Spectral Sparsification and Geometric Applications

Nikhil Srivastava

MSRI

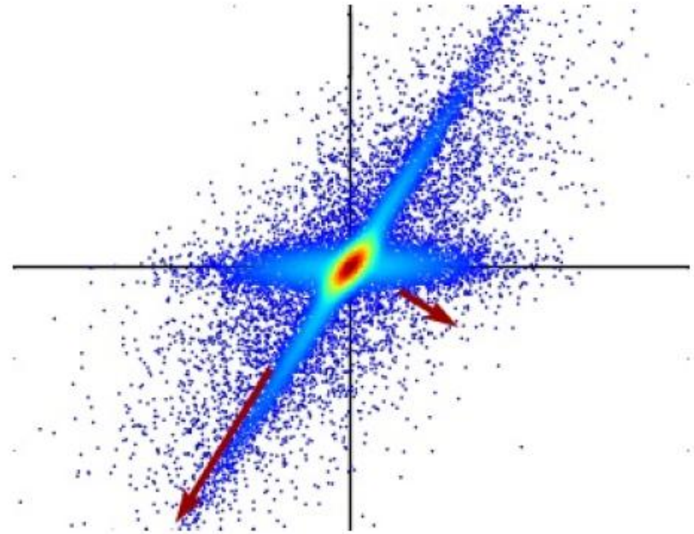
Part III: Covariance Estimation

Covariance Matrices

Random Vector $X \in \mathbf{R}^n$

Covariance Matrix

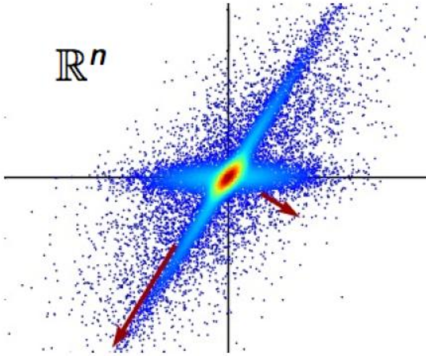
$$\Sigma = \mathbf{E} X X^T$$



Variances $u^T \Sigma u = \mathbf{E}\langle u, X\rangle^2 \quad u \in S^{n-1}$

PCA Find max variance directions = eigenvectors

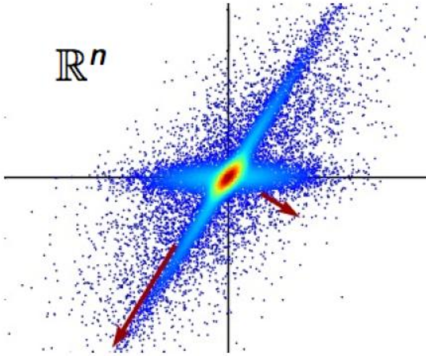
Covariance Estimation



Goal: Estimate $\sum$ given i.i.d. $X_1, \ldots, X_q$

Want: $\left\| \frac{1}{q} \sum_{i \leq q} X_i X_i^T - \Sigma \right\|_2 \leq \epsilon \|\Sigma\|$

Covariance Estimation



Goal: Estimate Σ given i.i.d. $X_1, \ldots, X_q$

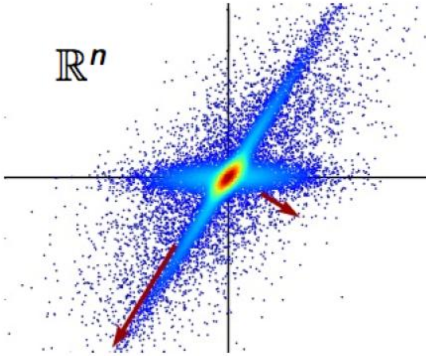
Want: $\left\| \frac{1}{q} \sum_{i \le q} X_i X_i^T - \Sigma \right\|_2 \le \epsilon \|\Sigma\|$

Sufficient to handle $\Sigma = I$ **isotropic position.**

$$\mathbf{E}\langle u, X\rangle^2 = 1 \quad \forall u \qquad \mathbf{E}\|X\|^2 = n$$

$$X_i' = \Sigma^{-1/2} X_i$$

Covariance Estimation



Goal: Estimate $\sum$ given i.i.d. $X_1, \ldots, X_q$

Want: $\left\| \frac{1}{q} \sum_{i \le q} X_i X_i^T - \Sigma \right\|_2 \le \epsilon \|\Sigma\|$

Sufficient to handle $\sum = I$ **isotropic position.**

$$\mathbf{E}\langle u, X \rangle^2 = 1 \quad \forall u \qquad\qquad \mathbf{E}\|X\|^2 = n$$

Given $\mathbf{E} X X^T = I$ how large is **q**?

[KLS,B,...Rudelson'99]

Isotropic random $X \in \mathbf{R}^n$ with $\|X\|_2 \leq O(\sqrt{n})$

If $q = \Omega(n \log n / \epsilon^2)$

Then $\left\| \frac{1}{q} \sum_{i \leq q} X_i X_i^T - I \right\|_2 \leq \epsilon$ whp.

[Rudelson'99]

Isotropic random $X \in \mathbf{R}^n$ with $\|X\|_2 \leq O(\sqrt{n})$

If $q = \Omega(n \log n / \epsilon^2)$

Then $\left\| \frac{1}{q} \sum_{i \leq q} X_i X_i^T - I \right\|_2 \leq \epsilon$ whp.

Cf' [SS08] Effective Resistance Sampling

[Rudelson'99]

Isotropic random $X \in \mathbf{R}^n$ with $\|X\|_2 \leq O(\sqrt{n})$

If $q = \Omega(n \log n/\epsilon^2)$

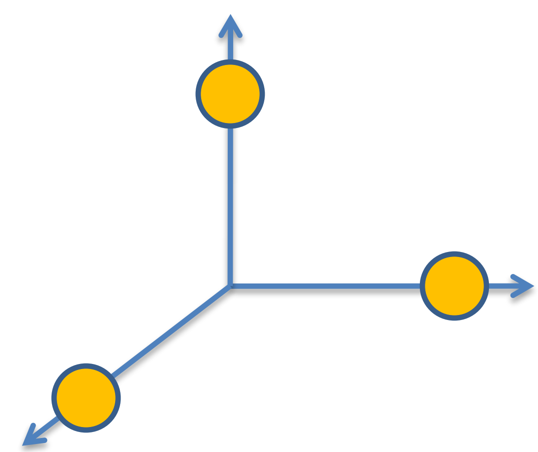
Then $\left\| \frac{1}{q} \sum_{i \leq q} X_i X_i^T - I \right\|_2 \leq \epsilon$ whp.

Tight example:

$X = \sqrt{n} e_i \qquad \text{w. prob. } 1/n$

$\mathbf{E} X X^T = (1/n) \sum_{i \leq n} n e_i e_i^T = I$

$\Sigma_q(i,i) = \text{num. of balls in bin } i$



[Rudelson'99]

Isotropic random $X \in \mathbf{R}^n$ with $\|X\|_2 \leq O(\sqrt{n})$

If $q = \Omega(n \log n / \epsilon^2)$

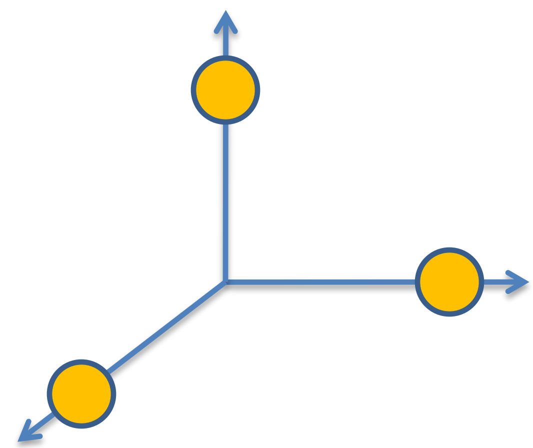
Then $\left\| \frac{1}{q} \sum_{i \leq q} X_i X_i^T - I \right\|_2 \leq \epsilon$ whp.

Tight example:

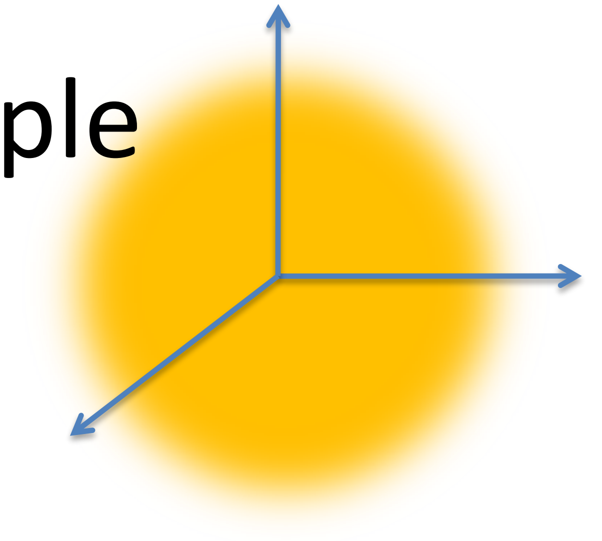
$X = \sqrt{n} e_i \quad$ w. prob. $1/n$

$\mathbf{E} X X^T = (1/n) \sum_{i \leq n} n e_i e_i^T = I$

$\Sigma_q(i,i) =$ num. of balls in bin i



A Good Example



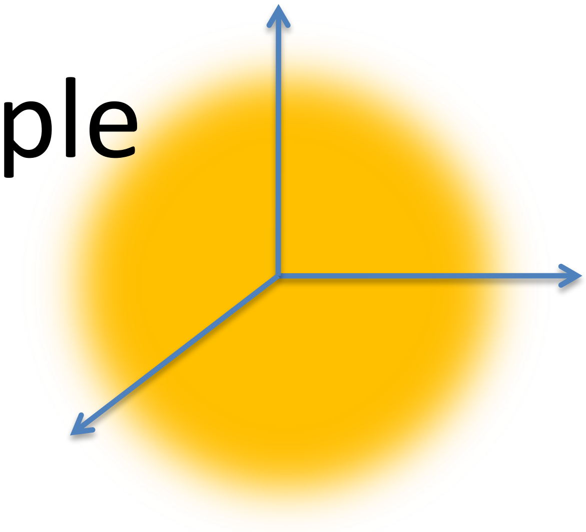
Standard Gaussian vector:

$$X \sim \mathcal{N}(0, I)$$

For any fixed direction

$$u \in S^{n-1} \qquad \langle u, X \rangle \sim \mathcal{N}(0, 1)$$

A Good Example



Standard Gaussian vector:

$$X \sim \mathcal{N}(0, I)$$

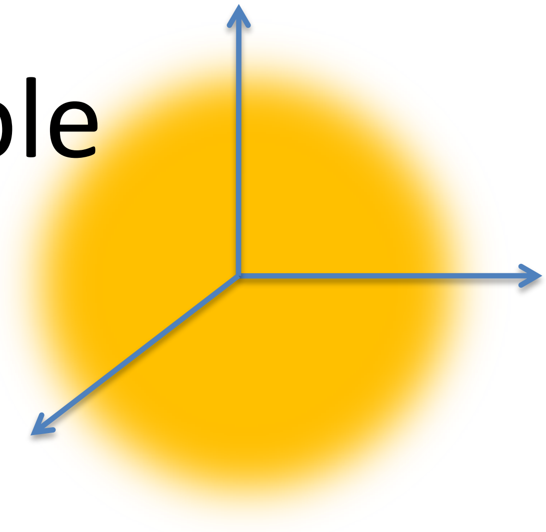
For any fixed direction

$$u \in S^{n-1} \qquad \langle u, X \rangle \sim \mathcal{N}(0, 1)$$

So for independent $\mathbf{X_1, \ldots X_q}$

$$u^T \Sigma_q u = \frac{1}{q} \sum_{i \leq q} \langle u, X_i \rangle^2 \sim \chi^2(q)$$

A Good Example



Standard Gaussian vector:

$$X \sim \mathcal{N}(0, I)$$

For any fixed direction

$$u \in S^{n-1} \qquad \langle u, X \rangle \sim \mathcal{N}(0, 1)$$

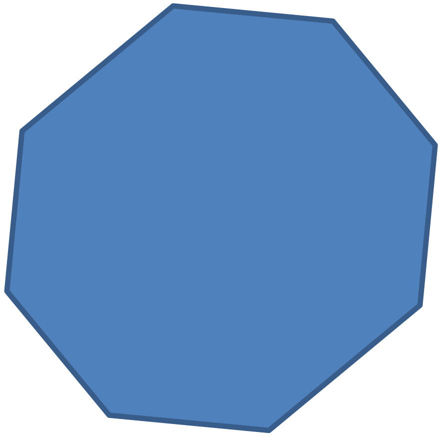
So for independent $\mathbf{X_1, \dots X_q}$

$$u^T \Sigma_q u = \frac{1}{q} \sum_{i \le q} \langle u, X_i \rangle^2 \sim \chi^2(q)$$

$$\mathbf{P}(|\frac{1}{q} \sum_{i \le q} \langle u, X_i \rangle^2 - 1| > \epsilon) \le \exp(-q\epsilon^2)$$

$$\text{Take } q \gg n/\epsilon^2, \text{ union bound.}$$

Convex Bodies [KLS'95]

$X \sim$ 

More generally **sub-exponential X**:

$$\forall u \in S^{n-1} : \quad \mathbf{P}(|\langle u, X\rangle| > t) \leq C \exp(-t)$$

[ALPT'09] **N=O(n)** whp, provided $\|X\|_2 \leq O(\sqrt{n})$

Heavier Tails

[S-Vershynin'11] Suppose isotropic **X** satisfies:

1D $\forall u \in S^{n-1} : \quad \mathbf{P}(|\langle u, X \rangle| > t) \leq C/t^{2+\eta}$

Heavier Tails

[S-Vershynin'11] Suppose isotropic **X** satisfies:

1D $\forall u \in S^{n-1} : \quad \mathbf{P}(|\langle u, X\rangle| > t) \leq C/t^{2+\eta}$

kD $\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\text{rank}(\Pi)}$

cf. $\mathbf{E}\|\Pi X\|^2 = \text{rank}(\Pi)$

Heavier Tails

[S-Vershynin'11] Suppose isotropic **X** satisfies:

