0 \\ B^* & \text{if } Z = 0 \end{cases}$$

Subdifferential are large sets at points where $\|.\|$ is not differentiable



$$\partial\|\cdot\|_1 \left(1,0\ldots 0\right)^\top = \left\{\left(1,x_2\ldots x_d\right) : |x_i| \leq 1\right\} \quad \Big| \quad \partial\|\cdot\|_1 \left(1\ldots 1\right)^\top = \left\{\left(1\ldots 1\right)^\top\right\}$$

For $\rho > 0$, we consider

$$\Gamma_{Z^*}(\rho) = \bigcup_{\|Z - Z^*\| \leq \rho/20} (\partial \|.\|)_Z,$$

For $\rho > 0$, we consider

$$\Gamma_{Z^*}(\rho) = \bigcup_{\|Z-Z^*\|\leq\rho/20} (\partial\|.\|)_Z,$$

We want $\Gamma_{Z^*}(\rho)$ to be a large subset of S^* (or B^*) when Z^* is structured or close to a structured element in H.

Definition | Sparsity equation

For $\rho > 0$ we define:

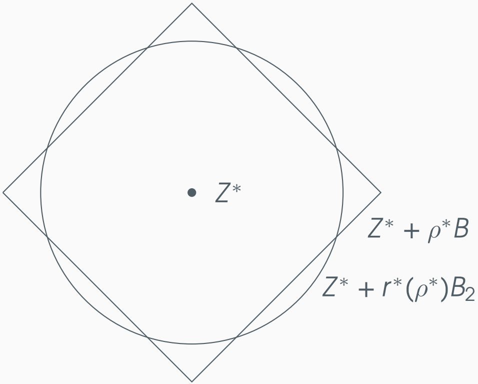
$$H_\rho := \left\{Z \in \mathcal{C} : \|Z - Z^*\| = \rho \;\; and \;\; \|Z - Z^*\|_2^2 \leq r^*(\rho)\right\}$$

and

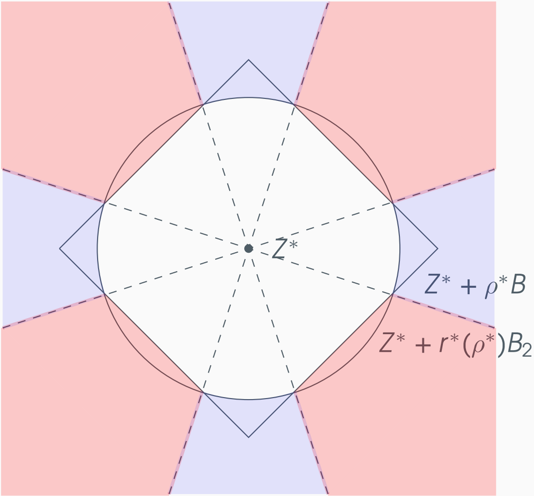
$$\Delta(\rho) := \inf_{Z\in H_\rho} \sup_{\Phi\in\Gamma_{Z^*}(\rho)} \langle \Phi, Z - Z^*\rangle.$$

We say that $\rho > 0$ satisfies the **sparsity equation** when $\Delta(\rho) \geq (4/5)\rho$.

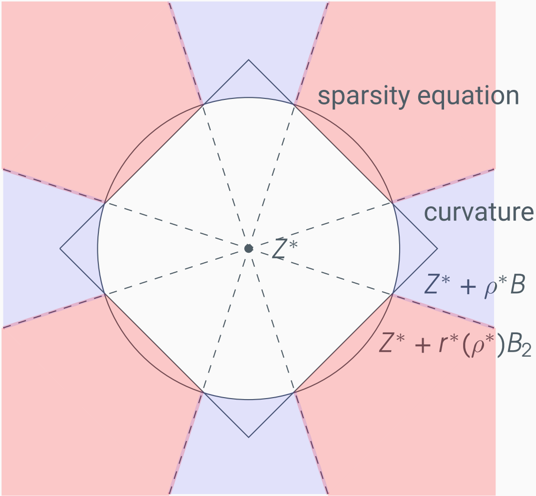
Sparsity equation and curvature



Sparsity equation and curvature



Sparsity equation and curvature



Assumptions

(i) $\mathcal{C}$ is star-shaped in Z^*

(ii) $\rho^* > 0$ s.t. the local curvature assumption holds and ρ^* satisfies the sparsity equation

(iii) $\lambda \simeq r^*(\rho^*)/\rho^*$

Theorem | Excess risk and estimation bounds for the RERM with ℓ_2-localization

With probability at least $1 - \delta$,

$$\|\hat{Z}^{\text{RERM}} - Z^*\| \leq \rho^* \quad , \quad \|\hat{Z}^{\text{RERM}} - Z^*\|_2^2 \leq r^*(\rho^*) \text{ and } P\mathcal{L}_{\hat{Z}^{\text{RERM}}} \leq \frac{r^*(\rho^*)}{A}.$$

- We observe $X_1, \ldots, X_N \in \mathbb{R}^d$, $i.i.d.$ centered and $\mathrm{cov}(X_1 X_1^\top) := \Sigma$,
- Aim[2]: find a 1st k-sparse principal component of Σ:

$$v^* \in \underset{\substack{v \in \mathbb{R}^d \\ \|v\|_2 = 1 \\ \|v\|_0 \leq k}}{\mathrm{argmax}} \langle \Sigma, vv^\top \rangle$$

[2] Johnstone and Lu, 09'; Berthet and Rigollet, 13'; Wang, Berthet and Samworth, 16'

- We observe $X_1, \dots, X_N \in \mathbb{R}^d$, *i.i.d.* centered and $\mathrm{cov}(X_1 X_1^\top) := \Sigma$,
- Aim[2]: find a 1st k-sparse principal component of Σ:

$$v^* \in \underset{\substack{v \in \mathbb{R}^d \\ \|v\|_2 = 1 \\ \|v\|_0 \leq k}}{\mathrm{argmax}} \langle \Sigma, vv^\top \rangle$$

Lifting
$$Z \in \underset{\substack{Z \succeq 0 \\ \mathrm{rank}(Z)=1 \text{ and } \|Z\|_0 \leq k^2 \\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z \rangle$$

[2] Johnstone and Lu, 09'; Berthet and Rigollet, 13'; Wang, Berthet and Samworth, 16'

- We observe $X_1, \dots, X_N \in \mathbb{R}^d$, $i.i.d.$ centered and $\mathrm{cov}(X_1 X_1^\top) := \Sigma$,
- Aim[2]: find a 1st k-sparse principal component of Σ:

$$v^* \in \underset{\substack{v \in \mathbb{R}^d \\ \|v\|_2 = 1 \\ \|v\|_0 \leq k}}{\mathrm{argmax}} \langle \Sigma, vv^\top \rangle$$

Lifting

$$Z \in \underset{\substack{Z \succeq 0 \\ \cancel{\mathrm{rank}(Z)=1 \text{ and }} \|Z\|_0 \leq k^2 \\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z \rangle$$

[2] Johnstone and Lu, 09'; Berthet and Rigollet, 13'; Wang, Berthet and Samworth, 16'

- We observe $X_1, \dots, X_N \in \mathbb{R}^d$, *i.i.d.* centered and $\mathrm{cov}(X_1 X_1^\top) := \Sigma$,
- Aim[2]: find a 1st k-sparse principal component of Σ:

$$v^* \in \underset{\substack{v\in\mathbb{R}^d\\ \|v\|_2=1\\ \|v\|_0\leq k}}{\mathrm{argmax}} \langle \Sigma, vv^\top\rangle$$

Lifting $$Z \in \underset{\substack{Z\succeq 0\\ \cancel{\mathrm{rank}(Z)=1 \text{ and } \|Z\|_0\leq k^2}\\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z\rangle$$

SDP Relaxation $$Z^* \in \underset{\substack{Z\succeq 0\\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z\rangle$$

[2] Johnstone and Lu, 09'; Berthet and Rigollet, 13'; Wang, Berthet and Samworth, 16'

- We observe $X_1, \dots, X_N \in \mathbb{R}^d$, *i.i.d.* centered and $\mathrm{cov}(X_1 X_1^\top) := \Sigma$,
- Aim[2]: find a 1^{st} k-sparse principal component of Σ:

$$v^* \in \underset{\substack{v \in \mathbb{R}^d \\ \|v\|_2 = 1 \\ \|v\|_0 \le k}}{\mathrm{argmax}} \langle \Sigma, vv^\top \rangle$$

Lifting $$Z \in \underset{\substack{Z \succeq 0 \\ \cancel{\mathrm{rank}(Z)=1 \text{ and } \|Z\|_0 \le k^2} \\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z \rangle$$

SDP Relaxation $$Z^* \in \underset{\substack{Z \succeq 0 \\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \langle \Sigma, Z \rangle$$

Estimator for Z^*: $$\hat{Z}_\lambda^{RERM} \in \underset{\substack{Z \succeq 0 \\ \mathrm{Tr}(Z)=1}}{\mathrm{argmax}} \left(\langle \hat{\Sigma}_N, Z \rangle + \lambda \|Z\|_1 \right) \text{ for } \hat{\Sigma}_N = \frac{1}{N} \sum_{i=1}^N X_i X_i^\top$$

Estimator for v^*: $\hat{v}$ a top unit-norm eigenvector of $\hat{Z}$

[2] Johnstone and Lu, 09'; Berthet and Rigollet, 13'; Wang, Berthet and Samworth, 16'

Spiked covariance model: $\Sigma = I_d + \theta\beta^*(\beta^*)^\top$ with $\beta^* \in \mathcal{S}_2^d$, k-sparse, $\theta > 0$

Exactness in the spiked covariance model

In the spiked covariance model, we have $Z^* = (\beta^*)(\beta^*)^\top$.

