Linear Size Sparsifier & Geometry of $\|\cdot\|_{\text{op}}$ Ball

Victor Reis

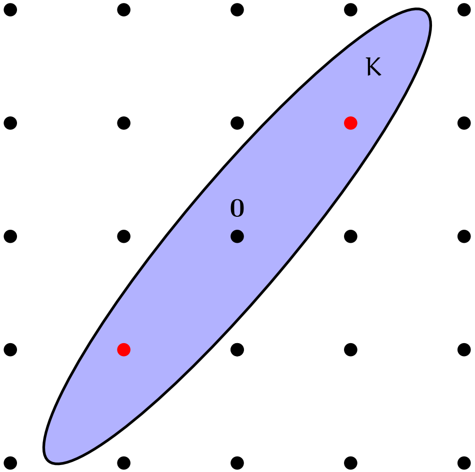
SODA 2020, January 8

Joint work with Thomas Rothvoss

UNIVERSITY *of* WASHINGTON

Warmup: Minkowski's Theorem (1889)

- Any symmetric convex $K \subset \mathbb{R}^m$ with volume $> 2^m$ intersects $\mathbb{Z}^m \setminus \{\mathbf{0}\}$



Non-constructive Partial Coloring

Theorem (Gluskin, 1989)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\mathrm{vol}(K \cap [-1,1]^m) \geqslant C^m$ for $C > 1$ there exists $\boldsymbol{y} \in 2K \cap \{-1,0,1\}^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$.

Non-constructive Partial Coloring

Theorem (Gluskin, 1989)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\mathrm{vol}(K \cap [-1,1]^m) \geqslant C^m$ for $C > 1$ there exists $\boldsymbol{y} \in 2K \cap \{-1,0,1\}^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$.

- Define $\gamma_m(K) = \Pr[gaussian \in K]$

Non-constructive Partial Coloring

Theorem (Gluskin, 1989)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\mathrm{vol}(K \cap [-1,1]^m) \geqslant C^m$ for $C > 1$ there exists $\boldsymbol{y} \in 2K \cap \{-1,0,1\}^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$.

- Define $\gamma_m(K) = \Pr[\mathit{gaussian} \in K]$
- Can show $\mathrm{vol}(K \cap [-1,1]^m) \geqslant 2^m \gamma_m(K)$

Non-constructive Partial Coloring

Theorem (Gluskin, 1989)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\mathrm{vol}(K \cap [-1,1]^m) \geqslant C^m$ for $C > 1$ there exists $\boldsymbol{y} \in 2K \cap \{-1,0,1\}^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$.

- Define $\gamma_m(K) = \Pr[\mathit{gaussian} \in K]$
- Can show $\mathrm{vol}(K \cap [-1,1]^m) \geqslant 2^m \gamma_m(K)$

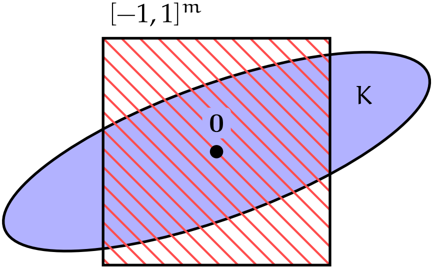
Corollary

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\gamma_m(K) \geqslant 0.51^m$ there exists $\boldsymbol{y} \in 2K \cap \{-1,0,1\}^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$.

Constructive Partial Coloring

Theorem (Rothvoss, FOCS 2014)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\gamma_m(K) \geqslant 0.999^m$ we can find $\boldsymbol{y} \in K \cap [-1,1]^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$ in **polynomial time**.



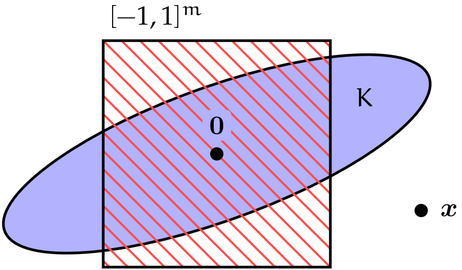
Constructive Partial Coloring

Theorem (Rothvoss, FOCS 2014)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\gamma_m(K) \geqslant 0.999^m$ we can find $\boldsymbol{y} \in K \cap [-1,1]^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$ in **polynomial time**.

Algorithm:

(1) Sample $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_m)$



Constructive Partial Coloring

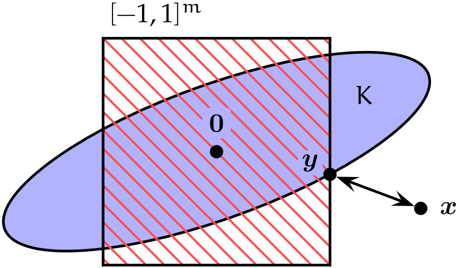
Theorem (Rothvoss, FOCS 2014)

Given symmetric convex $K \subseteq \mathbb{R}^m$ with $\gamma_m(K) \geqslant 0.999^m$ we can find $\boldsymbol{y} \in K \cap [-1,1]^m$ with $\Omega(m)$ coordinates in $\{-1,1\}$ in **polynomial time**.