1D $\forall u \in S^{n-1} : \quad \mathbf{P}(|\langle u, X\rangle| > t) \leq C/t^{2+\eta}$

kD $\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\mathrm{rank}(\Pi)}$

Then $\mathbf{E}\left\|\frac{1}{q}\sum_{i\leq q} X_i X_i^T - I\right\|_2 \leq \epsilon$ for $q = \Omega_{\epsilon,\eta}(n)$

Heavier Tails

[S-Vershynin'11] Suppose isotropic **X** satisfies:

1D $\forall u \in S^{n-1}: \quad \mathbf{P}(|\langle u, X\rangle| > t) \leq C/t^{2+\eta}$

kD $\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\mathrm{rank}(\Pi)}$

Then $\mathbf{E}\left\|\frac{1}{q}\sum_{i\leq q} X_i X_i^T - I\right\|_2 \leq \epsilon$ for $q = \Omega_{\epsilon,\eta}(n)$

Includes: log-concave **X** by [Paouris '07]

product **X** with bdd $4+\eta$ moments

cf. [Latala'05]

Heavier Tails

[S-Vershynin'11] Suppose isotropic **X** satisfies:

1D $\forall u \in S^{n-1}: \quad \mathbf{P}(|\langle u, X\rangle| > t) \leq C/t^{2+\eta}$

kD $\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\mathrm{rank}(\Pi)}$

Then $\mathbf{E}\left\|\frac{1}{q}\sum_{i\leq q} X_i X_i^T - I\right\|_2 \leq \epsilon$ for $q = \Omega_{\epsilon,\eta}(n)$

Lower edge is easier: Only require 1D for

$$\mathbf{E}\lambda_{min}(\Sigma_q) \geq 1 - \epsilon.$$

Sketch of the proof

Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$A_0 = 0$$

Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$A_1 = X_1 X_1^T$$

Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$A_2 = X_1 X_1^T + X_2 X_2^T$$

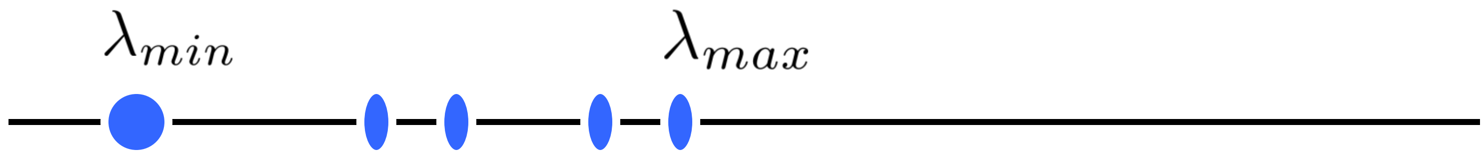
Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$A_3 = X_1 X_1^T + X_2 X_2^T + X_3 X_3^T$$

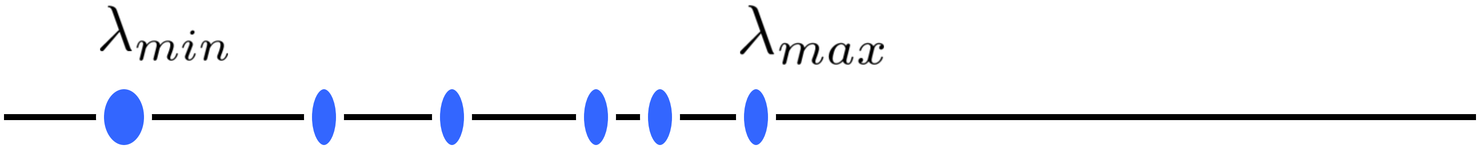
Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



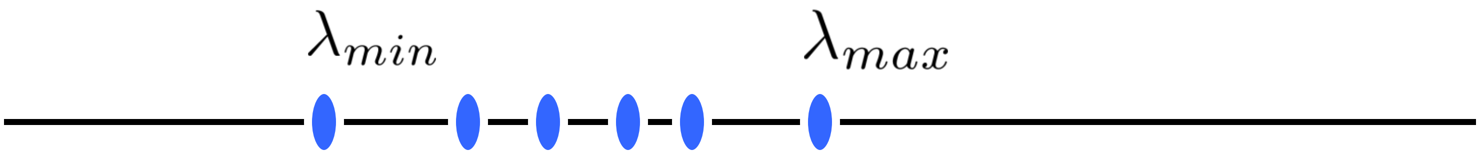
Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



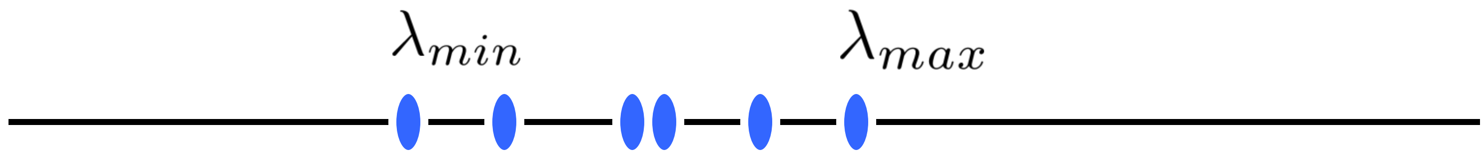
Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



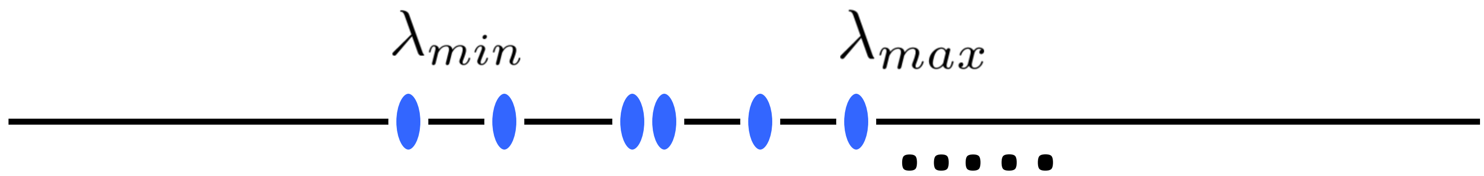
Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



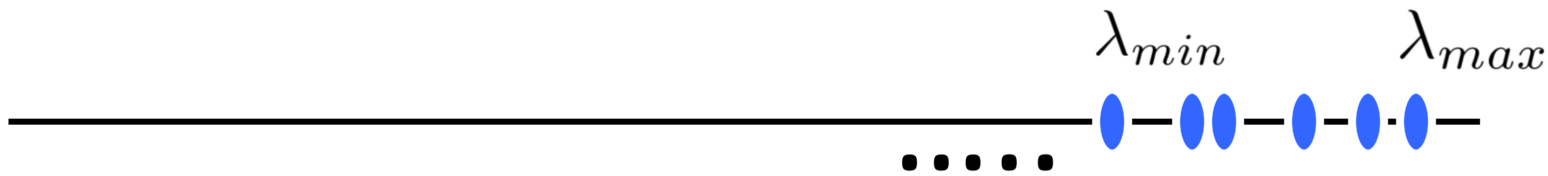
Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$A_q = X_1 X_1^T + X_2 X_2^T \ldots X_q X_q^T$$

Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$



$$(1-\epsilon)q \leq \mathbf{E}\lambda \leq (1+\epsilon)q$$

$$A_q = X_1 X_1^T + X_2 X_2^T \dots X_q X_q^T$$

Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$(1-\epsilon)q \le \mathbf{E}\lambda \le (1+\epsilon)q$$

$$A_q = X_1 X_1^T + X_2 X_2^T \ldots X_q X_q^T$$

$$\begin{aligned} \mathbf{E}\mathrm{Tr}(A_k) &= \mathrm{Tr}(A_{k-1}) + \mathrm{Tr}(\mathbf{E} X_k X_k^T) \\ &= \mathrm{Tr}(A_{k-1}) + n \end{aligned}$$

Basic Picture

$$A_k = \sum_{i \leq k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$(1-\epsilon)q \leq \mathbf{E}\lambda \leq (1+\epsilon)q$$

$$A_q = X_1 X_1^T + X_2 X_2^T \ldots X_q X_q^T$$

$$\mathbf{E}\lambda_{max}(A_k) \leq \lambda_{max}(A_{k-1}) + O(1)?$$

$$\mathbf{E}\mathrm{Tr}(A_k) =$$

$$= \mathrm{Tr}(A_{k-1}) + n$$

Basic Picture

$$A_k = \sum_{i \le k} X_i X_i^T = A_{k-1} + X_k X_k^T$$

$$(1-\epsilon)q \le \mathbf{E}\lambda \le (1+\epsilon)q$$

$$A_q = X_1 X_1^T + X_2 X_2^T \ldots X_q X_q^T$$

$$\mathbf{E}\mathrm{Tr}(A_k) =$$

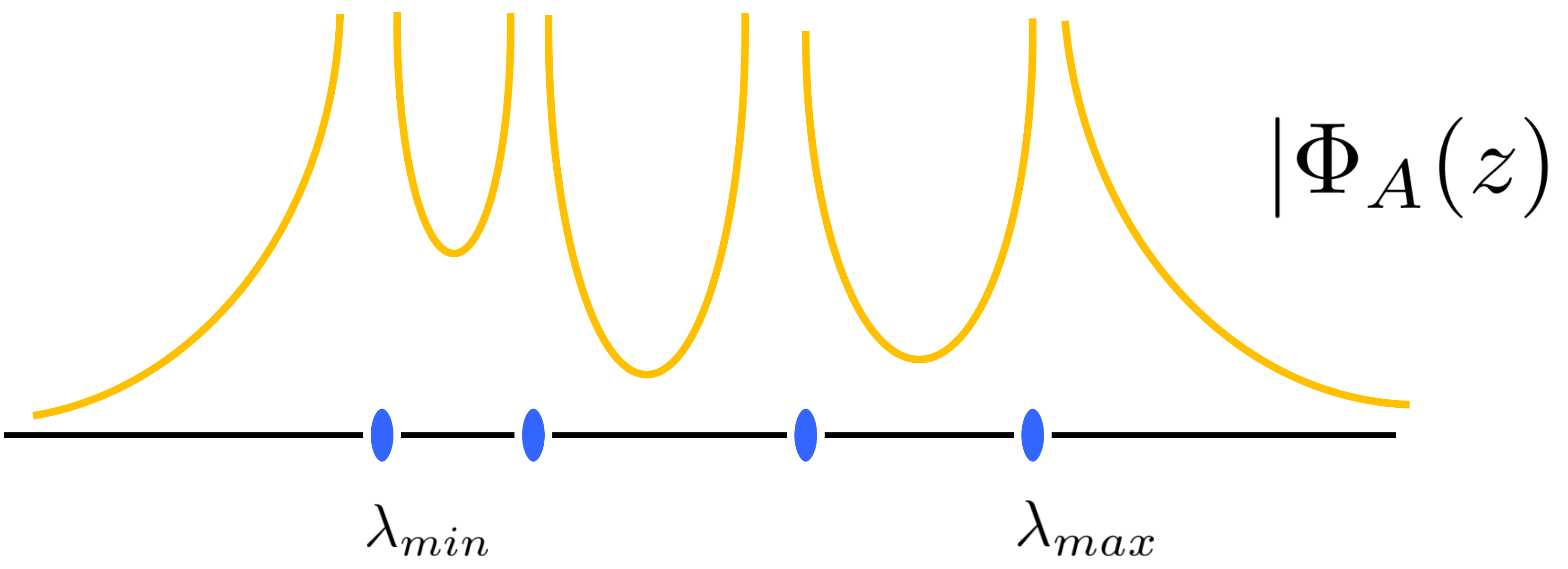
$$= \mathrm{Tr}(A_{k-1}) + n$$

$$\mathbf{E}\lambda_{max}(A_k) \le \lambda_{max}(A_{k-1}) + O(1)?$$

False

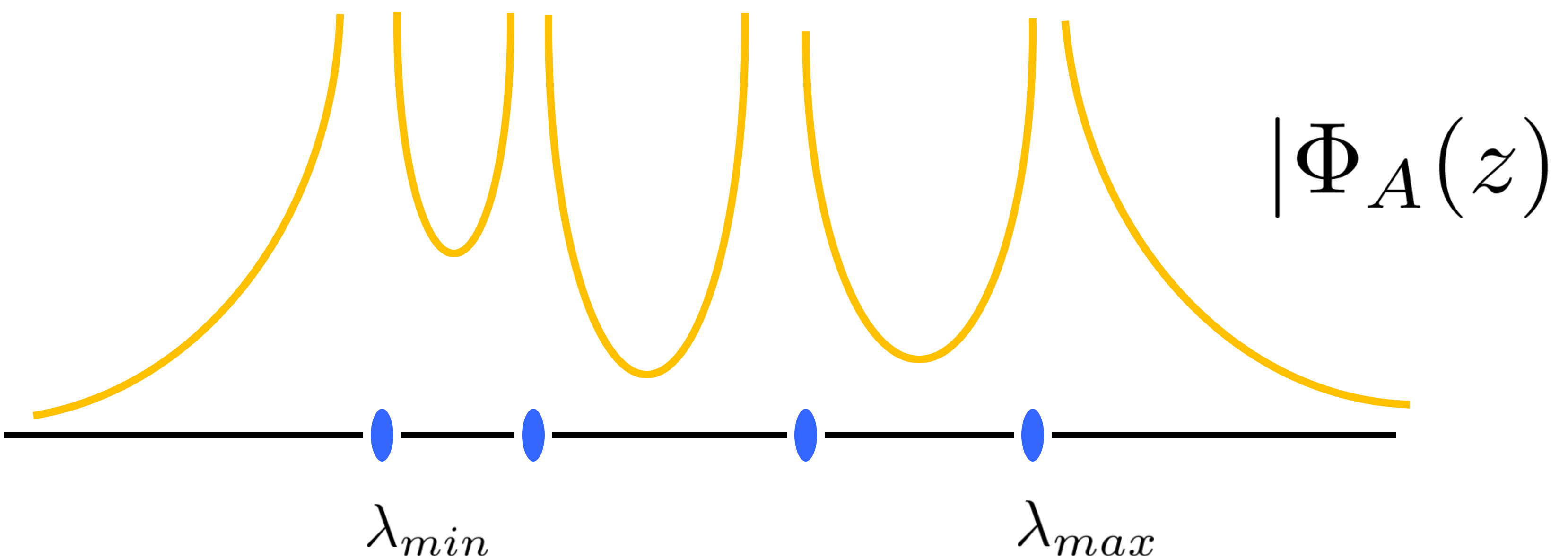
Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$