Spiked covariance model: $\Sigma = I_d + \theta\beta^*(\beta^*)^\top$ with $\beta^* \in \mathcal{S}_2^d$, k-sparse, $\theta > 0$

Exactness in the spiked covariance model

In the spiked covariance model, we have $Z^* = (\beta^*)(\beta^*)^\top$.

Curvature of the excess risk

In the spiked covariance model: for all $Z \in \mathcal{C} = \{Z \succeq 0 : \mathrm{Tr}(Z) = 1\}$,

$$P\mathcal{L}_Z = \langle \Sigma, Z^* - Z \rangle \geq (\theta/2)\|Z^* - Z\|_2^2.$$

$\rightarrow$ The local curvature assumption holds globally for $G : Z \rightarrow \|Z\|_2^2$.

Spiked covariance model: $\Sigma = I_d + \theta\beta^*(\beta^*)^\top$ with $\beta^* \in \mathcal{S}_2^d$, k-sparse, $\theta > 0$

Exactness in the spiked covariance model

In the spiked covariance model, we have $Z^* = (\beta^*)(\beta^*)^\top$.

Curvature of the excess risk

In the spiked covariance model: for all $Z \in \mathcal{C} = \{Z \succeq 0 : \mathrm{Tr}(Z) = 1\}$,

$$P\mathcal{L}_Z = \langle \Sigma, Z^* - Z\rangle \geq (\theta/2)\|Z^* - Z\|_2^2.$$

$\longrightarrow$ The local curvature assumption holds globally for $G : Z \to \|Z\|_2^2$.

Sparsity equation

ρ satisfies the sparsity equation when $\rho \geq 10k\sqrt{r^*(\rho)}$.

Assumptions:

i. For all $p, q \in [d]$ and all $2 \leq r \leq 3\log(ed/k)$, we have
$$\|X_{1p}X_{1q} - \mathbb{E}(X_{1p}X_{1q})\|_{L_r} \leq c_0 r$$

ii. $N \gtrsim \log(ed/k)$

iii. $\lambda \simeq \sqrt{(1/N)\log(d/k)}$

Theorem | Excess risk and estimation bounds for the RERM in sparse PCA

In the Spiked covariance model: $\Sigma = I_d + \theta\beta^*(\beta^*)^\top$, with probability at least $1 - 10k/ed$:

$$\|\hat{Z}_\lambda^{\mathrm{RERM}} - Z^*\|_1 \lesssim k^2\sqrt{\frac{1}{N\theta^2}\log\left(\frac{ed}{k}\right)}, \quad \|\hat{Z}_\lambda^{\mathrm{RERM}} - Z^*\|_2 \lesssim \sqrt{\frac{k^2}{N\theta^2}\log\left(\frac{ed}{k}\right)}$$

and

$$P\mathcal{L}_{\hat{Z}_\lambda^{\mathrm{RERM}}} \lesssim \frac{k^2}{N\theta}\log\left(\frac{ed}{k}\right).$$

Assumptions:

i. For all $p, q \in [d]$ and all $2 \leq r \leq 3\log(ed/k)$, we have

$$\|X_{1p}X_{1q} - \mathbb{E}(X_{1p}X_{1q})\|_{L_r} \leq c_0 r$$

ii. $N \gtrsim \log\left(ed/k\right)$

iii. $\lambda \simeq \sqrt{(1/N)\log(d/k)}$

Corollary | Estimation bounds for β^*

Let $\hat{\beta} \in \mathbb{R}^d$ be a leading unit length eigenvector of $\hat{Z}_\lambda^{\mathrm{RERM}}$. Applying the sin-$\theta$ theorem: with probability at least $1 - 10k/ed$:

$$\|\hat{\beta}\hat{\beta}^\top - \beta^*(\beta^*)^\top\|_2 \lesssim \sqrt{\frac{k^2}{N\theta^2}\log\left(\frac{ed}{k}\right)}.$$

This improves the result from Wang, Berthet and Samworth 16':

- $\log(d/k)$ rate instead of $\log(d)$, thanks to localization.
- we only need $\log(d/k)$ moments.

Adversarial contamination setup

good data $\widetilde{X}_1, \quad \widetilde{X}_2 \quad \dots \quad \widetilde{X}_p \quad \dots \quad \widetilde{X}_N \quad \in H, \quad i.i.d$

adversary $\widetilde{X}_1, \quad X_2 \quad \dots \quad X_p \quad \dots \quad \widetilde{X}_N \quad \in H$

$$\mathcal{I} := \left\{ i \in [N] : \widetilde{X}_i = X_i \right\} \qquad \mathcal{O} := [N] \backslash \mathcal{I}$$

$(X_i)_{i \in \mathcal{I}}$: inliers | $(X_i)_{i \in \mathcal{O}}$: outliers

$\mathcal{O} \cup \mathcal{I} = [N]$ is unknown to the statistician.

In what follows: $P\mathcal{L}_Z := \mathbb{E}[-\langle \widetilde{X}, Z \rangle], \quad Z^* \in \mathrm{argmin}_{Z \in \mathcal{C}} P\mathcal{L}_Z.$

MOM principle

- $[N] = B_1 \cup \ldots \cup B_K$, $|B_k| = N/K$

[3] Lecué and Lerasle '20

MOM principle

- $[N] = B_1 \cup \ldots \cup B_K$, $|B_k| = N/K$
- $P_{B_k}(g) = (K/N)\sum_{i\in B_k} g(X_i)$

[3] Lecué and Lerasle '20

MOM principle

- $[N] = B_1 \cup \ldots \cup B_K$, $|B_k| = N/K$
- $P_{B_k}(g) = (K/N) \sum_{i \in B_k} g(X_i)$
- $MOM_K(g) := Med\left(P_{B_1}(g), \ldots, P_{B_K}(g)\right)$

[3] Lecué and Lerasle '20

MOM principle

- $[N] = B_1 \cup \ldots \cup B_K$, $|B_k| = N/K$
- $P_{B_k}(g) = (K/N)\sum_{i\in B_k} g(X_i)$
- $MOM_K(g) := Med\left(P_{B_1}(g), \ldots, P_{B_K}(g)\right)$

Definition | Regularized minmax MOM estimator[3]

$$\hat{Z}^{\mathrm{RMOM}}_{K,\lambda} \in \underset{Z\in\mathcal{C}}{\mathrm{argmin}} \sup_{Z'\in\mathcal{C}} \left(\mathrm{MOM}_K(\ell_Z - \ell_{Z'}) + \lambda(\|Z\| - \|Z'\|)\right)$$

Idea: Use this procedure for the linear loss function.

[3] Lecué and Lerasle '20

We consider the case where the excess risk $P\mathcal{L}_Z$ is well described by $G : Z \to \|Z\|_2^2$.

Definition | Complexity fixed point parameter

Let $\sigma_1, \dots, \sigma_N$ be independent Rademacher variables independent of the $\tilde{X}_i$'s. The complexity fixed point for the robust structural learning problem is

$$r^*_{\mathrm{RMOM}}(\rho) := \inf \left\{ r > 0 \; : \; \max\left(E(r,\rho), V_K(r,\rho) \right) \leq r^2 \right\}$$

where, for all $r, \rho > 0$,

$$E(r,\rho) := \mathbb{E}\left[\sup_{Z \in \mathcal{C}_{\rho,r}} \left| \frac{1}{N} \sum_{i=1}^N \sigma_i \mathcal{L}_Z(\tilde{X}_i) \right| \right] \text{ and } V_K(r,\rho) := \sqrt{\frac{K}{N}} \sup_{Z \in \mathcal{C}_{\rho,r}} \sqrt{\mathrm{Var}(\mathcal{L}_Z(\tilde{X}))},$$

with $\mathcal{C}_{\rho,r} = \{Z \in \mathcal{C} : \|Z - Z^*\| \leq \rho, \|Z - Z^*\|_2^2 \leq r^2\}$.