Algorithm:

(1) Sample $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{I}_m)$

(2) Output the projection of $\boldsymbol{x}$ onto $K \cap [-1,1]^m$.



Motivating question

Question

How can we show good measure lower bounds for convex bodies?

Motivating question

Question

How can we show good measure lower bounds for convex bodies?

- For finite intersections can use $\gamma_m(K \cap L) \geqslant \gamma_m(K)\gamma(L)$

Motivating question

Question

How can we show good measure lower bounds for convex bodies?

- For finite intersections can use $\gamma_m(K \cap L) \geqslant \gamma_m(K)\gamma(L)$
- How about non-polyhedral sets?

Motivating question

Question

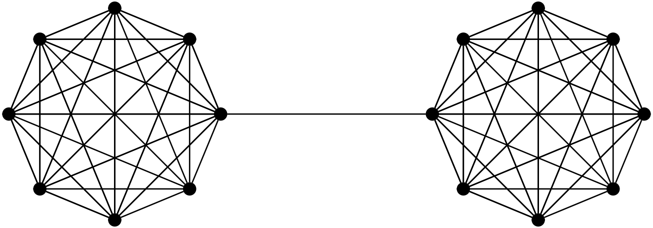
How can we show good measure lower bounds for convex bodies?

- For finite intersections can use $\gamma_m(K \cap L) \geqslant \gamma_m(K)\gamma(L)$
- How about non-polyhedral sets?
- This work: lower bounds for sets of sparsifying weights

Graph Sparsification

Theorem (Batson-Spielman-Srivastava '08)

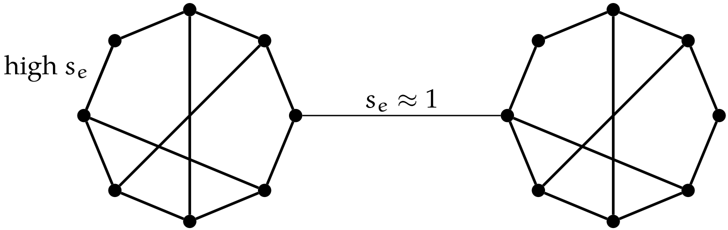
Given a graph $G = (V, E)$, we can find weights $\boldsymbol{s} \in \mathbb{R}^m_{\geqslant 0}$ in poly-time with $|\text{supp}(\boldsymbol{s})| \leqslant O(n/\varepsilon^2)$ so that $|\delta(U)| = (1 \pm \varepsilon) \cdot |\boldsymbol{s}(\delta(U))|$ for every $U \subseteq V$.



Graph Sparsification

Theorem (Batson-Spielman-Srivastava '08)

Given a graph $G = (V, E)$, we can find weights $\boldsymbol{s} \in \mathbb{R}^m_{\geqslant 0}$ in poly-time with $|\mathrm{supp}(\boldsymbol{s})| \leqslant O(n/\varepsilon^2)$ so that $|\delta(U)| = (1 \pm \varepsilon) \cdot |\boldsymbol{s}(\delta(U))|$ for every $U \subseteq V$.

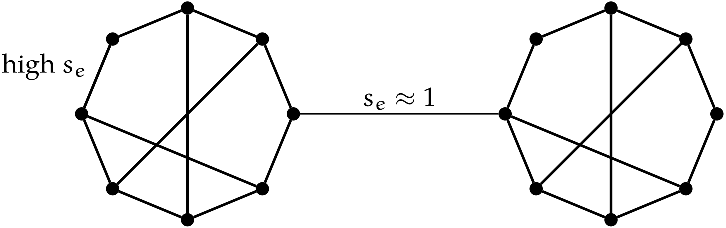


Graph Sparsification

Theorem (Batson-Spielman-Srivastava '08)

Given rank-one $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m \in \mathbb{R}^{n \times n}$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, we can find weights $\boldsymbol{s} \in \mathbb{R}^m_{\geqslant 0}$ in poly-time such that $|\mathrm{supp}(\boldsymbol{s})| \leqslant O(n/\varepsilon^2)$ and

$$\Big\| \sum_{i=1}^m s_i \boldsymbol{A}_i \Big\|_{\mathrm{op}} \in [1-\varepsilon, 1+\varepsilon]$$



A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^{m} \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