Softening the Edges

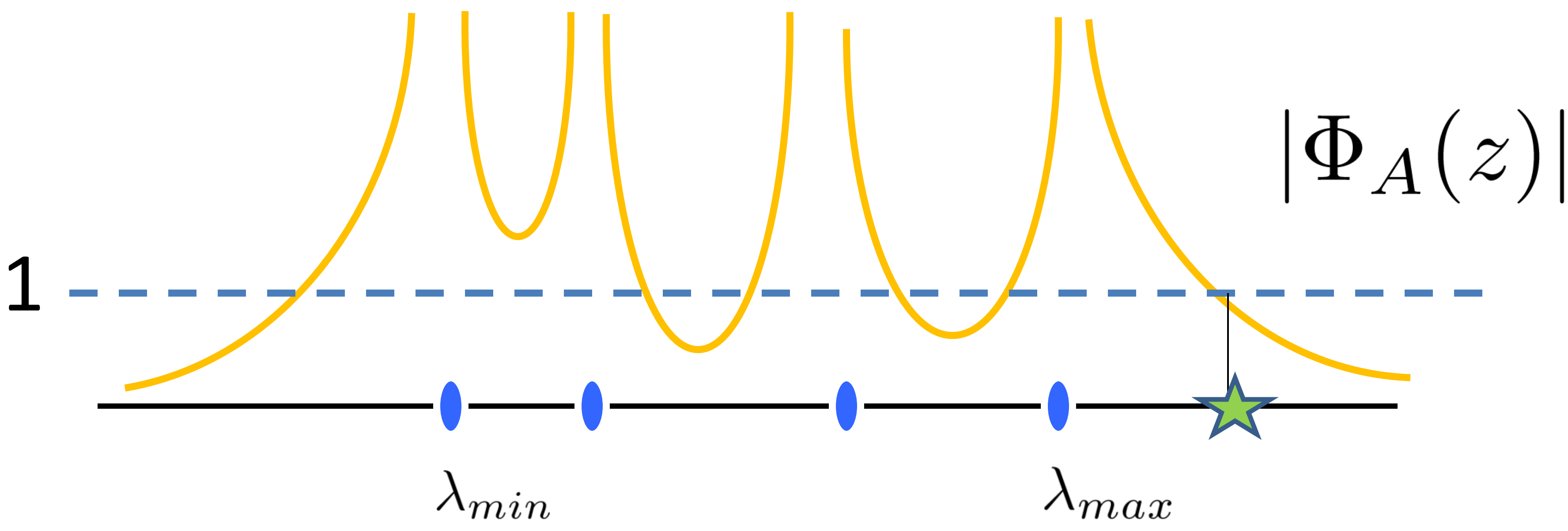
$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$



$$\lambda_{max} = \max\{z : \Phi_A(z) = \infty\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

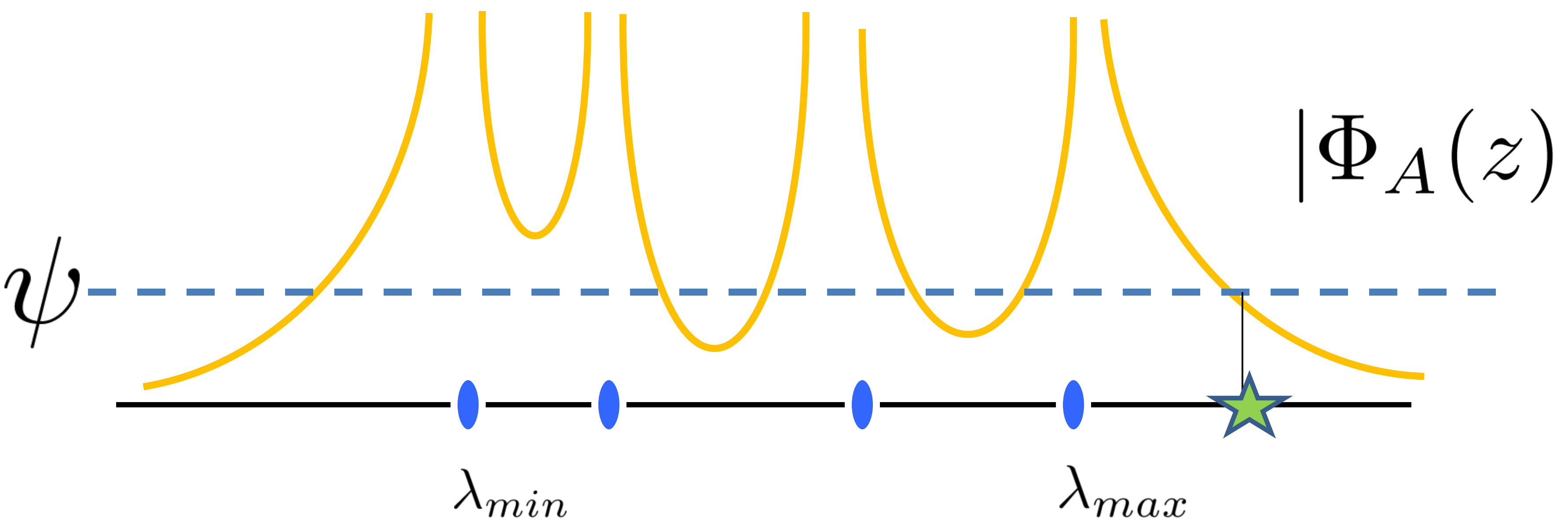


$$\lambda_{max} = \max\{z : \Phi_A(z) = \infty\}$$

$$s_{max} := \max\{z : \Phi_A(z) = 1\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

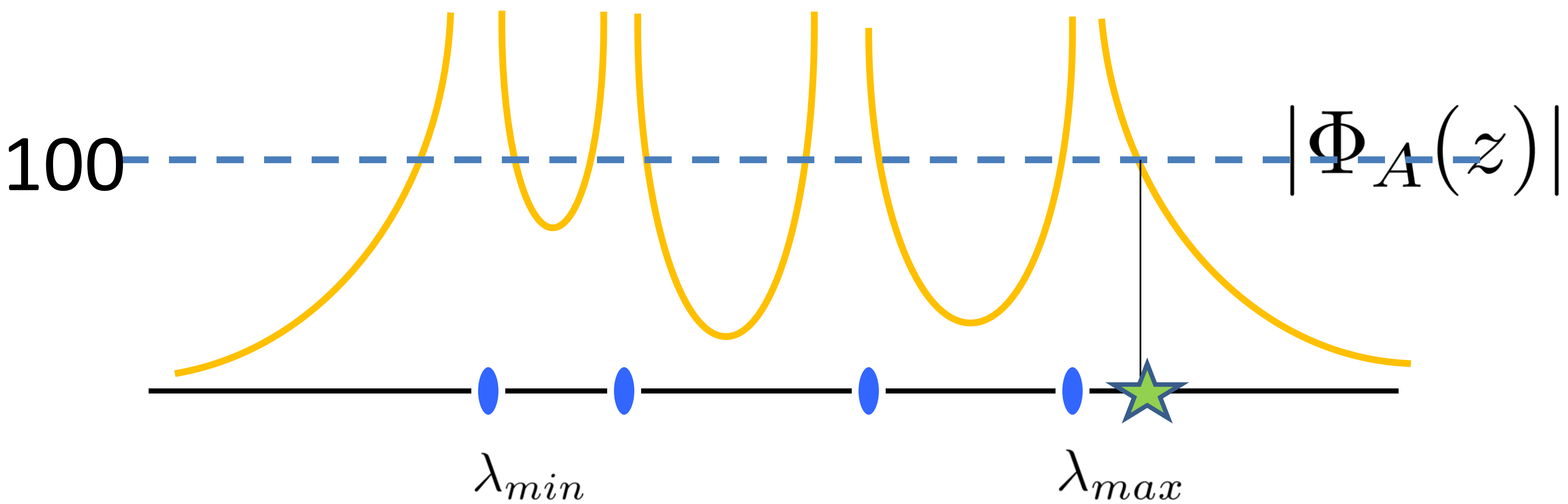


$$\lambda_{max} = \max\{z : \Phi_A(z) = \infty\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

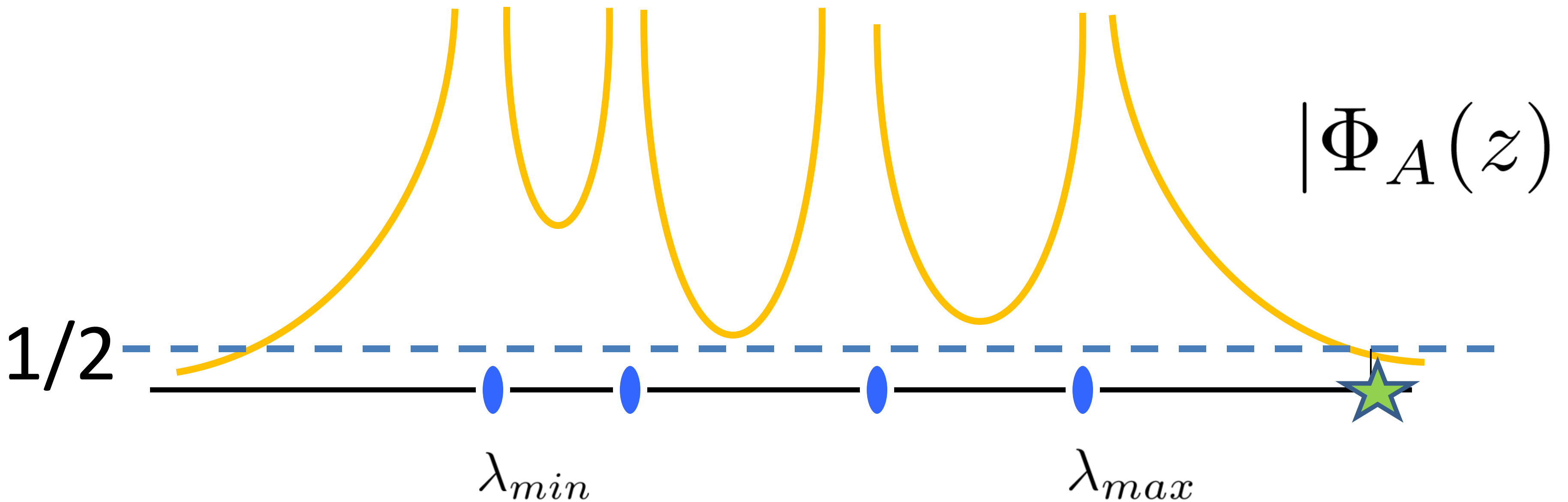


$$\lambda_{max} = \max\{z : \Phi_A(z) = \infty\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

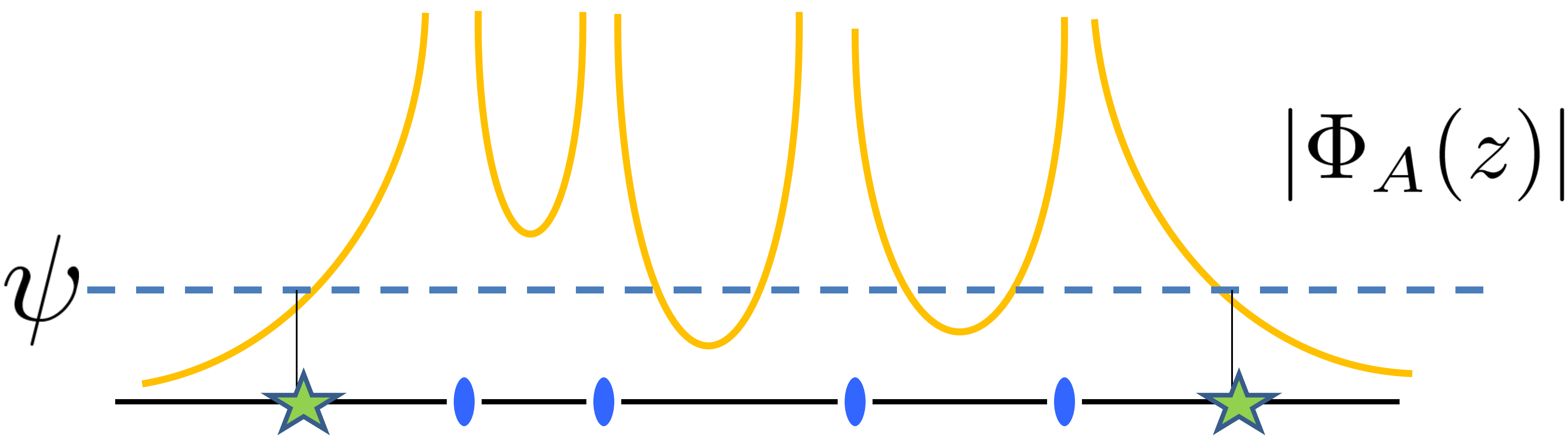


$$\lambda_{max} = \max\{z : \Phi_A(z) = \infty\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