Definition | Sparsity equation

Let

$$\bar{H}_\rho := \left\{Z \in \mathcal{C} : \|Z - Z^*\| = \rho \text{ and } \|Z - Z^*\|_2^2 \leq r^*_{\mathrm{RMOM}}(\rho)^2\right\}$$

and

$$\bar{\Delta}(\rho) := \inf_{Z \in \bar{H}_\rho} \sup_{\Phi \in \Gamma_{Z^*}(\rho)} \langle \Phi, Z - Z^* \rangle.$$

We say that ρ satisfies the **sparsity equation** if $\bar{\Delta}(\rho) \geq 4\rho/5$.

Assumption | Local curvature assumption

Let ρ^* satisfies the sparsity equation. We assume that for all $Z \in \mathcal{C}$, if $\|Z - Z^*\|_2^2 = r^*_{\mathrm{RMOM}}(\rho^*)^2$ and $\|Z - Z^*\| \leq \rho^*$, then $AP\mathcal{L}_Z \geq \|Z - Z^*\|_2^2$.

Assumptions

i. Adversarial contamination framework with $|\mathcal{O}|$ outliers
ii. $\mathcal{C}$ is star-shaped in Z^*
iii. $K \gtrsim |\mathcal{O}|$
iv. the local curvature assumption holds and ρ^* satisfies the sparsity equation
v. $\lambda \simeq r^*_{\text{RMOM}}(\rho^*)/\rho^*$

Theorem | Excess risk and estimation bounds for the RMOM with ℓ_2-localization

With probability at least $1 - 2\exp(-72K/625)$,

$$\|\hat{Z}_{K,\lambda}^{\text{RMOM}} - Z^*\| \leq \rho^* \quad , \quad \|\hat{Z}_{K,\lambda}^{\text{RMOM}} - Z^*\|_2^2 \leq r^*_{\text{RMOM}}(\rho^*)^2 \;\; \text{and} \;\; P\mathcal{L}_{\hat{Z}_{K,\lambda}^{\text{RMOM}}} \leq r^*_{\text{RMOM}}(\rho^*)^2.$$

Assumptions:
- Spiked covariance model: $\Sigma = I_d + \theta\beta^*(\beta^*)^\top$ with $\beta^* \in \mathcal{S}_2^d$, k-sparse, $\theta > 0$
- Adversarial contamination with $|\mathcal{O}|$ outliers
- For all $p, q \in [d]$ and all $2 \leq r \leq 2\log(ed/k)$, $\|\tilde{X}_{1p}\tilde{X}_{1q} - \mathbb{E}(\tilde{X}_{1p}\tilde{X}_{1q})\|_{L_r} \leq c_0 r$
- $N \gtrsim \log(d/k)$, $\theta \leq k$ and $K \gtrsim |\mathcal{O}|$

Theorem | Excess risk and estimation bounds for the RMOM in sparse PCA

Let $\lambda \simeq r^*_{\mathrm{RMOM}}(\rho^*)/\rho^*$. With probability at least $1 - \exp(-72K/625)$:

$$\|\hat{Z}_{K,\lambda}^{\mathrm{RMOM}} - Z^*\|_1 \lesssim \frac{k}{\sqrt{N\theta^2}} \max\left(k\sqrt{\log\left(\frac{ed}{k}\right)}; \sqrt{K}\right),$$

$$\|\hat{Z}_{K,\lambda}^{\mathrm{RMOM}} - Z^*\|_2 \lesssim \frac{1}{\sqrt{N\theta^2}} \max\left(k\sqrt{\log\left(\frac{ed}{k}\right)}; \sqrt{K}\right)$$

$$P\mathcal{L}_{\hat{Z}_{K,\lambda}^{\mathrm{RMOM}}} \lesssim \frac{1}{N\theta} \max\left(k^2\log\left(\frac{ed}{k}\right); K\right).$$

Thanks !

Example for the minmax MOM version of the LASSO:

$$\hat{u} \in \operatorname*{argmin}_{u \in \mathbb{R}^d} \sup_{u' \in \mathbb{R}^d} MOM_K(\ell_u - \ell_{u'}) + \lambda_K \left(\|u\|_1 - \|u'\|_1 \right)$$

where $\ell_u(x, y) = (y - \langle x, u \rangle)^2$ and $\lambda_K \sim \sigma \sqrt{(1/N) \log \left(\sigma^2 d / K \right)}$.

Example for the minmax MOM version of the LASSO:

$$\hat{u} \in \underset{u \in \mathbb{R}^d}{\operatorname{argmin}} \sup_{u' \in \mathbb{R}^d} MOM_K(\ell_u - \ell_{u'}) + \lambda_K \left(\|u\|_1 - \|u'\|_1 \right)$$

where $\ell_u(x, y) = (y - \langle x, u \rangle)^2$ and $\lambda_K \sim \sigma \sqrt{(1/N) \log \left(\sigma^2 d/K \right)}$.

At iteration $(u_t, u_{t'})$ we do:

u1 random partition: $\{1, \ldots, N\} = B_1 \cup \cdots \cup B_K$

u2 median block: $P_{B_k}(\ell_{u_t} - \ell_{u'_t}) = MOM_K(\ell_{u_t} - \ell_{u'_t})$

u3 descent direction: $\nabla_t := \nabla(u \to P_{B_k} \ell_u)_{|u=u_t}$

u4 $u_{t+1} = prox_{\lambda_K \|\cdot\|_1}(u_t - \eta_t \nabla_t)$

Example for the minmax MOM version of the LASSO:

$$\hat{u} \in \underset{u \in \mathbb{R}^d}{\text{argmin}} \sup_{u' \in \mathbb{R}^d} MOM_K(\ell_u - \ell_{u'}) + \lambda_K \left(\|u\|_1 - \|u'\|_1 \right)$$

where $\ell_u(x, y) = (y - \langle x, u \rangle)^2$ and $\lambda_K \sim \sigma\sqrt{(1/N) \log\left(\sigma^2 d/K\right)}$.

At iteration $(u_t, u_{t'})$ we do:

u1 random partition: $\{1, \ldots, N\} = B_1 \cup \cdots \cup B_K$
u2 median block: $P_{B_k}(\ell_{u_t} - \ell_{u'_t}) = MOM_K(\ell_{u_t} - \ell_{u'_t})$
u3 descent direction: $\nabla_t := \nabla(u \to P_{B_k}\ell_u)_{|u=u_t}$
u4 $u_{t+1} = prox_{\lambda_K \|\cdot\|_1}(u_t - \eta_t \nabla_t)$
$u'1$ random partition: $\{1, \ldots, N\} = B_1 \cup \cdots \cup B_K$
$u'2$ median block: $P_{B_{k'}}(\ell_{u_{t+1}} - \ell_{u'_t}) = MOM_K(\ell_{u_{t+1}} - \ell_{u'_t})$
$u'3$ ascent direction: $\nabla'_t := -\nabla(u \to P_{B_{k'}}\ell_u)_{|u=u'_t}$
$u'4$ $u'_{t+1} = prox_{\lambda_K \|\cdot\|_1}\left(u'_t + \eta_t \nabla'_t\right)$

Problem: Find a top unit-eigenvector of $A \in \mathcal{S}_d(\mathbb{R})$.

$$x^* \in \underset{x \in \mathbb{R}^d, \|x\|_2 = 1}{\operatorname{argmax}} \langle A, xx^\top \rangle \tag{1}$$

Problem: Find a top unit-eigenvector of $A \in \mathcal{S}_d(\mathbb{R})$.

$$x^* \in \underset{x \in \mathbb{R}^d, \|x\|_2 = 1}{\operatorname{argmax}} \langle A, xx^\top \rangle \tag{1}$$

Fact: $\{xx^\top : x \in \mathbb{R}^d, \|x\|_2 = 1\} = \{X \in \mathbb{R}^{d \times d} : X \succeq 0, \mathbf{Tr}(X) = 1, \mathbf{rank}(X) = 1\}$

Problem: Find a top unit-eigenvector of $A \in \mathcal{S}_d(\mathbb{R})$.

$$x^* \in \operatorname*{argmax}_{x \in \mathbb{R}^d, \|x\|_2 = 1} \langle A, xx^\top \rangle \tag{1}$$

Fact: $\{xx^\top : x \in \mathbb{R}^d, \|x\|_2 = 1\} = \{X \in \mathbb{R}^{d \times d} : X \succeq 0, \mathbf{Tr}(X) = 1, \mathbf{rank}(X) = 1\}$