(1) Set $[m] \leftarrow \mathrm{supp}(\boldsymbol{s})$, $\boldsymbol{A}'_i \leftarrow s_i \boldsymbol{A}_i$

A new algorithm

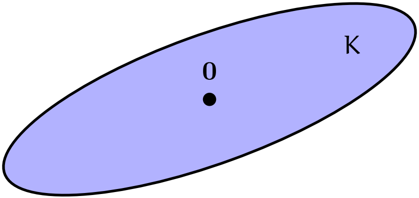
Given: PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^{m} \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

(1) Set $[m] \leftarrow \mathrm{supp}(\boldsymbol{s})$, $\boldsymbol{A}'_i \leftarrow s_i \boldsymbol{A}_i$

(2) Consider convex body $K = \left\{ \boldsymbol{x} \in \mathbb{R}^{[m]} \mid \left\| \sum_{i \in [m]} x_i \boldsymbol{A}'_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$



A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

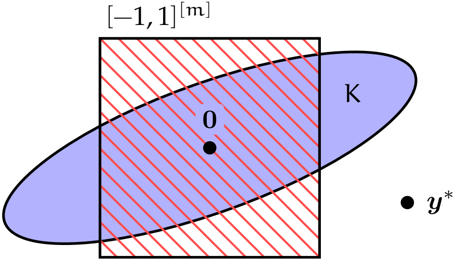
(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

(1) Set $[m] \leftarrow \mathrm{supp}(\boldsymbol{s})$, $\boldsymbol{A}'_i \leftarrow s_i \boldsymbol{A}_i$

(2) Consider convex body $K = \left\{ \boldsymbol{x} \in \mathbb{R}^{[m]} \mid \left\| \sum_{i \in [m]} x_i \boldsymbol{A}'_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$

(3) Sample $\boldsymbol{y}^* \sim N(\boldsymbol{0}, \boldsymbol{I}_{[m]})$



A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

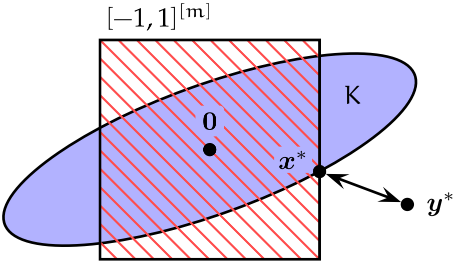
- Repeat until $\boldsymbol{s}$ sufficiently sparse:

(1) Set $[m] \leftarrow \mathrm{supp}(\boldsymbol{s})$, $\boldsymbol{A}_i' \leftarrow s_i \boldsymbol{A}_i$

(2) Consider convex body $K = \left\{ \boldsymbol{x} \in \mathbb{R}^{[m]} \mid \left\| \sum_{i \in [m]} x_i \boldsymbol{A}_i' \right\|_{op} \leqslant \sqrt{n/m} \right\}$

(3) Sample $\boldsymbol{y}^* \sim N(\boldsymbol{0}, \boldsymbol{I}_{[m]})$

(4) Let $\boldsymbol{x}^*$ be the projection of $\boldsymbol{y}^*$ onto $K \cap [-1,1]^{[m]}$



A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \dots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

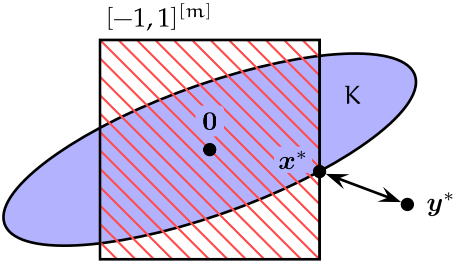
(1) Set $[m] \leftarrow \text{supp}(\boldsymbol{s})$, $\boldsymbol{A}_i' \leftarrow s_i \boldsymbol{A}_i$

(2) Consider convex body $K = \left\{ \boldsymbol{x} \in \mathbb{R}^{[m]} \mid \left\| \sum_{i \in [m]} x_i \boldsymbol{A}_i' \right\|_{\text{op}} \leqslant \sqrt{n/m} \right\}$

(3) Sample $\boldsymbol{y}^* \sim N(\boldsymbol{0}, \boldsymbol{I}_{[m]})$

(4) Let $\boldsymbol{x}^*$ be the projection of $\boldsymbol{y}^*$ onto $K \cap [-1,1]^{[m]}$

- $\boldsymbol{x}^*$ will have $\Omega(m)$ many $\{-1, 1\}$ coordinates with high probability



A new algorithm

Given: PSD matrices $\boldsymbol{A}_1, \dots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$ and $\varepsilon > 0$.