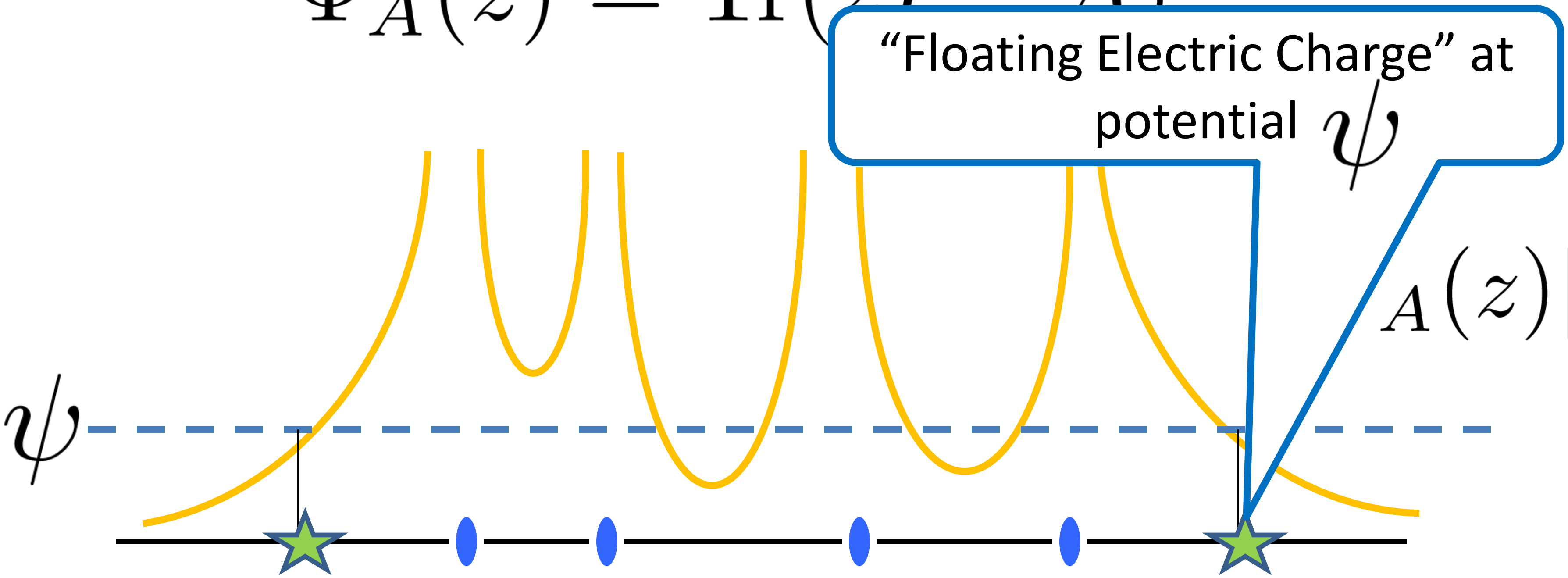


$$s_{min} := \min\{z : |\Phi_A(z)| = \psi\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

Softening the Edges

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$



$$s_{min} := \min\{z : |\Phi_A(z)| = \psi\}$$

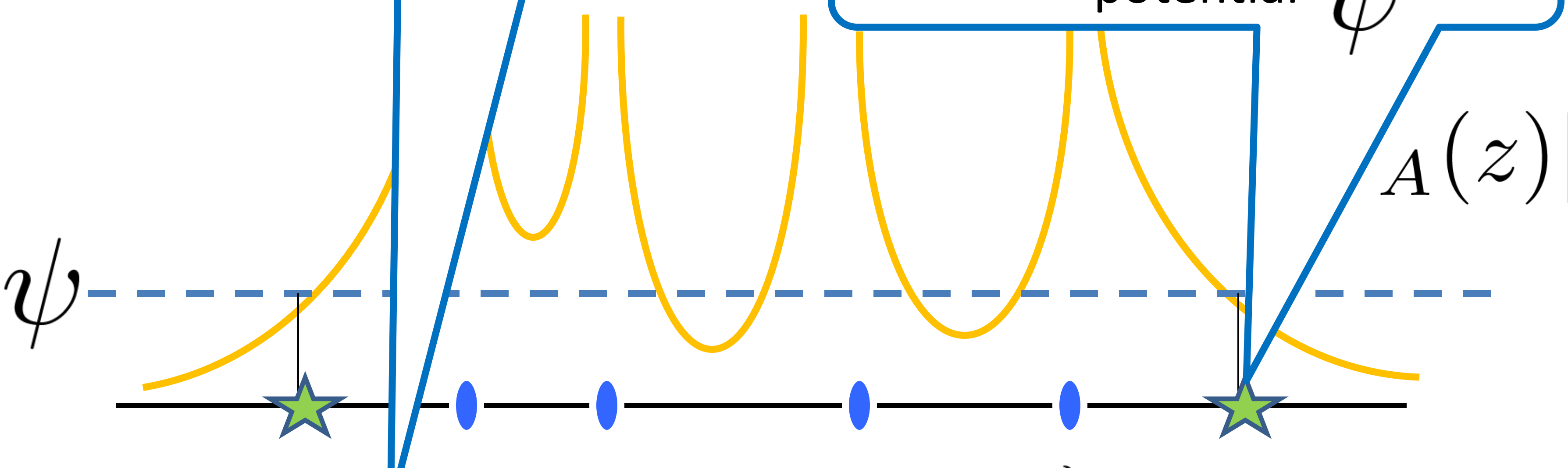
$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

…ng the Edges

Well-behaved + explicitly computable

$$\Phi_A(z) = \mathrm{Tr}(zI - A)^{-1}$$

“Floating Electric Charge” at potential ψ



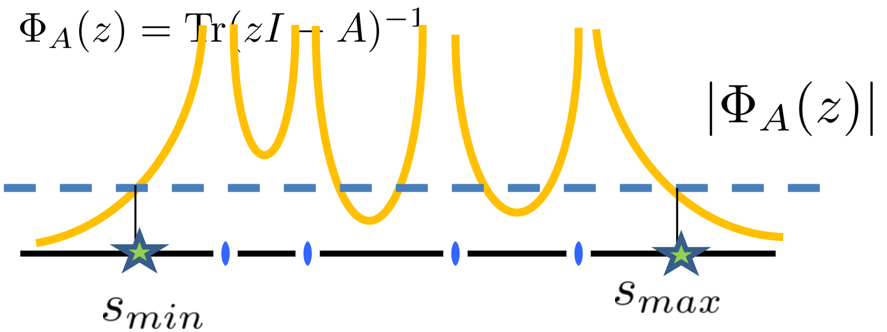
$$s_{min} := \min\{z : |\Phi_A(z)| = \psi\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$

Soft Spectral Edges

$$s_{min} := \min\{z : |\Phi_A(z)| = \psi\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$



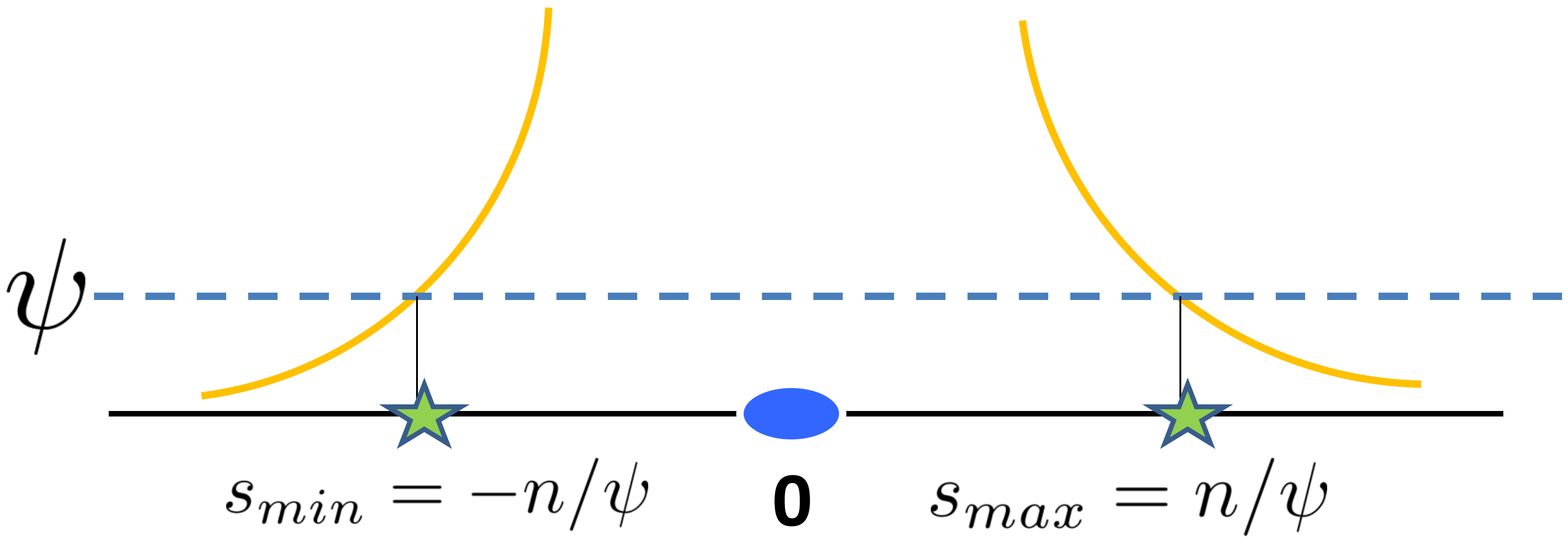
Main Lemma.

$A \succeq 0$, X **regular** isotropic random vector.
For all $\epsilon > 0$ there is $\psi = \psi(\epsilon)$ with

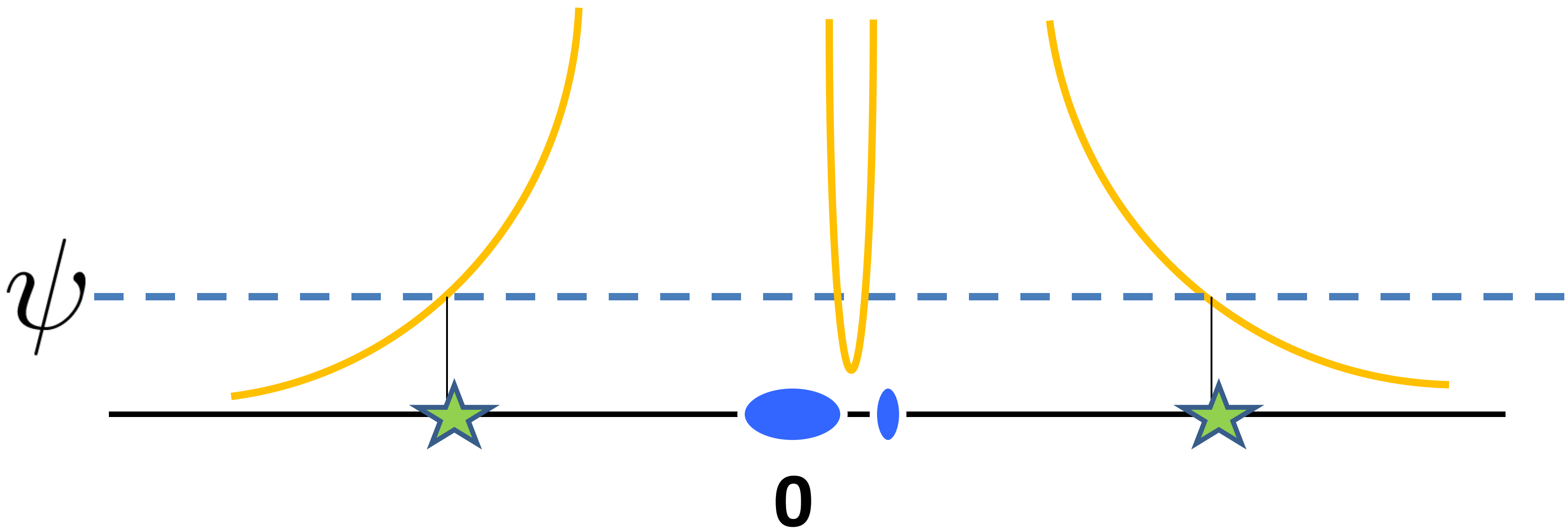
$$\mathbf{E} s_{max}(A + XX^T) \leq s_{max}(A) + 1 + \epsilon$$