Lifting: quadratic functions of x are linear functions of $xx^\top$

$$(1) \Leftrightarrow x^*(x^*)^\top \in \operatorname*{argmax}_{\substack{X \in \mathbb{R}^{d \times d}, \\ X \succeq 0, \\ \mathbf{Tr}(X) = 1, \\ \mathbf{rank}(X) = 1}} \langle A, X \rangle$$

Problem: Find a top unit-eigenvector of $A \in \mathcal{S}_d(\mathbb{R})$.

$$x^* \in \underset{x \in \mathbb{R}^d, \|x\|_2 = 1}{\operatorname{argmax}} \langle A, xx^\top \rangle \tag{1}$$

Fact: $\{xx^\top : x \in \mathbb{R}^d, \|x\|_2 = 1\} = \{X \in \mathbb{R}^{d \times d} : X \succeq 0, \mathbf{Tr}(X) = 1, \mathbf{rank}(X) = 1\}$

SDP Relaxation:

$$X \in \underset{\substack{X \in \mathbb{R}^{d \times d} \\ X \succeq 0 \\ \mathbf{Tr}(X) = 1 \\ \cancel{\mathbf{rank}(X) = 1}}}{\operatorname{argmax}} \langle A, X \rangle.$$

Problem: Find a top unit-eigenvector of $A \in \mathcal{S}_d(\mathbb{R})$.

$$x^* \in \underset{x \in \mathbb{R}^d, \|x\|_2 = 1}{\operatorname{argmax}} \langle A, xx^\top \rangle \tag{1}$$

Fact: $\{xx^\top : x \in \mathbb{R}^d, \|x\|_2 = 1\} = \{X \in \mathbb{R}^{d \times d} : X \succeq 0, \mathbf{Tr}(X) = 1, \mathbf{rank}(X) = 1\}$

SDP Relaxation:

$$X \in \underset{\substack{X \in \mathbb{R}^{d \times d} \\ X \succeq 0 \\ \mathbf{Tr}(X) = 1}}{\operatorname{argmax}} \langle A, X \rangle.$$

Lifting + SDP relaxation: a classical strategy for many problems (Goemans and Williamson, 95') and a source of many **ERMs w.r.t. a linear loss function** for us.

Problem: estimate d unknown angles $\theta_1, \cdots, \theta_d$.

Data: a noisy pairwise measurements of their offsets:

$$\delta_{ij} = (\theta_i - \theta_j + \sigma g_{ij})[2\pi]$$

Problem: estimate d unknown angles $\theta_1, \cdots, \theta_d$.

Data: a noisy pairwise measurements of their offsets:

$$\delta_{ij} = (\theta_i - \theta_j + \sigma g_{ij})[2\pi]$$

Equivalent problem: estimate $x^* \in \mathbb{R}^d$ with coordinates $x_j^* := e^{\iota\theta_j}$

$\rightarrow$ we define $A := \left(e^{\iota\delta_{ij}}\right)_{1\leq i,j,\leq d} = S \circ [x^*(\overline{x^*})^\top]$, with $S = (S_{ij})_{d\times d}, S_{ij} = \begin{cases} e^{\iota\sigma g_{ij}} & \text{if } i < j \\ 1 & \text{if } i = j \\ e^{-\iota\sigma g_{ij}} & \text{if } i > j \end{cases}$

We have

$$x^* \in \underset{x\in\mathbb{C}^d:|x_i|=1}{\text{argmax}} \left\{\langle \mathbb{E}A, x\bar{x}^\top \rangle\right\} = \left\{(e^{\iota(\theta_i+\theta_0)})_{i=1}^d : \theta_0 \in [0, 2\pi)\right\}.$$

We have

$$x^* \in \underset{x\in\mathbb{C}^d : |x_i|=1}{\operatorname{argmax}} \left\{ \langle \mathbb{E}A, x\bar{x}^\top \rangle \right\} = \left\{ (e^{\iota(\theta_i+\theta_0)})_{i=1}^d : \theta_0 \in [0, 2\pi) \right\}.$$

Lifting

$$\left\{ x\bar{x}^\top : \; x \in \mathbb{C}^d : |x_i| = 1 \right\} = \{ Z \in \mathbb{H}_d : Z \succeq 0, \operatorname{diag}(Z) = \mathbf{1}_d, \operatorname{rank}(Z) = 1 \}$$

We have

$$x^* \in \underset{x\in\mathbb{C}^d:|x_i|=1}{\operatorname{argmax}} \left\{\langle \mathbb{E}A, x\bar{x}^\top\rangle\right\} = \left\{\left(e^{\iota(\theta_i+\theta_0)}\right)_{i=1}^d : \theta_0 \in [0,2\pi)\right\}.$$

Lifting

$$\left\{x\bar{x}^\top : x \in \mathbb{C}^d : |x_i| = 1\right\} = \{Z \in \mathbb{H}_d : Z \succeq 0, \operatorname{diag}(Z) = \mathbf{1}_d, \operatorname{rank}(Z) = 1\}$$

Relaxation

$$Z^* \in \underset{\substack{Z\in\mathbb{H}_d,\\ Z\succeq 0,\\ \operatorname{diag}(Z)=\mathbf{1}_d}}{\operatorname{argmax}} \langle \mathbb{E}A, Z\rangle$$

x^* can be obtain as a first eigen-vector of Z^* up to a global rotational shift $e^{\iota\theta_0}$ of the coordinates.

We have

$$x^* \in \operatorname*{argmax}_{x\in\mathbb{C}^d : |x_i|=1} \left\{\langle \mathbb{E}A, x\bar{x}^\top\rangle\right\} = \left\{(e^{\iota(\theta_i+\theta_0)})_{i=1}^d : \theta_0 \in [0, 2\pi)\right\}.$$

Lifting

$$\left\{x\bar{x}^\top : x \in \mathbb{C}^d : |x_i| = 1\right\} = \{Z \in \mathbb{H}_d : Z \succeq 0, \operatorname{diag}(Z) = \mathbf{1}_d, \operatorname{rank}(Z) = 1\}$$

Estimator

$$\hat{Z} \in \operatorname*{argmax}_{\substack{Z\in\mathbb{H}_d,\\ Z\succeq 0,\\ \operatorname{diag}(Z)=\mathbf{1}_d}} \langle A, Z\rangle$$

(An ERM w.r.t. a linear loss function).

- We observe $X = |Bx| \in \mathbb{R}^N$, with $B \in \mathbb{C}^{N\times d}$, random matrix.
- Aim: recover x from the observation of X.
- Equivalent problem: (d'Aspremont et al.)

$$z^* \in \underset{\substack{z\in\mathbb{C}^N \\ |z_i|=1,\forall i\in[N]}}{\operatorname{argmin}} \langle \mathbb{E}[A], z\bar{z}^\top \rangle,$$

where $A := \left(XX^\top\right) \circ \left(I_N - BB^+\right)$, B^+ pseudo-inverse of B.

- We observe $X = |Bx| \in \mathbb{R}^N$, with $B \in \mathbb{C}^{N\times d}$, random matrix.
- Aim: recover x from the observation of X.
- Equivalent problem: (d'Aspremont et al.)

$$z^* \in \underset{\substack{z\in\mathbb{C}^N \\ |z_i|=1\,,\forall i\in[N]}}{\operatorname{argmin}} \langle \mathbb{E}[A], z\bar{z}^\top \rangle,$$

where $A := (XX^\top) \circ (I_N - BB^+)$, B^+ pseudo-inverse of B.

- Lifting

$$\min_{\substack{Z\succeq 0 \\ Z_{ii}=1 \forall i \\ rank(Z)=1}} \langle \mathbb{E}[A], Z \rangle$$

- We observe $X = |Bx| \in \mathbb{R}^N$, with $B \in \mathbb{C}^{N\times d}$, random matrix.
- Aim: recover x from the observation of X.
- Equivalent problem: (d'Aspremont et al.)

$$z^* \in \underset{\substack{z\in\mathbb{C}^N \\ |z_i|=1\,,\forall i\in[N]}}{\operatorname{argmin}} \left\langle \mathbb{E}[A], z\bar{z}^\top \right\rangle,$$

where $A := \left(XX^\top\right) \circ \left(I_N - BB^+\right)$, B^+ pseudo-inverse of B.