(0) Initialize weights $s_i := 1$ for $i \in [m]$

- Repeat until $\boldsymbol{s}$ sufficiently sparse:

(1) Set $[m] \leftarrow \mathrm{supp}(\boldsymbol{s})$, $\boldsymbol{A}'_i \leftarrow s_i \boldsymbol{A}_i$

(2) Consider convex body $K = \left\{ \boldsymbol{x} \in \mathbb{R}^{[m]} \mid \left\| \sum_{i \in [m]} x_i \boldsymbol{A}'_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$

(3) Sample $\boldsymbol{y}^* \sim N(\boldsymbol{0}, \boldsymbol{I}_{[m]})$

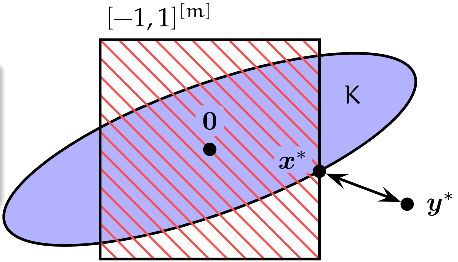
(4) Let $\boldsymbol{x}^*$ be the projection of $\boldsymbol{y}^*$ onto $K \cap [-1,1]^{[m]}$

- $\boldsymbol{x}^*$ will have $\Omega(m)$ many $\{-1,1\}$ coordinates with high probability

(5) Update weights $s_i \leftarrow s_i(1 + x_i^*)$.

(R., Rothvoss '19)

$O(\log m)$ iterations suffice and output $1 \pm O(\varepsilon)$ sparsifier with high probability.



Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{\text{op}} \leqslant \sqrt{n/m} \right\}$$

Question: Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$?

Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

Question: Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$? We don't know!

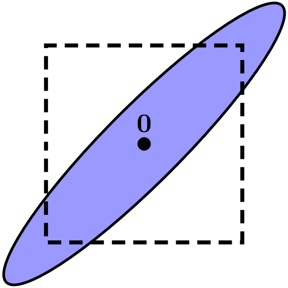
Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

Question: Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$? We don't know!

- Weaker condition to get $\Omega(m)$ coordinates $\boldsymbol{x}_i \in \{-1, 1\}$:



Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

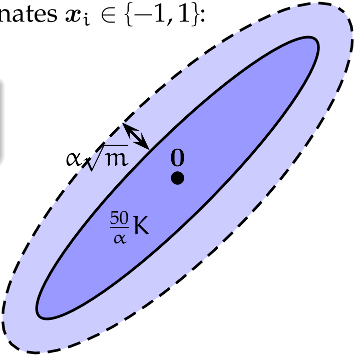
$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

Question: Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$? We don't know!

- Weaker condition to get $\Omega(m)$ coordinates $\boldsymbol{x}_i \in \{-1, 1\}$:

Theorem (R., Rothvoss '19)

$\gamma_m\left(\frac{50}{\alpha}K + \alpha\sqrt{m}B_2^m\right) \geqslant \frac{1}{2}\ \forall \alpha \in (0, 1)$



Main technical results

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

Question: Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$? We don't know!

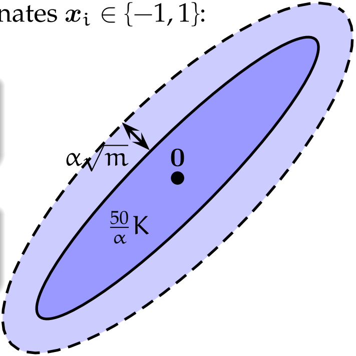
- Weaker condition to get $\Omega(m)$ coordinates $\boldsymbol{x}_i \in \{-1, 1\}$:

Theorem (R., Rothvoss '19)

$\gamma_m(\frac{50}{\alpha}K + \alpha\sqrt{m}B_2^m) \geqslant \frac{1}{2}\ \forall \alpha \in (0, 1)$

Corollary (R., Rothvoss '19)

The **mean width** of K is $\Omega(\sqrt{m})$



Corollary: Spectral Partial Coloring

Corollary (R., Rothvoss '19)

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

We can find a point $\boldsymbol{x}^* \in 500\mathsf{K} \cap [-1,1]^m$ with $\geqslant \frac{m}{9000}$ indices $\boldsymbol{x}^*_i \in \{-1, 1\}$

Proof idea

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

$$B_2^m = \{ \boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\|_2 \leqslant 1 \}$$

Theorem (R., Rothvoss '19)

We have $\gamma_m(\frac{50}{\alpha}K + \alpha\sqrt{m}B_2^m) \geqslant \frac{1}{2}$ for any choice of $\alpha \in (0, 1)$

- That is:

a Gaussian $\sim N(\boldsymbol{0}, \boldsymbol{I}_m)$ is close to $\frac{50}{\alpha}K$ with prob $\geqslant 1/2$

Proof idea

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

$$B_2^m = \{ \boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\|_2 \leqslant 1 \}$$