$$\mathbf{E} s_{min}(A + XX^T) \geq s_{min}(A) + 1 - \epsilon.$$

The Sensitivity Tradeoff

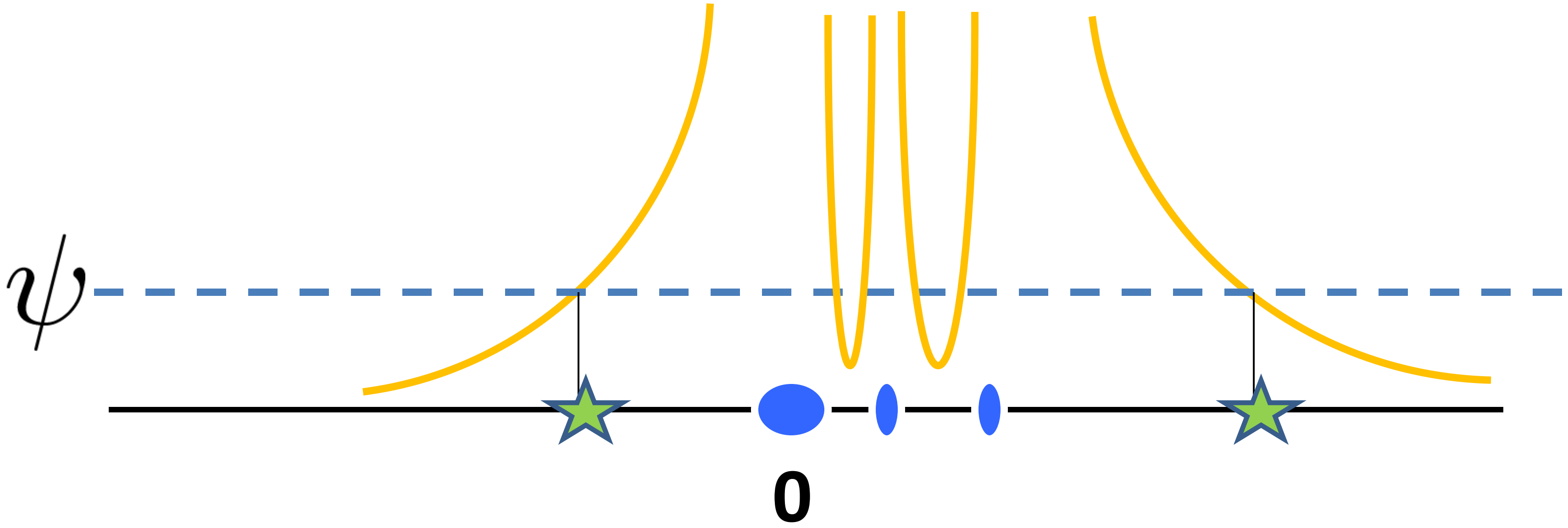


The Sensitivity Tradeoff



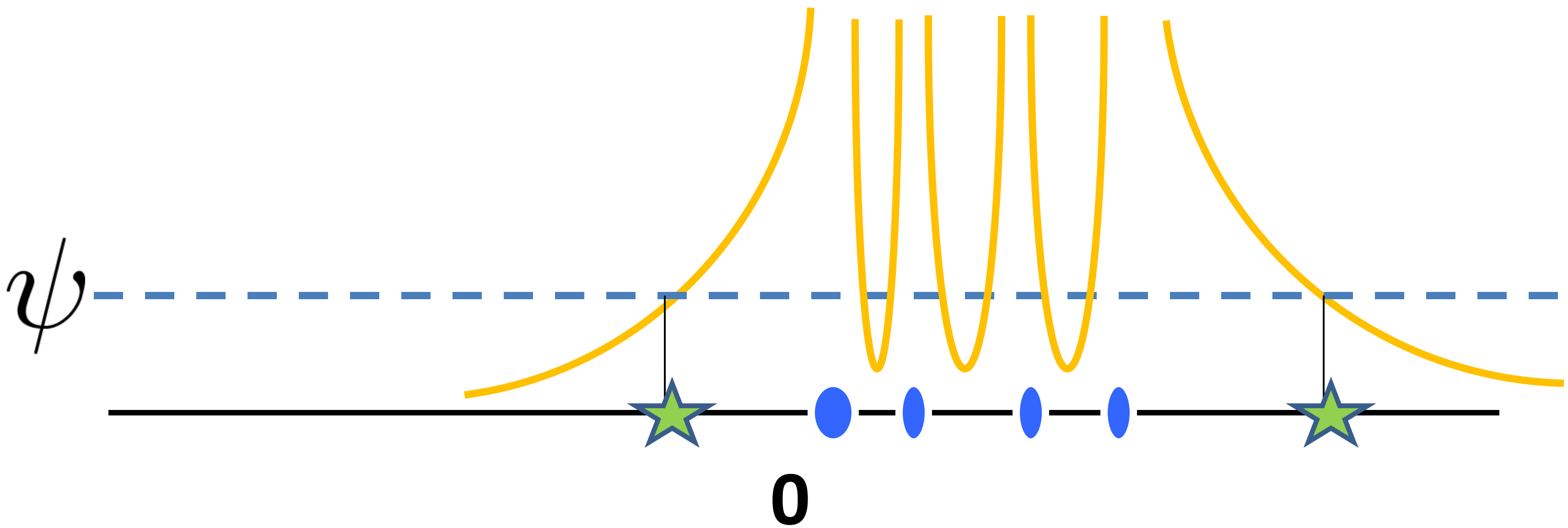
$$A_k = A_{k-1} + X_k X_k^T$$

The Sensitivity Tradeoff



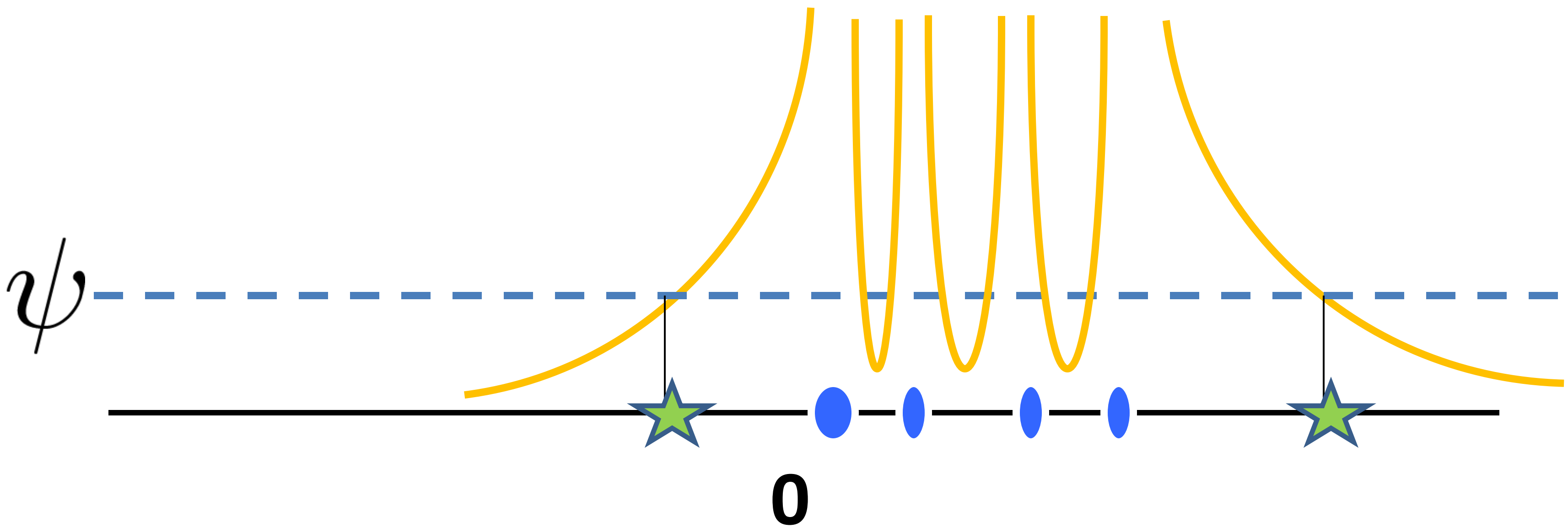
$$A_k = A_{k-1} + X_k X_k^T$$

The Sensitivity Tradeoff



$$A_k = A_{k-1} + X_k X_k^T$$

The Sensitivity Tradeoff

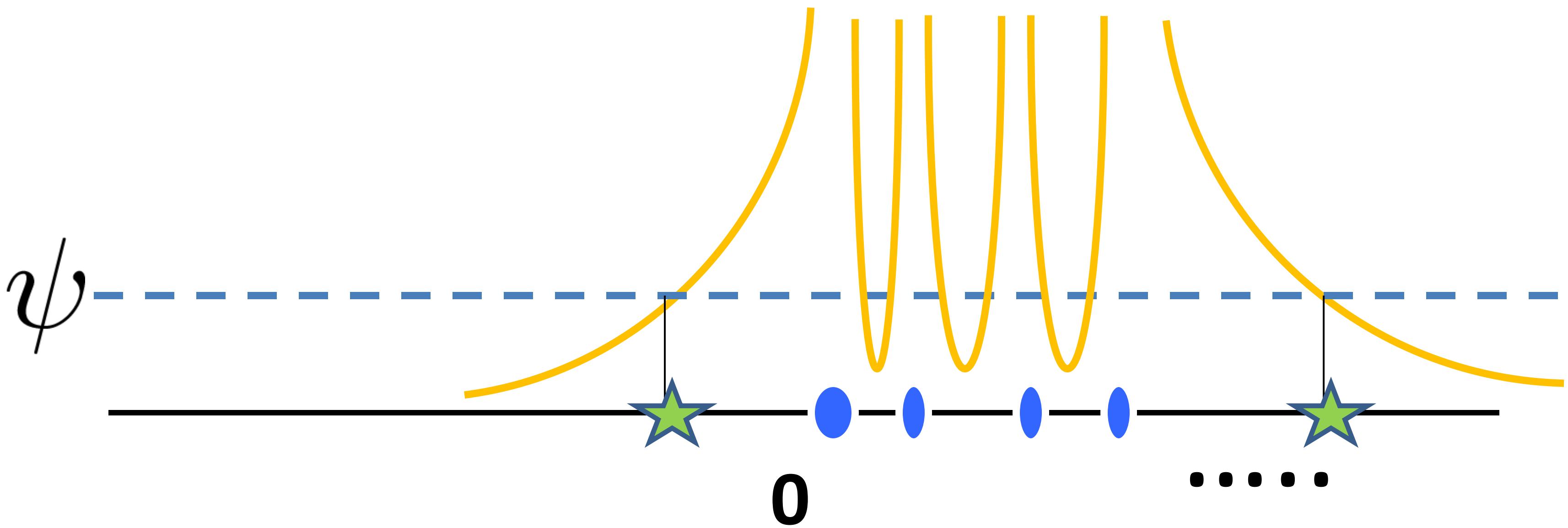


$$A_k = A_{k-1} + X_k X_k^T$$

Lemma.

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + k(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + k(1+\epsilon).$$

The Sensitivity Tradeoff

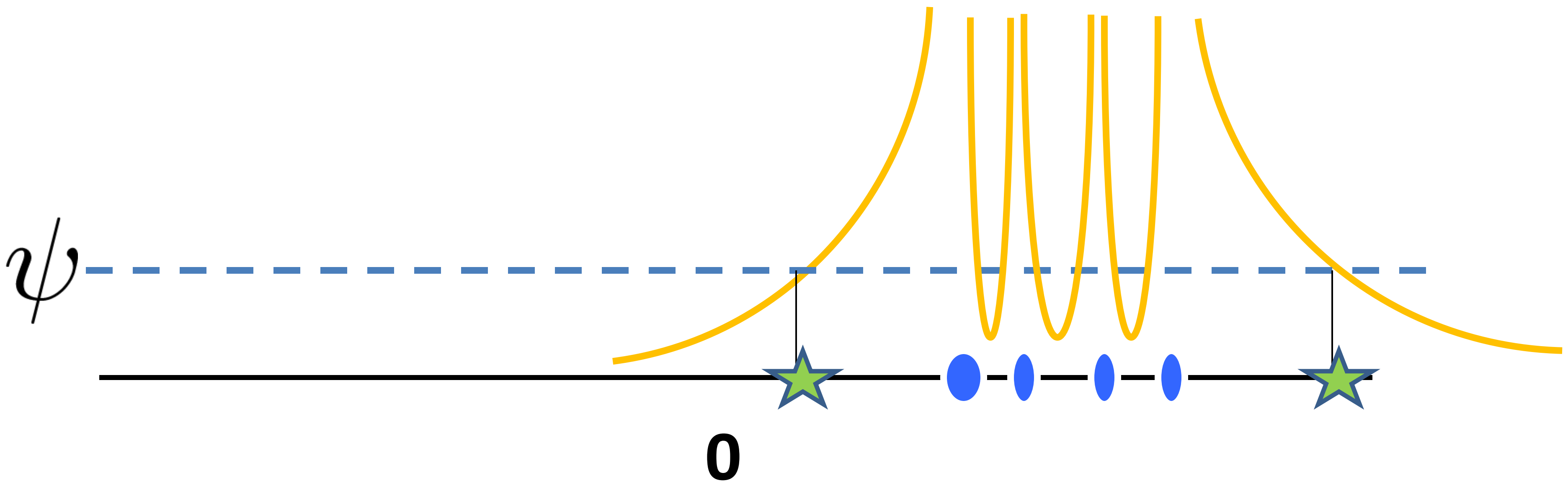


$$A_k = A_{k-1} + X_k X_k^T$$

Lemma.

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + k(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + k(1+\epsilon).$$