- Relaxation:

$$Z^* \in \underset{\substack{Z\succeq 0 \\ Z_{ii}=1\forall i}}{\operatorname{argmin}} \langle \mathbb{E}[A], Z\rangle \quad \Big| \quad z^* : 1^{st} \text{ eigen vector of } Z^*$$

- Estimator

$$\hat{Z} \in \underset{\substack{Z\succeq 0 \\ Z_{ii}=1\forall i}}{\operatorname{argmin}} \langle A, Z\rangle \quad \Big| \quad \hat{z} : 1^{st} \text{ eigen vector of } \hat{Z}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- We observe $X_1, \cdots, X_N$, independant copies of $X \in \mathbb{R}^{d\times d}$
  $\Sigma := cov(X)$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- We observe $X_1, \cdots, X_N$, independant copies of $X \in \mathbb{R}^{d\times d}$
  $\Sigma := cov(X)$
- Aim: find a partition $G = \{G_1, \dots, G_K\}$ of $\{1, \dots, d\}$ that separates the components of X.

Membership matrix:

$$Q := \left(1_{\{a\in G_k\}}\right)_{\substack{1\le a\le d\\ 1\le k\le K}}$$

Partnership matrix:

$$Z := \left(|G_k|^{-1} 1_{\{i\in G_k, j\in G_k\}}\right)_{1\le i,j\le d}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- We observe $X_1, \cdots, X_N$, independant copies of $X \in \mathbb{R}^{d\times d}$ $\Sigma := cov(X)$
- Aim: find a partition $G = \{G_1, \dots, G_K\}$ of $\{1, \dots, d\}$ that separates the components of X.

Membership matrix:

$$Q := \left(1_{\{a \in G_k\}}\right)_{\substack{1\leq a\leq d \\ 1\leq k\leq K}}$$

Partnership matrix:

$$Z := \left(|G_k|^{-1} 1_{\{i\in G_k, j\in G_k\}}\right)_{1\leq i,j\leq d}$$

- exact G-block covariance model: $\Sigma = QCQ^\top + \Gamma$, with $C \in \mathcal{S}_K(\mathbb{R})$ and Γ diagonal.

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle \mathbb{E} A, Z \rangle$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \operatorname*{argmax}_{Z \in \mathcal{C}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \operatorname*{argmax}_{Z \in \mathcal{C}} \langle \mathbb{E} A, Z \rangle$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle \mathbb{E} A, Z \rangle$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle \mathbb{E}A, Z \rangle$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \underset{Z \in \mathcal{C}}{\operatorname{argmax}} \langle \mathbb{E}A, Z \rangle$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_j \\ \vdots \\ X_d \end{bmatrix}$$

- Estimator (Verzelen, Giraud et al.)

$$\hat{Z} \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle A, Z \rangle, \qquad \mathcal{C} := \left\{ Z \succeq 0, Z \geq 0, \sum_j Z_{ij} = 1 \forall i, Tr\left(Z\right) = K \right\}$$

- Noiseless case:

$$Z^* \in \underset{Z \in \mathcal{C}}{\mathrm{argmax}} \langle \mathbb{E}A, Z \rangle$$

3 | Key technical tools.

Reminders

Complexity fixed point parameter

$$r_G^*(\delta) = \inf\left(r > 0 : \mathbb{P}\left[\sup_{Z\in\mathcal{C}:G(Z-Z^*)\leq r}(P_N - P)\mathcal{L}_Z \leq \frac{r}{2}\right] \geq 1-\delta\right)$$

Local curvature assumption

For all $Z \in \mathcal{C}$, if $P\mathcal{L}_Z \leq r_G^*(\delta)$ then $P\mathcal{L}_Z \geq G(Z^* - Z)$.

Theorem | Excess risk for the ERM estimator with G-localization

We assume that $\mathcal{C}$ is star-shaped in Z^* and that the local curvature Assumption holds for some $0 < \delta < 1$. Then, with probability at least $1-\delta$, it holds true that

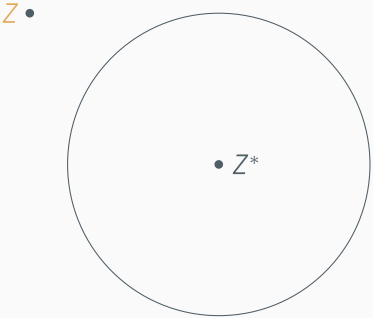
$$G(Z^* - \hat{Z}) \leq P\mathcal{L}_{\hat{Z}} \leq r_G^*(\delta).$$

Assume $r^* > 0$

- Define $\Omega^* := \{\forall Z \in \mathcal{C} \text{ s.t } \langle \mathbb{E}[A], Z^* - Z\rangle \leq r^*, \langle A - \mathbb{E}[A], Z - Z^*\rangle \leq r^*/2\}$
  By definition of r^*, Ω^* holds with probability at least $1 - \delta$.

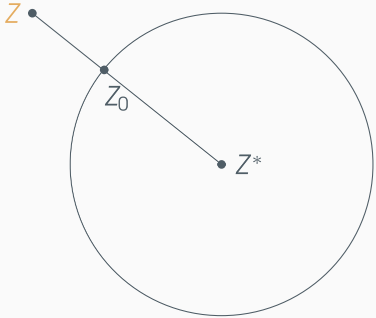
Assume $r^* > 0$

- Define $\Omega^* := \{\forall Z \in \mathcal{C} \text{ s.t } \langle \mathbb{E}[A], Z^* - Z\rangle \leq r^*, \langle A - \mathbb{E}[A], Z - Z^*\rangle \leq r^*/2\}$
  By definition of r^*, Ω^* holds with probability at least $1 - \delta$.
- Consider $Z \in \mathcal{C}$ such that $\langle \mathbb{E}[A], Z^* - Z\rangle > r^*$.



Assume $r^* > 0$

- Define $\Omega^* := \{\forall Z \in \mathcal{C} \text{ s.t } \langle \mathbb{E}[A], Z^* - Z\rangle \leq r^*, \langle A - \mathbb{E}[A], Z - Z^*\rangle \leq r^*/2\}$
  By definition of r^*, Ω^* holds with probability at least $1 - \delta$.
- Consider $Z \in \mathcal{C}$ such that $\langle \mathbb{E}[A], Z^* - Z\rangle > r^*$.

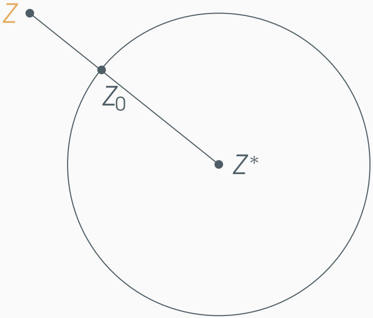


Star-shaped property of $\mathcal{C}$:

$\exists\, 0 < \lambda < 1$, $Z_0 \in \mathcal{C}$ s.t $Z_0 - Z^* = \lambda(Z - Z^*)$ and $\langle \mathbb{E}[A], Z^* - Z_0\rangle = r^*$.

Assume $r^* > 0$

- Define $\Omega^* := \{\forall Z \in \mathcal{C} \text{ s.t } \langle \mathbb{E}[A], Z^* - Z\rangle \leq r^*, \langle A - \mathbb{E}[A], Z - Z^*\rangle \leq r^*/2\}$
  By definition of r^*, Ω^* holds with probability at least $1 - \delta$.
- Consider $Z \in \mathcal{C}$ such that $\langle \mathbb{E}[A], Z^* - Z\rangle > r^*$.



Star-shaped property of $\mathcal{C}$:

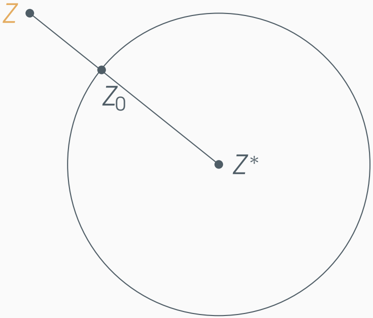
$\exists\, 0 < \lambda < 1,\ Z_0 \in \mathcal{C}$ s.t $Z_0 - Z^* = \lambda(Z - Z^*)$ and $\langle \mathbb{E}[A], Z^* - Z_0\rangle = r^*$.

Homogeneity of the loss function:

On $\Omega^* \Rightarrow \langle A - \mathbb{E}[A], Z_0 - Z^*\rangle \leq r^*/2$
and then $\langle A, Z - Z^*\rangle \leq -r^*/2 < 0$

Assume $r^* > 0$

- Define $\Omega^* := \{\forall Z \in \mathcal{C} \text{ s.t } \langle \mathbb{E}[A], Z^* - Z\rangle \leq r^*, \langle A - \mathbb{E}[A], Z - Z^*\rangle \leq r^*/2\}$
  By definition of r^*, Ω^* holds with probability at least $1-\delta$.
- Consider $Z \in \mathcal{C}$ such that $\langle \mathbb{E}[A], Z^* - Z\rangle > r^*$.