Theorem (R., Rothvoss '19)

We have $\gamma_m(\frac{50}{\alpha}K + \alpha\sqrt{m}B_2^m) \geqslant \frac{1}{2}$ for any choice of $\alpha \in (0,1)$

- That is:

  a Gaussian $\sim N(\mathbf{0}, \boldsymbol{I}_m)$ is close to $\frac{50}{\alpha}K$ with prob $\geqslant 1/2$
- Equivalently:

  construct random $\boldsymbol{x}$ with $\mathbb{P}[\boldsymbol{x} \in \frac{50}{\alpha}K$ and $\boldsymbol{x}$ close to Gaussian$] \geqslant 1/2$

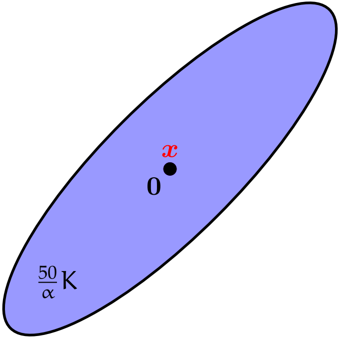
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

(3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\text{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

(4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



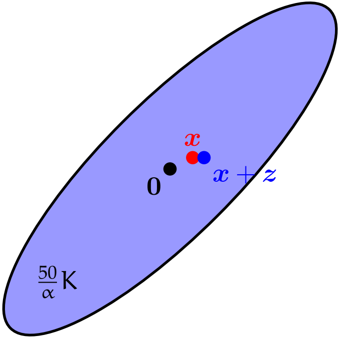
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

    (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

    (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



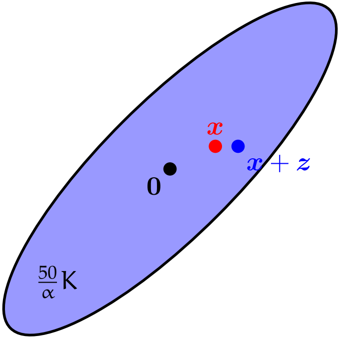
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

   (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

   (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



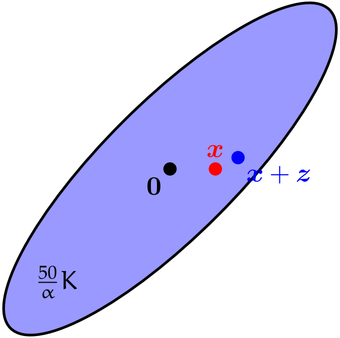
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

 (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

 (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



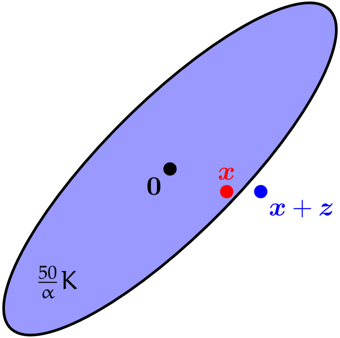
A coupling argument

(1) Set δ := tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

  (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

  (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



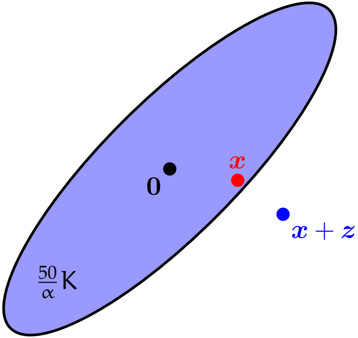
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

  (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1 - \alpha^2)m$

  (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



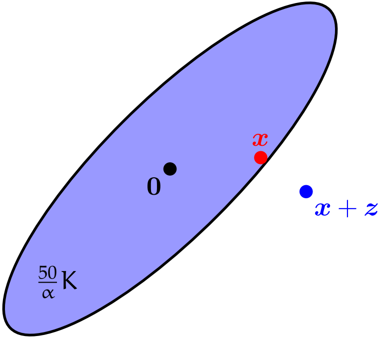
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

  (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

  (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



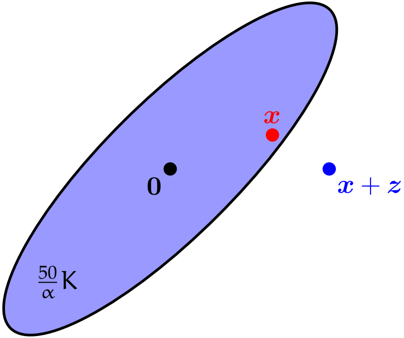
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \boldsymbol{0}, \boldsymbol{z} := \boldsymbol{0}$

(2) FOR $T := \delta^{-2}$ iterations:

  (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

  (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\boldsymbol{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\boldsymbol{0}, \boldsymbol{I}_m - \boldsymbol{X})$



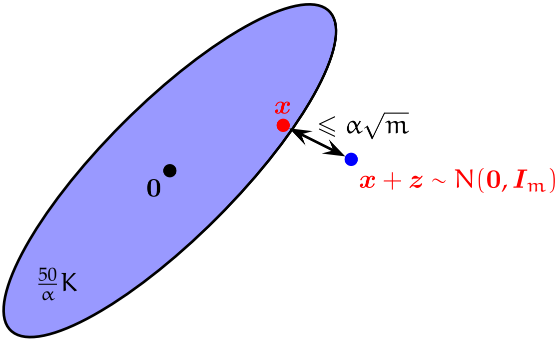
A coupling argument

(1) Set $\delta :=$ tiny step size and $\boldsymbol{x} := \mathbf{0}, \boldsymbol{z} := \mathbf{0}$

(2) FOR $T := \delta^{-2}$ iterations:

  (3) Find "appropriate" PSD $\boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

  (4) Update $\boldsymbol{x} := \boldsymbol{x} + \delta \cdot N(\mathbf{0}, \boldsymbol{X})$ and $\boldsymbol{z} := \boldsymbol{z} + \delta \cdot N(\mathbf{0}, \boldsymbol{I}_m - \boldsymbol{X})$



One update step

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

- Want $\boldsymbol{x} \in K$: can reduce to upper bounding $\lambda_{\max}(\sum_{i=1}^{m} x_i \boldsymbol{A}_i)$
- Use **potential function**

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}(\boldsymbol{x})^{-1}] \quad \text{where} \quad \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

One update step

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{\mathrm{op}} \leqslant \sqrt{n/m} \right\}$$

- Want $\boldsymbol{x} \in \mathsf{K}$: can reduce to upper bounding $\lambda_{\max}(\sum_{i=1}^m x_i \boldsymbol{A}_i)$
- Use **potential function**

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}(\boldsymbol{x})^{-1}] \quad \text{where} \quad \boldsymbol{A}(\boldsymbol{x}) := (\mathsf{C} + \mathsf{D}\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Keep $\Phi(\boldsymbol{x})$ bounded above so that at each step

$$\boldsymbol{A}(\boldsymbol{x}) \succ 0 \implies \lambda_{\max}\Big(\sum_{i=1}^{m} x_i \boldsymbol{A}_i \Big) < \mathsf{C} + \mathsf{D}\|\boldsymbol{x}\|_2^2$$

One update step

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^{m} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

- Want $\boldsymbol{x} \in K$: can reduce to upper bounding $\lambda_{max}(\sum_{i=1}^{m} x_i \boldsymbol{A}_i)$
- Use **potential function**

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}(\boldsymbol{x})^{-1}] \quad \text{where} \quad \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Keep $\Phi(\boldsymbol{x})$ bounded above so that at each step

$$\boldsymbol{A}(\boldsymbol{x}) \succ 0 \implies \lambda_{max}\left(\sum_{i=1}^{m} x_i \boldsymbol{A}_i \right) < C + D\|\boldsymbol{x}\|_2^2$$

Main Lemma

If $\Phi(\boldsymbol{x}) \leqslant \frac{D m^2 \alpha^2}{10}$, there is $\mathbf{0} \prec \boldsymbol{X} \prec \boldsymbol{I}_m$ with $\mathrm{Tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$ so that

$$\mathop{\mathbb{E}}_{\boldsymbol{y} \sim N(\mathbf{0}, \boldsymbol{X})} [\Phi(\boldsymbol{x} + \delta \boldsymbol{y})] \leqslant \Phi(\boldsymbol{x})$$

One update step

- Recall

$$\Phi(\boldsymbol{x}) := \operatorname{tr}[\boldsymbol{A}^{-1}] \text{ where } \boldsymbol{A} := \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

One update step

- Recall

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}^{-1}] \text{ where } \boldsymbol{A} := \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Pick $\boldsymbol{X}$: $\boldsymbol{y} \sim N(\mathbf{0}, \boldsymbol{X})$ is orthogonal to $\boldsymbol{x}$ and to $< \alpha^2 m$ other vectors

One update step

- Recall

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}^{-1}] \text{ where } \boldsymbol{A} := \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Pick $\boldsymbol{X}$: $\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{X})$ is orthogonal to $\boldsymbol{x}$ and to $< \alpha^2 m$ other vectors
- Set $\boldsymbol{B} := \sum_{i=1}^{m} y_i \boldsymbol{A}_i - \delta D\|\boldsymbol{y}\|_2^2 \boldsymbol{I}_n$.
- Change of potential function is

$$\mathrm{tr}[(\boldsymbol{A} - \delta \boldsymbol{B})^{-1}] - \mathrm{tr}[\boldsymbol{A}^{-1}]$$