The Sensitivity Tradeoff

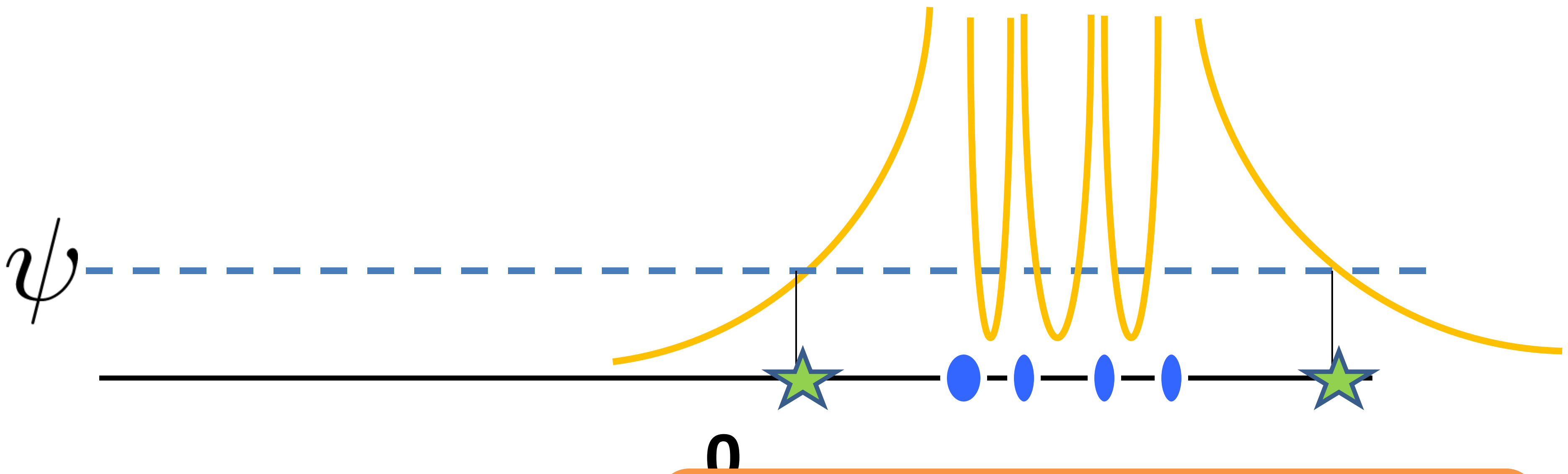


$$A_q = X_1 X_1^T + \ldots, X_q X_q^T$$

After **q** steps

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

The Sensitivity Tradeoff



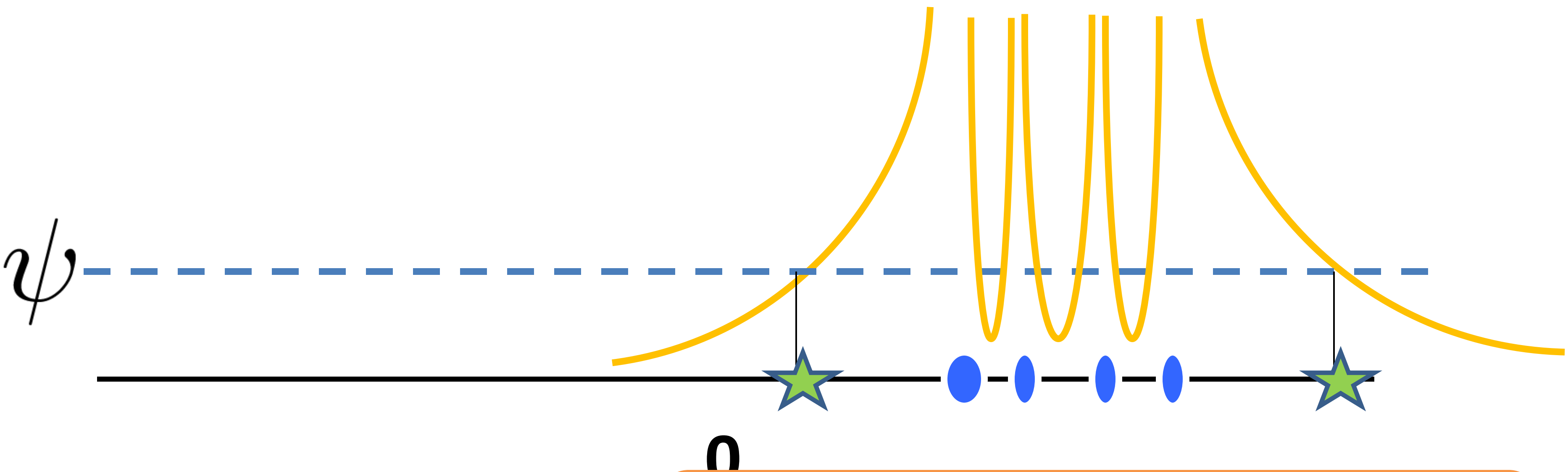
$$A_q = X_1 \ldots$$

Set $q > \frac{n}{\psi \epsilon} = c(\epsilon)n.$

After **q** steps

$$\mathbf{E}s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E}s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

The Sensitivity Tradeoff



$A_q = X_1$ …

Set $q > \frac{n}{\psi \epsilon} = c(\epsilon) n.$

After **q** steps

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

□

Proof of the Main Lemma

$A \succeq 0$, X **regular** isotropic random vector.
For all $\epsilon > 0$ there is $\psi = \psi(\epsilon)$ with

$$\mathbf{E} s_{max}(A + XX^T) \leq s_{max}(A) + 1 + \epsilon$$

$$\mathbf{E} s_{min}(A + XX^T) \geq s_{min}(A) + 1 - \epsilon.$$

For fixed **A**, **X**, how do we certify

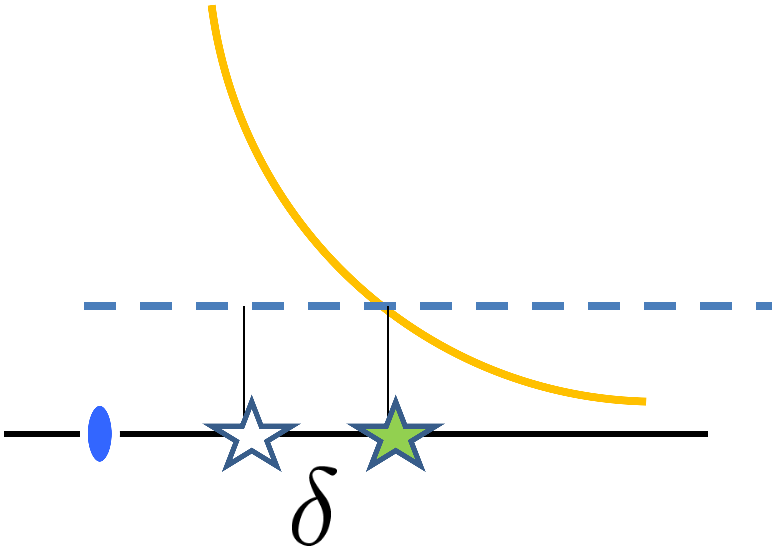
$$s_{max}(A + XX^T) \leq s_{max}(A) + \delta \text{ ?}$$

Upper Barrier Shifts

Let $s_{max}(A) = s.$

$$s_{max}(A + XX^T) \leq s + \delta$$

$$\iff \Phi_{A+XX^T}(s+\delta) \leq \psi$$



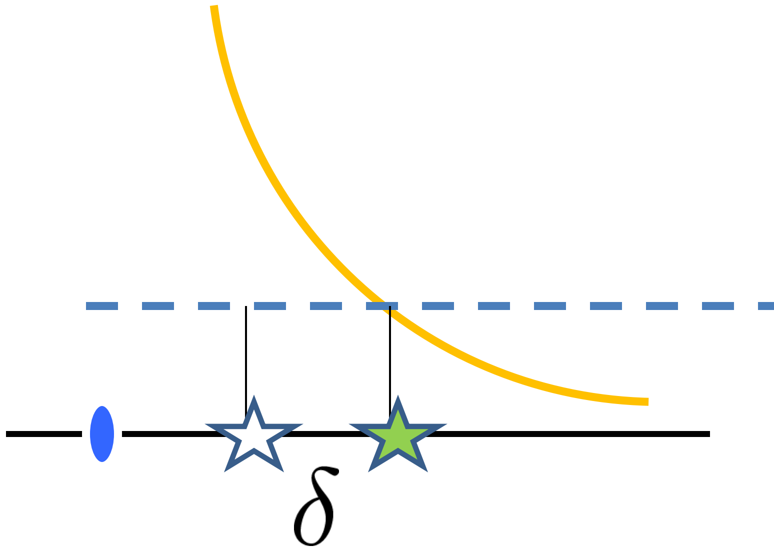
Upper Barrier Shifts

Let $s_{max}(A) = s.$

$s_{max}(A + XX^T) \leq s + \delta$

$\iff \Phi_{A+XX^T}(s+\delta) \leq \psi$

$\iff \Phi_{A+XX^T}(s+\delta) \leq \Phi_A(s)$



Upper Barrier Shifts

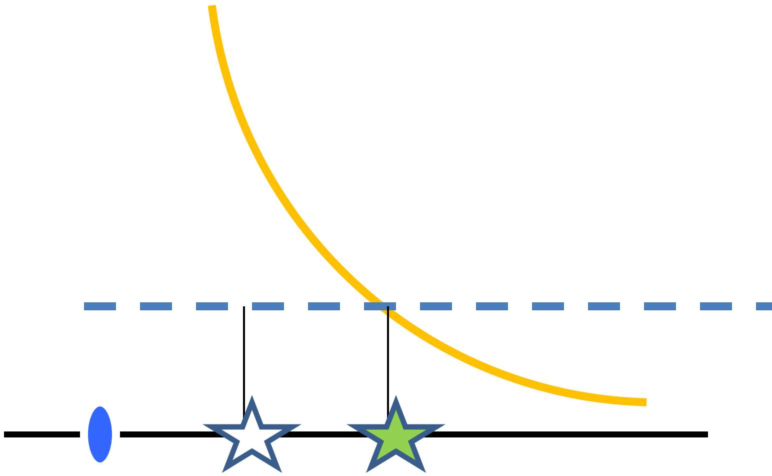
Let $s_{max}(A) = s.$

$$s_{max}(A + XX^T) \le s + \delta$$

$$\iff \Phi_{A+XX^T}(s+\delta) \le \psi$$

$$\iff \Phi_{A+XX^T}(s+\delta) \le \Phi_A(s)$$

$$\iff \frac{X^T(s+\delta-A)^{-2}X}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + X^T(s+\delta-A)^{-1}X \le 1.$$



Upper Barrier Shifts

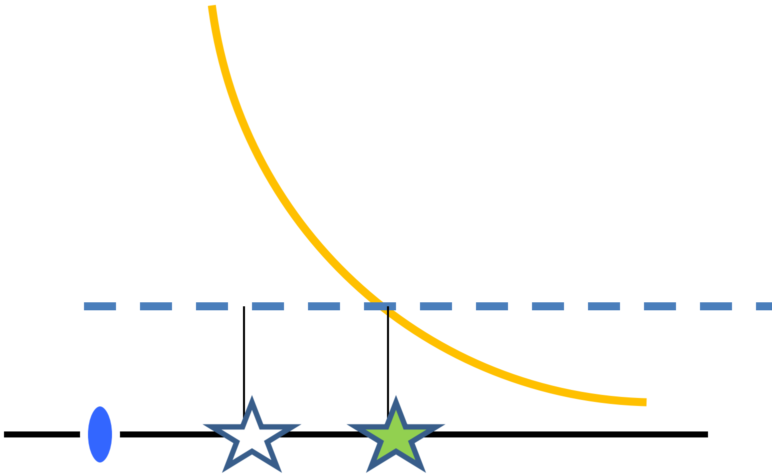
Let $s_{max}(A) = s.$

$$s_{max}(A + XX^T) \le s + \delta$$

$$\iff \Phi_{A+XX^T}(s+\delta) \le \psi$$

$$\iff \Phi_{A+XX^T}(s+\delta) \le \Phi_A(s)$$

$$\iff \frac{X^T(s+\delta-A)^{-2}X}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + X^T(s+\delta-A)^{-1}X \le 1.$$



$$\delta(X) := \min\{\delta : U_A(\delta, X) \le 1\}$$

$$\mathbf{E} s_{max}(A + XX^T) = s_{max}(A) + \mathbf{E}\delta(X)$$

"Heuristic" bound on $\mathbf{E}\delta(X)$

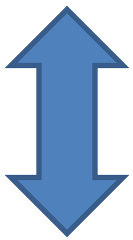
$$s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\Updownarrow$$

$$\frac{X^T(s+\delta-A)^{-2}X}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + X^T(s+\delta-A)^{-1}X \leq 1.$$

“Heuristic” bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$



$$\mathbf{E}\frac{X^T(s+\delta-A)^{-2}X}{\delta\mathrm{Tr}(s+\delta-A)^{-2}} + \mathbf{E}X^T(s+\delta-A)^{-1}X \leq 1.$$

"Heuristic" bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\Updownarrow$$

$$\frac{(s+\delta-A)^{-2} \bullet \mathbf{E}XX^T}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + (s+\delta-A)^{-1} \bullet \mathbf{E}XX^T \leq 1.$$

"Heuristic" bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\Updownarrow$$

$$\frac{(s+\delta-A)^{-2} \bullet \mathbf{E}XX^T}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + (s+\delta-A)^{-1} \bullet \mathbf{E}XX^T \leq 1.$$

$$\mathbf{E}XX^T = I$$

“Heuristic” bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\Updownarrow$$

$$\frac{\mathrm{Tr}(s+\delta-A)^{-2}}{\delta\mathrm{Tr}(s+\delta-A)^{-2}} + \mathrm{Tr}(s+\delta-A)^{-1} \leq 1.$$

“Heuristic” bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\Uparrow$$

$$\frac{1}{\delta} + \mathrm{Tr}(s + \delta - A)^{-1} \leq 1.$$

“Heuristic” bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\frac{1}{\delta} + \psi \leq 1.$$

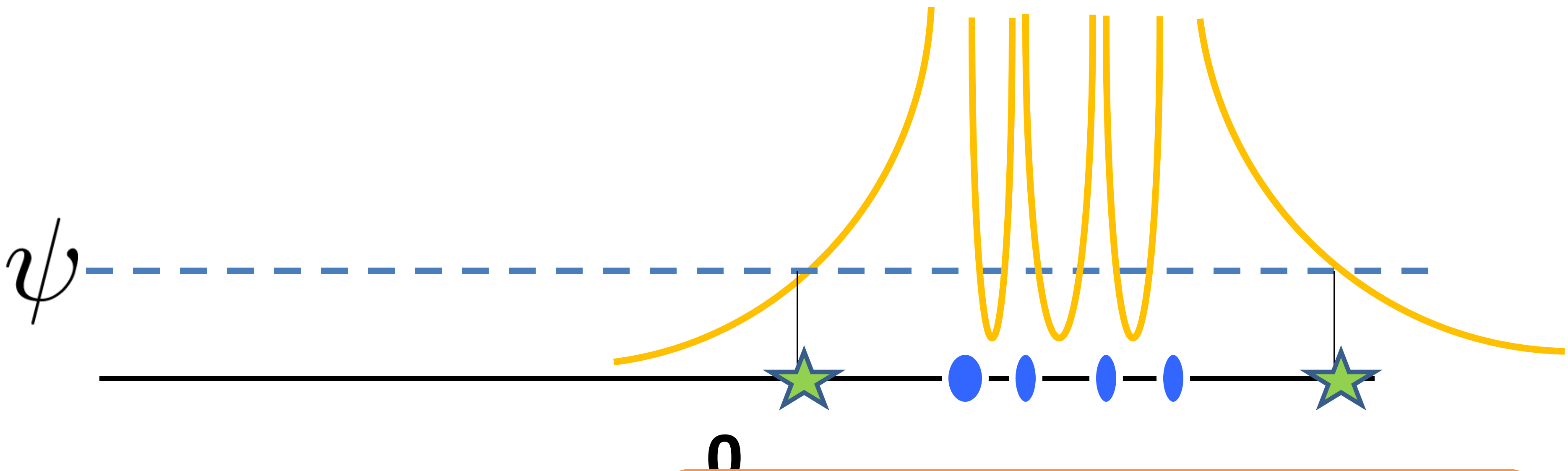
"Heuristic" bound on $\mathbf{E}\delta(X)$

$$\mathbf{E}s_{max}(A + XX^T) \leq s_{max}(A) + \delta$$

$$\frac{1}{\delta} + \psi \leq 1.$$

$$\psi \leq \frac{\delta - 1}{\delta} \approx \epsilon, \qquad \delta = 1 + \epsilon$$