Star-shaped property of $\mathcal{C}$:

$\exists\, 0 < \lambda < 1$, $Z_0 \in \mathcal{C}$ s.t $Z_0 - Z^* = \lambda(Z - Z^*)$ and $\langle \mathbb{E}[A], Z^* - Z_0\rangle = r^*$.

Homogeneity of the loss function:
On $\Omega^* \Rightarrow \langle A - \mathbb{E}[A], Z_0 - Z^*\rangle \leq r^*/2$
and then $\langle A, Z - Z^*\rangle \leq -r^*/2 < 0$
$\rightarrow Z$ cannot be a maximizer of $Z \rightarrow P_N \ell_Z$ on $\mathcal{C}$

- Necessarily, $\langle \mathbb{E}[A], Z^* - \hat{Z} \rangle \leq r^*$.

- Necessarily, $\langle \mathbb{E}[A], Z^* - \hat{Z} \rangle \leq r^*$.
- From the curvature equation, we get that $G(Z^* - \hat{Z}) \leq P\mathcal{L}_{\hat{Z}}$.

- Necessarily, $\langle \mathbb{E}[A], Z^* - \hat{Z} \rangle \leq r^*$.
- From the curvature equation, we get that $G(Z^* - \hat{Z}) \leq P\mathcal{L}_{\hat{Z}}$.

Finally :

$$G(Z^* - \hat{Z}) \leq P\mathcal{L}_{\hat{Z}} \leq r_G^*(\delta)$$

.

4 | Appendix.

Reminders

Assumption | Local curvature assumption

There exists $A > 0$, $\gamma > 0$ and $\rho^* > 0$ such that ρ^* satisfies the spartsity equation and for both $b \in \{1, 2\}$ and all $Z \in \mathcal{C}$, if $\|Z - Z^*\|_2^2 = r^*_{\mathrm{RMOM},\mathrm{G}}(\gamma^*, b\rho^*)^2$ and $\|Z - Z^*\| \leq b\rho^*$, then $AP\mathcal{L}_Z \geq \|Z - Z^*\|_2^2$.

Theorem | Excess risk and estimation bounds for the RMOM with ℓ_2-localization

Grant Assumptions (i) to (v). Then, with probability at least $1 - 2\exp(-72K/625)$, the following bounds hold for the RMOM estimator with ℓ_2-localization:

$$\|\hat{Z}^{\mathrm{RMOM}}_{K,\lambda} - Z^*\| \leq 2\rho^* \quad , \quad \|\hat{Z}^{\mathrm{RMOM}}_{K,\lambda} - Z^*\|_2^2 \leq r^*_{\mathrm{RMOM},\mathrm{G}}(\gamma, 2\rho^*)^2 \ \ and \ \ P\mathcal{L}_{\hat{Z}^{\mathrm{RMOM}}_{K,\lambda}} \leq \frac{93}{100} r^*_{\mathrm{RMOM},\mathrm{G}}(\gamma, 2\rho^*)^2.$$

Structure of the proof

- For $b \in \{1, 2\}$, we define $r_b^* = r^*_{\mathrm{RMOM},G}(\gamma, b\rho^*)$, and $\mathcal{B}_b := \{Z \in \mathcal{C} : G(Z - Z^*) \leq (r_b^*)^2 \text{ and } \|Z - Z^*\| \leq b\rho^*\}$.
- We consider the event

$$\Omega_{K,G} = \left\{ \forall b \in \{1,2\}, \forall Z \in \mathcal{B}_b,\ \sum_{k=1}^{K} 1\left(|(P_{B_k} - P)\mathcal{L}_Z| \leq \frac{1}{20}(r_b^*)^2 \right) > \frac{K}{2} \right\}.$$

- Two steps

1| $\hat{Z}$ has the desired properties on $\Omega_{K,G}$.

2| $\Omega_{K,G}$ holds with high probability.

Step 1 |

Lemma 1.

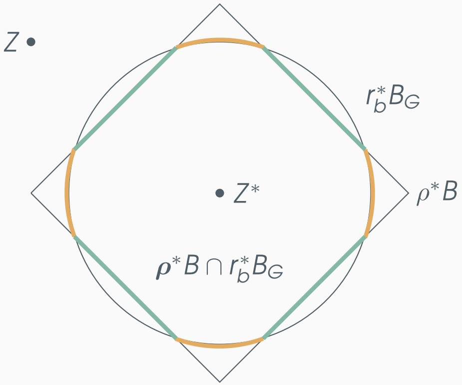
If there exists $\eta > 0$ such that

$$\sup_{Z \in \mathcal{C} \setminus \mathcal{B}_2} \mathrm{MOM}_K(\ell_{Z^*} - \ell_Z) + \lambda(\|Z^*\| - \|Z\|) < -\eta \tag{2}$$

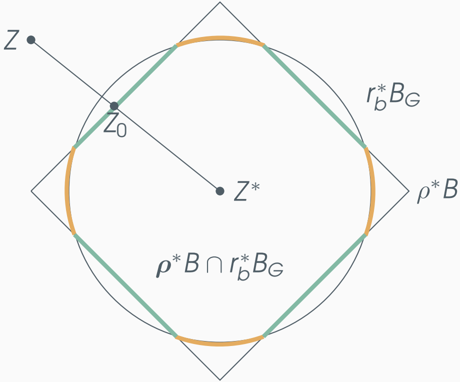
and

$$\sup_{Z \in \mathcal{C}} \mathrm{MOM}_K(\ell_{Z^*} - \ell_Z) + \lambda(\|Z^*\| - \|Z\|) \leq \eta \tag{3}$$

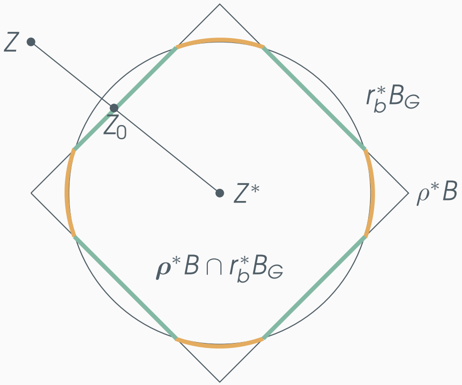
then $\|\hat{Z} - Z^*\| \leq 2\rho^*$ and $G(\hat{Z} - Z^*) \leq r^*_{\mathrm{RMOM},G}(\gamma, 2\rho^*)^2$.



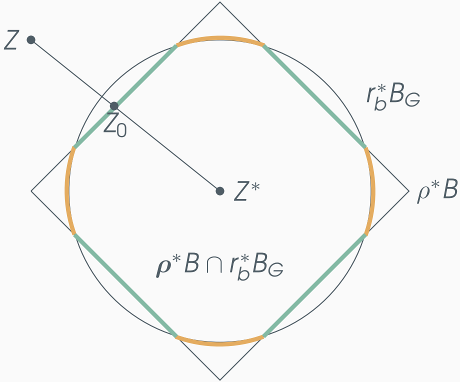
- Consider $Z \in \mathcal{C} \backslash \mathcal{B}_b$.



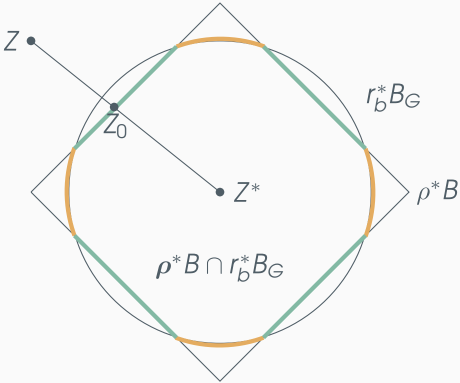
- Consider $Z \in \mathcal{C} \backslash \mathcal{B}_b$.
- Regularity of G and star-shaped property of $\mathcal{C} \to \exists Z_0 \in \partial \mathcal{B}_b$ and $\alpha > 1$ such that
$$Z = Z^* + \alpha(Z_0 - Z^*).$$



- Consider $Z \in \mathcal{C} \backslash \mathcal{B}_b$.
- Regularity of G and star-shaped property of $\mathcal{C} \rightarrow$ $\exists Z_0 \in \partial \mathcal{B}_b$ and $\alpha > 1$ such that
$$Z = Z^* + \alpha(Z_0 - Z^*).$$
- Linearity of ℓ_Z and convexity of $\|\cdot\|$:
$$\forall k \in [K], P_{B_k} \mathcal{L}_Z^\lambda \geq \alpha P_{B_k} \mathcal{L}_{Z_0}^\lambda. \tag{4}$$