One update step

- Recall

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}^{-1}] \text{ where } \boldsymbol{A} := \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Pick $\boldsymbol{X}$: $\boldsymbol{y} \sim N(\mathbf{0}, \boldsymbol{X})$ is orthogonal to $\boldsymbol{x}$ and to $< \alpha^2 m$ other vectors
- Set $\boldsymbol{B} := \sum_{i=1}^{m} y_i \boldsymbol{A}_i - \delta D \|\boldsymbol{y}\|_2^2 \boldsymbol{I}_n$.
- Change of potential function is

$$\begin{aligned} & \mathrm{tr}[(\boldsymbol{A} - \delta \boldsymbol{B})^{-1}] - \mathrm{tr}[\boldsymbol{A}^{-1}] \\ \overset{\text{Taylor}}{\leqslant} \quad & \delta \cdot \mathrm{tr}[\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}] + 2\delta^2 \cdot \mathrm{tr}[\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}] \end{aligned}$$

One update step

- Recall

$$\Phi(\boldsymbol{x}) := \mathrm{tr}[\boldsymbol{A}^{-1}] \text{ where } \boldsymbol{A} := \boldsymbol{A}(\boldsymbol{x}) := (C + D\|\boldsymbol{x}\|_2^2) \cdot \boldsymbol{I}_n - \sum_{i=1}^{m} x_i \boldsymbol{A}_i$$

- Pick $\boldsymbol{X}$: $\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{X})$ is orthogonal to $\boldsymbol{x}$ and to $< \alpha^2 m$ other vectors
- Set $\boldsymbol{B} := \sum_{i=1}^{m} y_i \boldsymbol{A}_i - \delta D\|\boldsymbol{y}\|_2^2 \boldsymbol{I}_n$.
- Change of potential function is

$$\begin{aligned} & \mathrm{tr}[(\boldsymbol{A} - \delta\boldsymbol{B})^{-1}] - \mathrm{tr}[\boldsymbol{A}^{-1}] \\ \overset{\text{Taylor}}{\leqslant} \quad & \delta \cdot \mathrm{tr}[\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}] + 2\delta^2 \cdot \mathrm{tr}[\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}] \\ \approx \quad & -\delta^2 D\|\boldsymbol{y}\|_2^2 \mathrm{tr}[\boldsymbol{A}^{-2}] + 2\delta^2 \mathrm{tr}\Big[\boldsymbol{A}^{-1}\Big(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\Big)\boldsymbol{A}^{-1}\Big(\sum_{j=1}^{m} y_j \boldsymbol{A}_j\Big)\boldsymbol{A}^{-1}\Big] \end{aligned}$$

One update step

Key Claim

$$\mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{X})}\left[\mathrm{tr}\left[\boldsymbol{A}^{-1}\left(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\right)\boldsymbol{A}^{-1}\left(\sum_{j=1}^{m} y_j \boldsymbol{A}_j\right)\boldsymbol{A}^{-1}\right]\right] \leqslant \frac{2}{\alpha^2 m}\mathrm{tr}[\boldsymbol{A}^{-1}]\mathrm{tr}[\boldsymbol{A}^{-2}]$$

One update step

Key Claim

$$\mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{X})}\left[\mathrm{tr}\left[\boldsymbol{A}^{-1}\left(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\right)\boldsymbol{A}^{-1}\left(\sum_{j=1}^{m} y_j \boldsymbol{A}_j\right)\boldsymbol{A}^{-1}\right]\right] \leqslant \frac{2}{\alpha^2 m}\mathrm{tr}[\boldsymbol{A}^{-1}]\mathrm{tr}[\boldsymbol{A}^{-2}]$$

- Key idea: Pick $y \perp \boldsymbol{e}_i$ for i such that $\mathrm{tr}[\boldsymbol{A}^{-1}\boldsymbol{A}_i] > \frac{2}{\alpha^2 m}\mathrm{tr}[\boldsymbol{A}^{-1}]$

One update step

Key Claim

$$\mathop{\mathbb{E}}_{\boldsymbol{y}\sim N(\boldsymbol{0},\boldsymbol{X})}\left[\operatorname{tr}\left[\boldsymbol{A}^{-1}\left(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\right)\boldsymbol{A}^{-1}\left(\sum_{j=1}^{m} y_j \boldsymbol{A}_j\right)\boldsymbol{A}^{-1}\right]\right] \leqslant \frac{2}{\alpha^2 m}\operatorname{tr}[\boldsymbol{A}^{-1}]\operatorname{tr}[\boldsymbol{A}^{-2}]$$