The Sensitivity Tradeoff



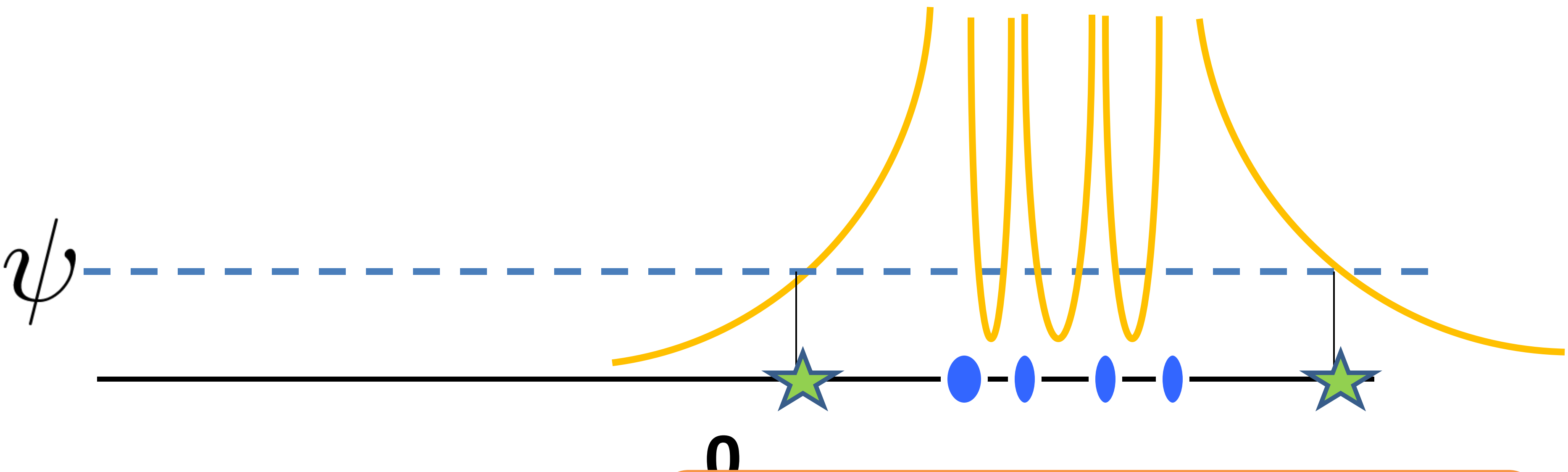
$A_q = X_1$ …

Set $q > \frac{n}{\psi \epsilon} = c(\epsilon) n.$

After **q** steps

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + q(1 - \epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + q(1 + \epsilon).$$

The Sensitivity Tradeoff



$A_q = X_1$ …

Set $q > \frac{n}{\psi \epsilon} = n/\epsilon^2$.

After **q** steps

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

Actual Proof

Need to bound $\mathbf{E}_X \delta(X)$ for δ satisfying

$$\frac{X^T(s+\delta-A)^{-2}X}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + X^T(s+\delta-A)^{-1}X \leq 1.$$

Insufficient to bound "in expectation"

Actual Proof

Need to bound $\mathbf{E}_X \delta(X)$ for δ satisfying

$$\frac{X^T(s+\delta-A)^{-2}X}{\delta \mathrm{Tr}(s+\delta-A)^{-2}} + \boxed{X^T(s+\delta-A)^{-1}X \leq 1.}$$

Actual Proof

Need to bound $\mathbf{E}_X \delta(X)$ for δ satisfying

$$\sum_i \frac{\langle X, u_i \rangle^2}{s + \delta - \lambda_i} \le 1.$$

$$\left(A = \sum_i \lambda_i u_i u_i^T \right)$$

Actual Proof

Need to bound $\mathbf{E}_X \delta(X)$ for δ satisfying

$$\sum_i \frac{\langle X, u_i \rangle^2}{s + \delta - \lambda_i} \leq 1.$$

Know:

$$\mathbf{E}\langle X, u_i \rangle^2 = 1 \qquad \sum_i \frac{1}{s - \lambda_i} = \Phi_A(s) \leq \psi$$

Abstract Problem

Need to bound $\mathbf{E}_Y \delta$ for δ satisfying

$$\sum_i \frac{Y_i}{\mu_i + \delta} \leq 1.$$

Given:

$$\mathbf{E} Y_i = 1 \qquad \sum_i \frac{1}{\mu_i} \leq \psi$$

$$Y_i = \langle X, u_i \rangle^2$$
$$\mu_i = s - \lambda_i$$

Example

Need to bound $\mathbf{E}_Y \delta$ for δ satisfying

$$\sum_i \frac{Y_i}{\mu_i + \delta} \leq 1.$$

Given:

$$\frac{1}{4} + \frac{1}{4} + \frac{1}{2} = 1$$

Example

Need

$$\frac{1}{4+\delta} + \frac{3}{4+\delta} + \frac{0.5}{2+\delta} \leq 1$$

$$\delta = 1$$

Given:

$$\frac{1}{4} + \frac{1}{4} + \frac{1}{2} = 1$$

An Important Case

$$\text{Subset } |S| = k \text{ with } \mu_i^{-1} = \frac{\mathbf{1}_{i \in S}}{2k}.$$

$$\delta :=$$

$$\sum_i \frac{Y_i}{\mu_i + \delta} \le 1.$$

$$\mathbf{E} Y_i = 1$$

$$\sum_i \frac{1}{\mu_i} = \frac{1}{2}$$

An Important Case

Subset $|S| = k$ with $\mu_i^{-1} = \frac{\mathbf{1}_{i \in S}}{2k}$.

$$\delta := \frac{\sum_{i \in S} Y_i}{2k + \delta} \leq 1.$$

$$\mathbf{E} Y_i = 1 \qquad \sum_i \frac{1}{\mu_i} = \frac{1}{2}$$

An Important Case

$$\text{Subset } |S| = k \text{ with } \mu_i^{-1} = \frac{\mathbf{1}_{i \in S}}{2k}.$$

$$\delta = \left(\sum_{i \in S} Y_i - 2k \right)_+ = \left(\sum_{i \in S} Y_i - 2 \sum_{i \in S} \mathbf{E} Y_i \right)_+$$

$$\mathbf{E} Y_i = 1$$

$$\sum_i \frac{1}{\mu_i} = \frac{1}{2}$$

An Important Case

$$\delta = \left(\sum_{i \in S} Y_i - 2k \right)_+ = \left(\sum_{i \in S} Y_i - 2 \sum_{i \in S} \mathbf{E} Y_i \right)_+$$

$$\mathbf{P}(\delta > t) = \mathbf{P} \left(\sum_{i \in S} Y_i > 2\mathbf{E} \sum_{i \in S} Y_i + t \right)$$

An Important Case

$$\delta = \left(\sum_{i \in S} Y_i - 2k \right)_+ = \left(\sum_{i \in S} Y_i - 2 \sum_{i \in S} \mathbf{E} Y_i \right)_+$$

$$\mathbf{P}(\delta > t) = \mathbf{P} \left(\sum_{i \in S} Y_i > 2\mathbf{E} \sum_{i \in S} Y_i + t \right)$$

$$\asymp \mathbf{P} \left(\sum_{i \in S} Y_i > t \right) \qquad t > 4\mathbf{E} \sum_{i \in S} Y_i$$

An Important Case

$$\delta = \left(\sum_{i \in S} Y_i - 2k \right)_+ = \left(\sum_{i \in S} Y_i - 2 \sum_{i \in S} \mathbf{E} Y_i \right)_+$$

$$\mathbf{P}(\delta > t) = \mathbf{P} \left(\sum_{i \in S} Y_i > 2\mathbf{E} \sum_{i \in S} Y_i + t \right)$$

$$\asymp \mathbf{P} \left(\sum_{i \in S} Y_i > t \right) \qquad t > 4\mathbf{E} \sum_{i \in S} Y_i$$

$$\text{need} \quad \lesssim \frac{1}{t^{1+\eta}} \qquad \text{for } \mathbf{E}\delta = O(1)$$

The Regularity Assumption

$$\mathbf{P}(\delta > t) \asymp \mathbf{P}\left(\sum_{i \in S} Y_i > t\right) \qquad t > 4\mathbf{E}\sum_{i \in S} Y_i \qquad \lesssim \frac{1}{t^{1+\eta}}$$

The Regularity Assumption

$$\mathbf{P}(\delta > t) \asymp \mathbf{P}\left(\sum_{i\in S} Y_i > t\right) \qquad t > 4\mathbf{E}\sum_{i\in S} Y_i \qquad \lesssim \frac{1}{t^{1+\eta}}$$

$$\sum_{i\in S} Y_i = \sum_{i\in S} \langle X, u_i\rangle^2 = \|\Pi_S X\|^2$$

$$\mathbf{E}\sum_{i\in S} Y_i = \mathrm{rank}(\Pi)$$

$$\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\mathrm{rank}(\Pi)}$$

The Regularity Assumption

$$\mathbf{P}(\delta > t) \asymp \mathbf{P}\left(\sum_{i \in S} Y_i > t\right) \qquad t > 4\mathbf{E}\sum_{i\in S} Y_i \qquad \lesssim \frac{1}{t^{1+\eta}}$$

$$\sum_{i\in S} Y_i = \sum_{i\in S} \langle X, u_i \rangle^2 = \|\Pi_S X\|^2$$

$$\mathbf{E}\sum_{i\in S} Y_i = \operatorname{rank}(\Pi)$$

$$\forall \Pi \quad \mathbf{P}(\|\Pi X\|_2 > t) \leq C/t^{2+\eta}, \quad t > C\sqrt{\operatorname{rank}(\Pi)}$$

$$\mathbf{E}\delta(X) = \int_0^\infty \mathbf{P}(\delta > t) = O(1)$$

The General Case

Split μ_i into blocks by magnitude.

$\delta :=$

$$\sum_i \frac{Y_i}{\mu_i + \delta} \leq 1.$$

Handle other quadratic form by 1D + Holder.

$$\frac{X^T (s + \delta - A)^{-2} X}{\delta \mathrm{Tr}(s + \delta - A)^{-2}} + X^T (s + \delta - A)^{-1} X \leq 1.$$

The Lower Shift

Need to show $\mathbf{E}\delta(X) \geq 1 - \epsilon$

Use truncation

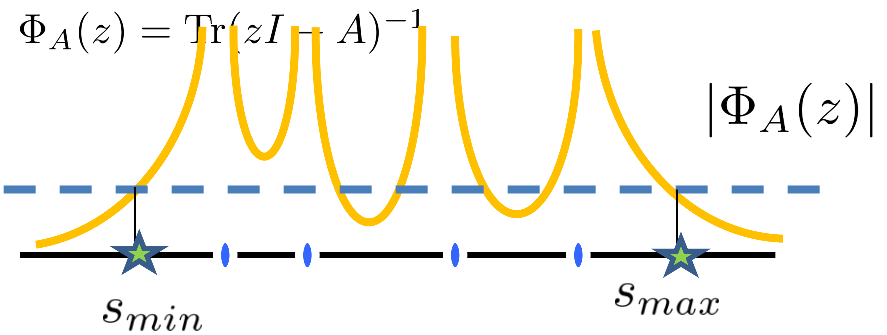
$$\mathbf{E}\delta(X) \geq \mathbf{E}\delta(X)\mathbf{1}_{\{\text{all } \langle X, u_i \rangle^2 \text{ are small}\}}$$

Don't need to handle rare events, 1D + Paley-Zygmund / Holder suffices.

Main Lemma

$$s_{min} := \min\{z : |\Phi_A(z)| = \psi\}$$

$$s_{max} := \max\{z : \Phi_A(z) = \psi\}$$



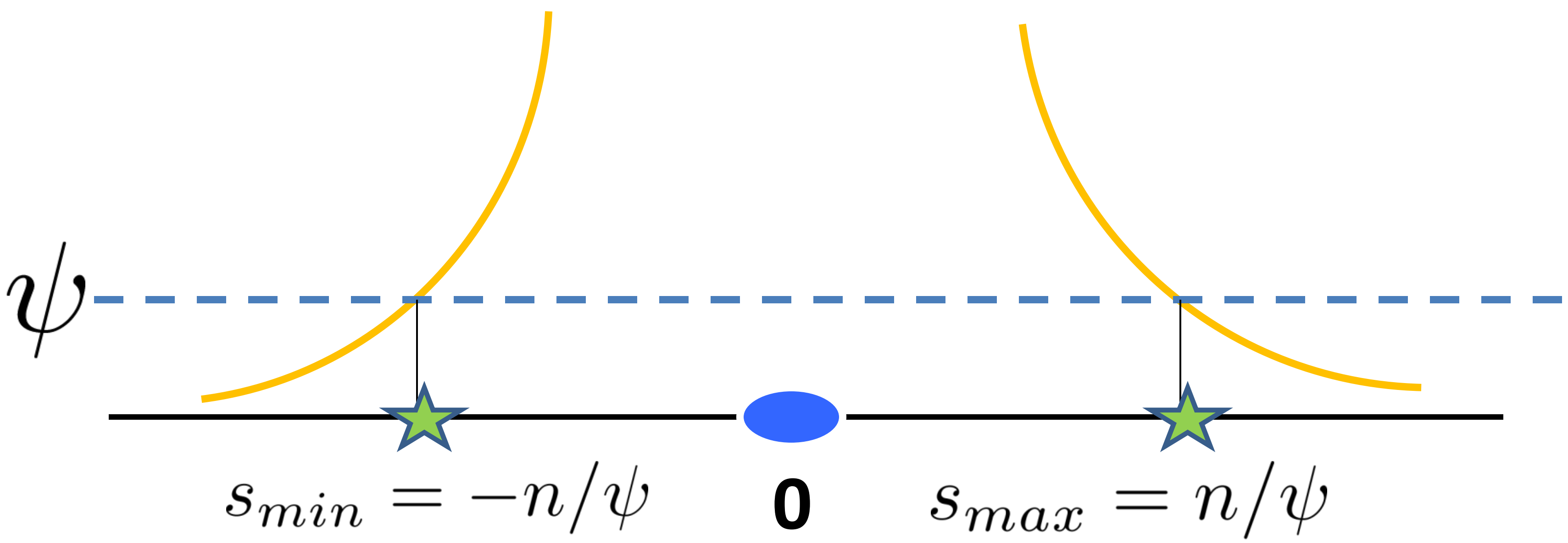
Main Lemma.