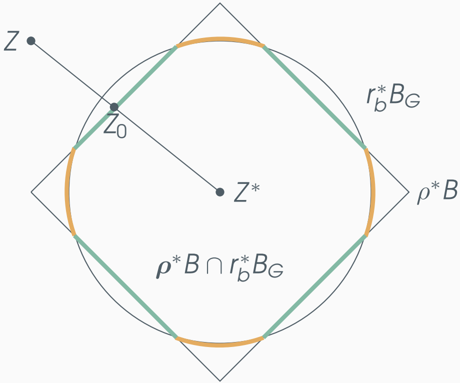


- Consider $Z \in \mathcal{C} \backslash \mathcal{B}_b$.
- Regularity of G and star-shaped property of $\mathcal{C} \to \exists Z_0 \in \partial \mathcal{B}_b$ and $\alpha > 1$ such that
$$Z = Z^* + \alpha(Z_0 - Z^*).$$
- Linearity of ℓ_Z and convexity of $\|\cdot\|$:
$$\forall k \in [K], P_{B_k} \mathcal{L}_Z^\lambda \geq \alpha P_{B_k} \mathcal{L}_{Z_0}^\lambda. \tag{4}$$
- Either a $\Big|\, G(Z_0 - Z^*) = (r_b^*)^2$ and $\|Z_0 - Z^*\| < b\rho^*$
  or b $\Big|\, G(Z_0 - Z^*) < (r_b^*)^2$ and $\|Z_0 - Z^*\| = b\rho^*$.



a | On $\Omega_{K,G}$, using the local curvature assumption, we find at least $K/2$ blocks on which

$$P_{B_k} \mathcal{L}_{Z_0}^{\lambda} \geq \begin{cases} (r_2^*)^2/5 & \text{for } b = 1 \\ 2(r_2^*)^2/5 & \text{for } b = 2 \end{cases} \tag{5}$$

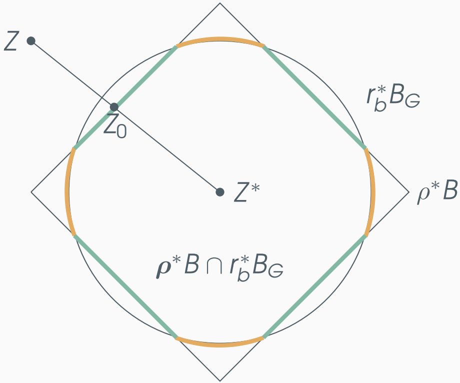


a | On $\Omega_{K,G}$, using the local curvature assumption, we find at least $K/2$ blocks on which

$$P_{B_k}\mathcal{L}_{Z_0}^{\lambda} \geq \begin{cases} (r_2^*)^2/5 & \text{for } b = 1 \\ 2(r_2^*)^2/5 & \text{for } b = 2 \end{cases} \tag{5}$$

b | Here, $Z_0 \in \bar{H}_{b\rho^*}$. The sparsity equation also holds for $\rho = b\rho^*$. Then, on $\Omega_{K,G}$, we find at least $K/2$ blocks on which

$$P_{B_k}\mathcal{L}_{Z_0}^{\lambda} \geq \begin{cases} 57(r_2^*)^2/400 & \text{for } b = 1 \\ 134(r_2^*)^2/400 & \text{for } b = 2 \end{cases} \tag{6}$$



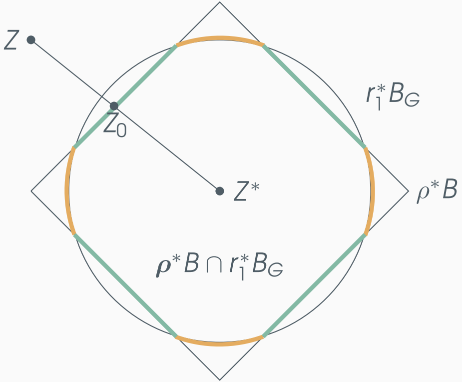
a | On $\Omega_{K,G}$, using the local curvature assumption, we find at least $K/2$ blocks on which

$$P_{B_k}\mathcal{L}_{Z_0}^{\lambda} \geq \begin{cases} (r_2^*)^2/5 & \text{for } b=1 \\ 2(r_2^*)^2/5 & \text{for } b=2 \end{cases} \tag{5}$$

b | Here, $Z_0 \in \bar{H}_{b\rho^*}$. The sparsity equation also holds for $\rho = b\rho^*$. Then, on $\Omega_{K,G}$, we find at least $K/2$ blocks on which

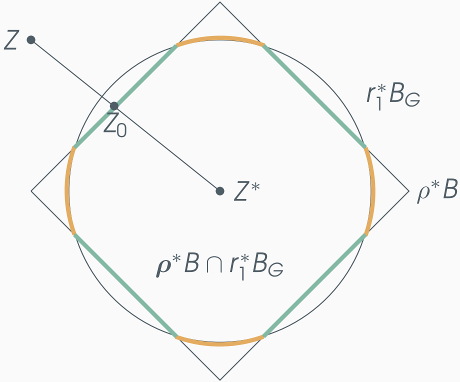
$$P_{B_k}\mathcal{L}_{Z_0}^{\lambda} \geq \begin{cases} 57(r_2^*)^2/400 & \text{for } b=1 \\ 134(r_2^*)^2/400 & \text{for } b=2 \end{cases} \tag{6}$$

- From (4), (5) and (6) with $b=2$, we conclude that (2) holds for any $\eta < (134/400)(r_2^*)^2$

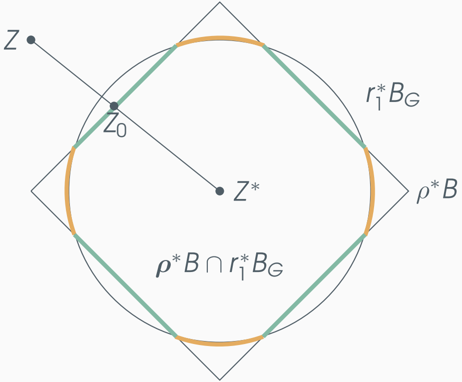


- Consider $Z \in \mathcal{B}_1$. On $\Omega_{K,G}$, we find at least $K/2$ blocks on which

$$P_{B_k} \mathcal{L}_Z^\lambda \geq -13(r_2^*)^2/40 \tag{7}$$



- Consider $Z \in \mathcal{B}_1$. On $\Omega_{K,G}$, we find at least $K/2$ blocks on which
$$P_{B_k} \mathcal{L}_Z^\lambda \geq -13(r_2^*)^2/40 \tag{7}$$
- From (4), (5) and (6) with $b = 1$, we conclude that (3) holds with $\eta = (13/40)(r_2^*)^2$



- Consider $Z \in \mathcal{B}_1$. On $\Omega_{K,G}$, we find at least $K/2$ blocks on which

$$P_{B_k} \mathcal{L}_Z^\lambda \geq -13(r_2^*)^2/40 \tag{7}$$

- From (4), (5) and (6) with $b = 1$, we conclude that (3) holds with $\eta = (13/40)(r_2^*)^2$

Finally, both (2) and (3) hold for $\eta = (33/100)(r_2^*)^2$.

$\rightarrow$ We conclude from Lemma 1. that $\|\hat{Z} - Z^*\| \leq 2\rho^*$ and $G(\hat{Z} - Z^*) \leq r^*_{\mathrm{RMOM},G}(\gamma, 2\rho^*)^2$.

Bound on $P\mathcal{L}_{\hat{Z}}$

- From the above: $\hat{Z} \in \mathcal{B}_2 \to$ on $\Omega_{K,G}$ there exist stricly more than $K/2$ blocks on which

$$P\mathcal{L}_{\hat{Z}} \leq P_{B_k}\mathcal{L}_{\hat{Z}} + (r_2^*)^2/(20). \tag{8}$$

Bound on $P\mathcal{L}_{\hat{Z}}$

- From the above: $\hat{Z} \in \mathcal{B}_2 \to$ on $\Omega_{K,\mathcal{G}}$ there exist stricly more than $K/2$ blocks on which

$$P\mathcal{L}_{\hat{Z}} \leq P_{B_k}\mathcal{L}_{\hat{Z}} + (r_2^*)^2/(20). \tag{8}$$

- From (3) with $\eta = (33/100)(r_2^*)^2$ and the definition of $\hat{Z}$, we find at least $K/2$ blocks B_k on which

$$P_{B_k}\mathcal{L}_{\hat{Z}} \leq \frac{88}{100}(r_2^*)^2. \tag{9}$$

Bound on $P\mathcal{L}_{\hat{Z}}$

- From the above: $\hat{Z} \in \mathcal{B}_2 \to$ on $\Omega_{K,G}$ there exist stricly more than $K/2$ blocks on which

$$P\mathcal{L}_{\hat{Z}} \leq P_{B_k}\mathcal{L}_{\hat{Z}} + (r_2^*)^2/(20). \tag{8}$$