- Key idea: Pick $y \perp \boldsymbol{e}_i$ for i such that $\operatorname{tr}[\boldsymbol{A}^{-1}\boldsymbol{A}_i] > \frac{2}{\alpha^2 m}\operatorname{tr}[\boldsymbol{A}^{-1}]$
- By Markov: at most $\frac{\alpha^2 m}{2}$ such constraints

One Update Step

- Putting everything together..

$$
\begin{aligned}
& \mathbb{E}[\mathrm{tr}[(\boldsymbol{A} - \delta \boldsymbol{B})^{-1}]] - \mathrm{tr}[\boldsymbol{A}^{-1}] \\
\leqslant & -\delta^2 D\, \mathbb{E}[\|\boldsymbol{y}\|_2^2]\mathrm{tr}[\boldsymbol{A}^{-2}] + 2\delta^2\, \mathbb{E}\left[\mathrm{tr}\left[\boldsymbol{A}^{-1}\Big(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\Big)\boldsymbol{A}^{-1}\Big(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\Big)\boldsymbol{A}^{-1}\right]\right] \\
\leqslant & \quad \delta^2 \cdot \mathrm{tr}[\boldsymbol{A}^{-2}] \cdot \Big(- 0.4Dm + \frac{4}{\alpha^2 m}\mathrm{tr}[\boldsymbol{A}^{-1}]\Big) \leqslant 0
\end{aligned}
$$

One Update Step

- Putting everything together..

$$
\begin{aligned}
&\mathbb{E}[\mathrm{tr}[(\boldsymbol{A} - \delta \boldsymbol{B})^{-1}]] - \mathrm{tr}[\boldsymbol{A}^{-1}] \\
\leqslant &-\delta^2 D\, \mathbb{E}[\|\boldsymbol{y}\|_2^2]\mathrm{tr}[\boldsymbol{A}^{-2}] + 2\delta^2\, \mathbb{E}\left[\mathrm{tr}\left[\boldsymbol{A}^{-1}\Big(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\Big)\boldsymbol{A}^{-1}\Big(\sum_{i=1}^{m} y_i \boldsymbol{A}_i\Big)\boldsymbol{A}^{-1}\right]\right] \\
\leqslant &\quad \delta^2 \cdot \mathrm{tr}[\boldsymbol{A}^{-2}] \cdot \Big(-0.4Dm + \frac{4}{\alpha^2 m}\mathrm{tr}[\boldsymbol{A}^{-1}]\Big) \leqslant 0
\end{aligned}
$$

- Only $\leqslant \alpha^2 m$ constraints needed: $\mathrm{tr}[\boldsymbol{X}] \geqslant (1-\alpha^2)m$

Conclusion

Given PSD matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m$ with $\sum_{i=1}^m \boldsymbol{A}_i = \boldsymbol{I}_n$, define the set

$$K = \left\{ \boldsymbol{x} \in \mathbb{R}^m \mid \left\| \sum_{i=1}^m x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n/m} \right\}$$

Theorem (R., Rothvoss '19)

We have $\gamma_m(\frac{50}{\alpha}K + \alpha\sqrt{m}B_2^m) \geqslant \frac{1}{2}\ \forall \alpha \in (0,1)$

Corollary (R., Rothvoss '19)

The **mean width** of K is $\Omega(\sqrt{m})$

- Do we have $\gamma_m(K) \geqslant 2^{-O(m)}$? Does the Theorem already imply this?

More Open Problems

- Can we show similar measure lower bounds for other problems?
  - Matrix Spencer: given $\boldsymbol{A}_i \in \mathbb{R}^{n \times n}$ with $\|\boldsymbol{A}_i\|_{op} \leqslant 1$

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left\| \sum_{i=1}^{n} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n} \right\}$$

    known for $O(\sqrt{n \log n})$

  - Matoušek's: given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with all |subdeterminants| $\leqslant 1$

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{A}\boldsymbol{x}\|_\infty \leqslant \log n \right\}$$

    known for $O(\sqrt{\log m} \cdot \log n)$

More Open Problems

- Can we show similar measure lower bounds for other problems?
  - Matrix Spencer: given $\boldsymbol{A}_i \in \mathbb{R}^{n \times n}$ with $\|\boldsymbol{A}_i\|_{op} \leqslant 1$

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \left\| \sum_{i=1}^{n} x_i \boldsymbol{A}_i \right\|_{op} \leqslant \sqrt{n} \right\}$$

    known for $O(\sqrt{n \log n})$

  - Matoušek's: given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with all $|\text{subdeterminants}| \leqslant 1$

$$\mathsf{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{A}\boldsymbol{x}\|_\infty \leqslant \log n \right\}$$

    known for $O(\sqrt{\log m} \cdot \log n)$

Thanks for your attention!