- From (3) with $\eta = (33/100)(r_2^*)^2$ and the definition of $\hat{Z}$, we find at least $K/2$ blocks B_k on which

$$P_{B_k}\mathcal{L}_{\hat{Z}} \leq \frac{88}{100}(r_2^*)^2. \tag{9}$$

- Then, there exists at least one on which (8) and (9) hold simultaneously, which gives

$$P\mathcal{L}_{\hat{Z}} \leq \frac{93}{100}(r_2^*)^2.$$

Bound on $P\mathcal{L}_{\hat{Z}}$

- From the above: $\hat{Z} \in \mathcal{B}_2 \rightarrow$ on $\Omega_{K,G}$ there exist stricly more than $K/2$ blocks on which

$$P\mathcal{L}_{\hat{Z}} \leq P_{B_k}\mathcal{L}_{\hat{Z}} + (r_2^*)^2/(20). \tag{8}$$

- From (3) with $\eta = (33/100)(r_2^*)^2$ and the definition of $\hat{Z}$, we find at least $K/2$ blocks B_k on which

$$P_{B_k}\mathcal{L}_{\hat{Z}} \leq \frac{88}{100}(r_2^*)^2. \tag{9}$$

- Then, there exists at least one on which (8) and (9) hold simultaneously, which gives

$$P\mathcal{L}_{\hat{Z}} \leq \frac{93}{100}(r_2^*)^2.$$

$\rightarrow$ On $\Omega_{K,G}$, $\hat{Z}$ has the announced properties.

Step 2 |

Lemma

Assume that $K \geq 100|\mathcal{O}|$, *and let* $\rho^* > 0$ *be such that it satisfies the sparsity equation. Then,* $\Omega_{K,G}$ *holds with probability at least* $1 - 2\exp(-72K/625)$.

Definitions

- For $k \in [K]$:
  $W_k := \{X_i : i \in B_k\}$ and $F_Z(W_k) = (P_{B_k} - P)\mathcal{L}_Z$.
  $\widetilde{W}_k := \left\{\widetilde{X}_i : i \in B_k\right\}$ and $F_Z(\widetilde{W}_k) = (\widetilde{P_{B_k}} - \widetilde{P})\mathcal{L}_Z$, their non-corrupted counterparts.
- For $b \in \{1, 2\}$:

$$Z \to \psi_b(Z) = \sum_{k \in [K]} 1_{\left\{|F_Z(W_k)| \leq \frac{(r_b^*)^2}{20}\right\}}.$$

- $\phi : t \in \mathbb{R} \to 1_{\{t \geq 1\}} + 2(t - 1/2)1_{\{1/2 \leq t \leq 1\}}$.
- We show that

$$(\Omega_{K,G} \text{ holds w.h.p}) \iff \left(\psi_b(Z) \leq \frac{49K}{400} \text{ w.h.p for } Z \in \mathcal{B}_b \text{ and } b \in \{1, 2\}\right).$$

Definitions

- For $k \in [K]$:
  $W_k := \{X_i : i \in B_k\}$ and $F_Z(W_k) = (P_{B_k} - P)\mathcal{L}_Z$.
  $\widetilde{W}_k := \left\{\widetilde{X}_i : i \in B_k\right\}$ and $F_Z(\widetilde{W}_k) = (\widetilde{P_{B_k}} - \widetilde{P})\mathcal{L}_Z$, their non-corrupted counterparts.
- For $b \in \{1, 2\}$:

$$Z \to \psi_b(Z) = \sum_{k \in [K]} 1_{\left\{|F_Z(W_k)| \leq \frac{(r_b^*)^2}{20}\right\}}.$$

- $\phi : t \in \mathbb{R} \to 1_{\{t \geq 1\}} + 2(t - 1/2)1_{\{1/2 \leq t \leq 1\}}$.
- We show that

$$(\Omega_{K,G} \text{ holds w.h.p}) \iff (\psi_b(Z) \leq \tfrac{49K}{400} \text{ w.h.p for } Z \in \mathcal{B}_b \text{ and } b \in \{1, 2\}).$$

$\rightarrow$ We bound $\psi_b(Z)$ with high probability.

Bounding $\psi_b(Z)$

In what follows, we consider $b \in \{1, 2\}$ and $Z \in \mathcal{B}_b$.

Bounding $\psi_b(Z)$

In what follows, we consider $b \in \{1, 2\}$ and $Z \in \mathcal{B}_b$.

Basic manipulations lead to

$$\psi_b(Z) \leq \sup_{Z \in \mathcal{B}_b} \left(\sum_{k \in [K]} \phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2} \right) - \mathbb{E}\left[\phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2} \right) \right] \right) + \sum_{k \in [K]} \mathbb{P}\left(|F_Z(\widetilde{W}_k)| > \frac{(r_b^*)^2}{40} \right)$$
$$= S_b^{(1)}(Z) + S_b^{(2)}(Z). \tag{10}$$

Bounding $\psi_b(Z)$

In what follows, we consider $b \in \{1, 2\}$ and $Z \in \mathcal{B}_b$.

Basic manipulations lead to

$$\psi_b(Z) \leq \sup_{Z \in \mathcal{B}_b} \left(\sum_{k \in [K]} \phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2} \right) - \mathbb{E}\left[\phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2} \right) \right] \right) + \sum_{k \in [K]} \mathbb{P}\left(|F_Z(\widetilde{W}_k)| > \frac{(r_b^*)^2}{40} \right)$$
$$= S_b^{(1)}(Z) + S_b^{(2)}(Z). \tag{10}$$

We bound each term separately.

Bounding $\psi_b(Z)$ | Bounding $S_b^{(2)}(Z)$

For each $k \in [K]$, Markov's inequality and the definition of r_b^* yield to

$$\mathbb{P}\left(|F_Z(\widetilde{W}_k)| > \frac{(r_b^*)^2}{40}\right) \leq \frac{1600^2}{(r_b^*)^4}\left(V_K(r_b^*)\right)^2 \leq \frac{1}{200},$$

so that

$$S_b^{(2)}(Z) \leq \frac{K}{200}. \tag{11}$$

Bounding $\psi_b(Z)$ | Bounding $S_b^{(1)}(Z)$

- Mc Diarmind inequality with $t = 12/25 \to$ w.p.a.l $1 - e^{-72K/625}$

$$S_b^{(1)}(Z) \leq \frac{12K}{25} + \mathbb{E}\left[\sup_{Z \in \mathcal{B}_b} \sum_{k\in[K]} \phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2}\right) - \mathbb{E}\phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2}\right)\right]$$
$$:= \frac{12K}{25} + T_b^{(1)}(Z). \tag{12}$$

- $\epsilon_1, \ldots, \epsilon_K$, Rademacher variables independant from the $\widetilde{X}_i$'s. By the symmetrization Lemma:

$$T_b^{(1)}(Z) \leq 2\mathbb{E}\left[\sup_{Z \in \mathcal{B}_b} \sum_{k\in[K]} \epsilon_k \phi\left(\frac{20|F_Z(\widetilde{W}_k)|}{(r_b^*)^2}\right)\right] := 2T_b^{(2)}(Z). \tag{13}$$

- ϕ is Lipschitz with $\phi(0) = 0 \rightarrow$ by the contraction Lemma:

$$T_b^{(2)}(Z) \leq \frac{40}{(r_b^*)^2}\mathbb{E}\left[\sup_{Z\in\mathcal{B}_b}\sum_{k\in[K]}\epsilon_k(\widetilde{P_{B_k}} - \widetilde{P})\mathcal{L}_Z\right] := \frac{40}{(r_b^*)^2}T_b^{(3)}(Z) \tag{14}$$

- $(\sigma_i)_{i=1,\dots,N}$ Rademacher variables independant from the $\widetilde{X}_i$'s and the ϵ_i's. $\rightarrow$ By the symmetrization Lemma:

$$T_b^{(3)}(Z) \leq 2KE_G(r_b^*, b\rho^*) \leq 2K\gamma(r_b^*)^2. \tag{15}$$

Combining (12), (13), (14) and (15), we get that w.p.a.l $1 - e^{-72K/625}$

$$S_b^{(1)}(Z) \leq \left(\frac{12}{25} + 160\gamma\right)K. \tag{16}$$

Bounding $\psi_b(Z)$

Finally, combining (10), (11) and (16), we get that with probability at least $1 - \exp(-72K/625)$:

$$\psi_b(Z) \leq \frac{49}{400}K.$$

This must be verified for both $b = 1$ and 2, so the final probability is $1 - 2\exp(-72K/625)$, which concludes the proof.