$A \succeq 0$, X **regular** isotropic random vector.
For all $\epsilon > 0$ there is $\psi = \psi(\epsilon)$ with

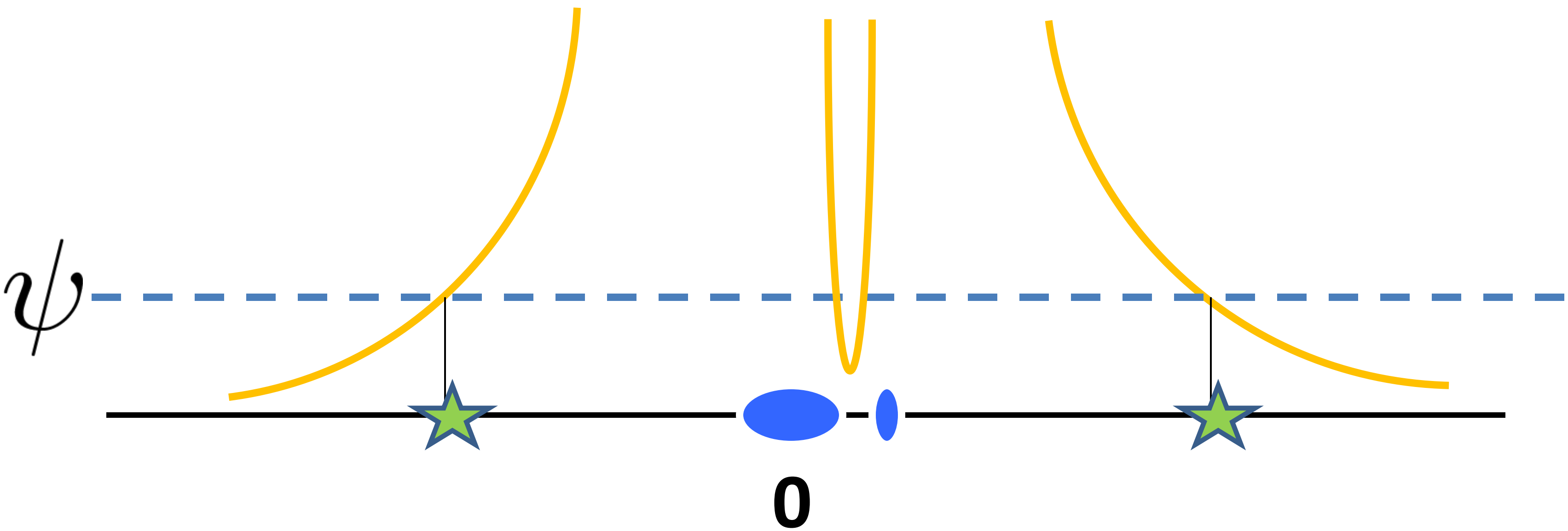
$$\mathbf{E} s_{max}(A + XX^T) \leq s_{max}(A) + 1 + \epsilon$$

$$\mathbf{E} s_{min}(A + XX^T) \geq s_{min}(A) + 1 - \epsilon.$$

End of the Proof

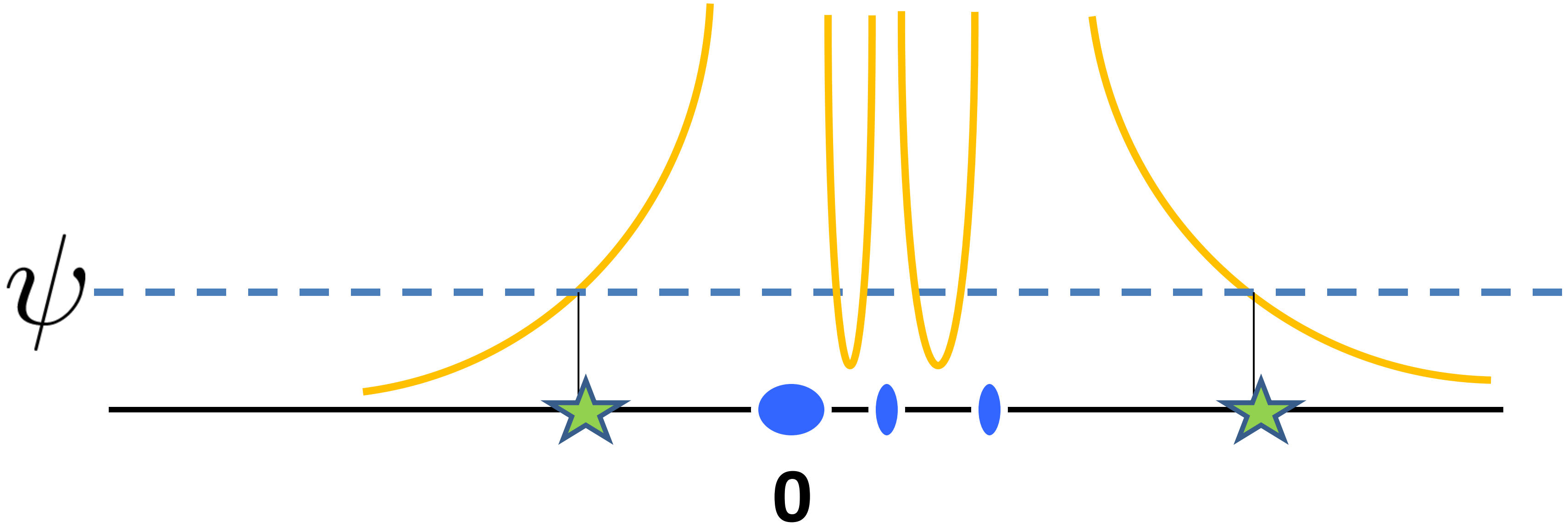


End of the Proof



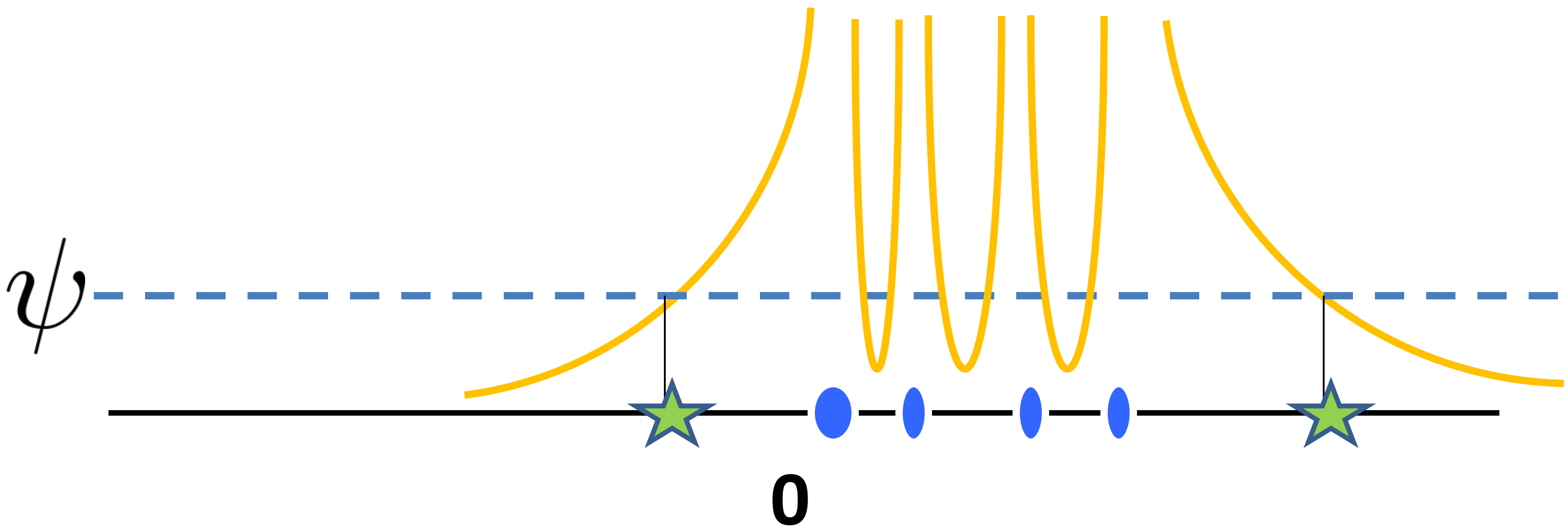
$$A_k = A_{k-1} + X_k X_k^T$$

End of the Proof



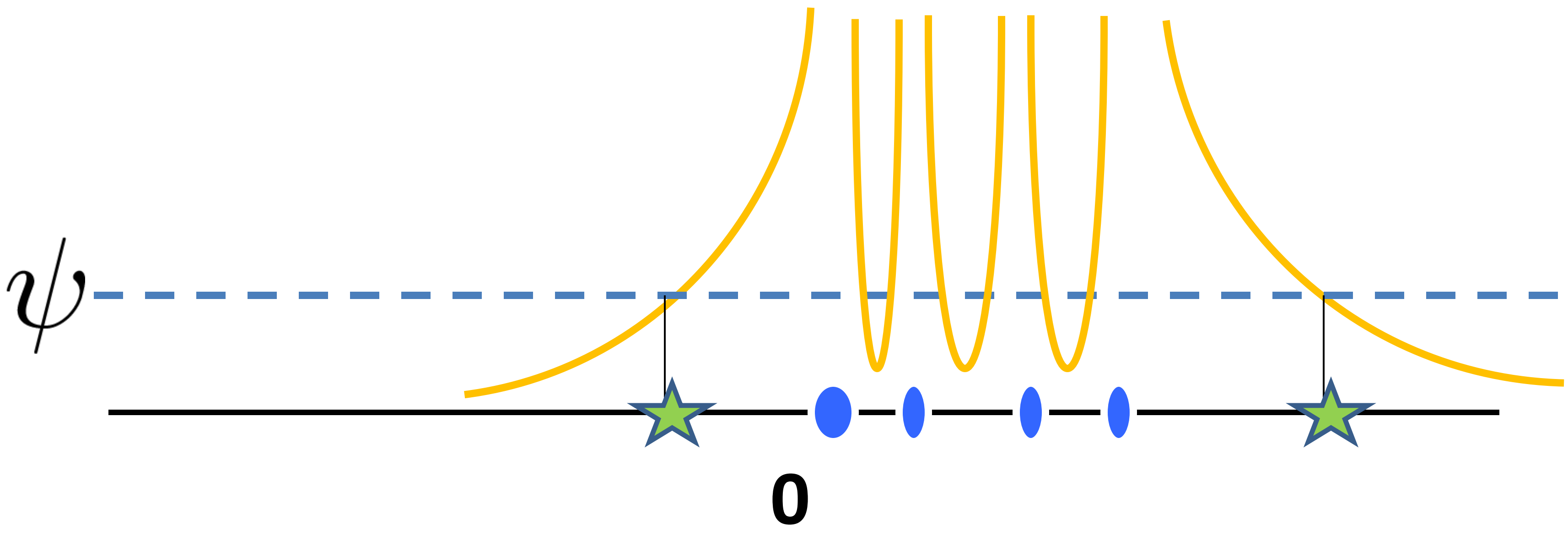
$$A_k = A_{k-1} + X_k X_k^T$$

End of the Proof



$$A_k = A_{k-1} + X_k X_k^T$$

End of the Proof

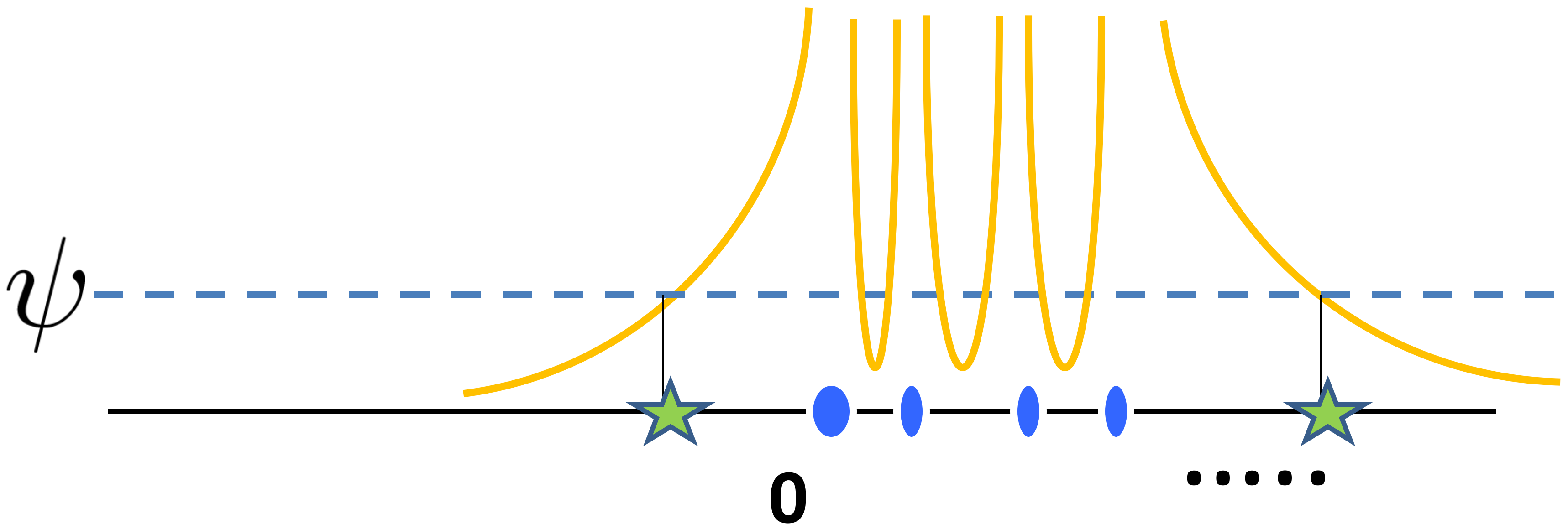


$$A_k = A_{k-1} + X_k X_k^T$$

Lemma.

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + k(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + k(1+\epsilon).$$

End of the Proof

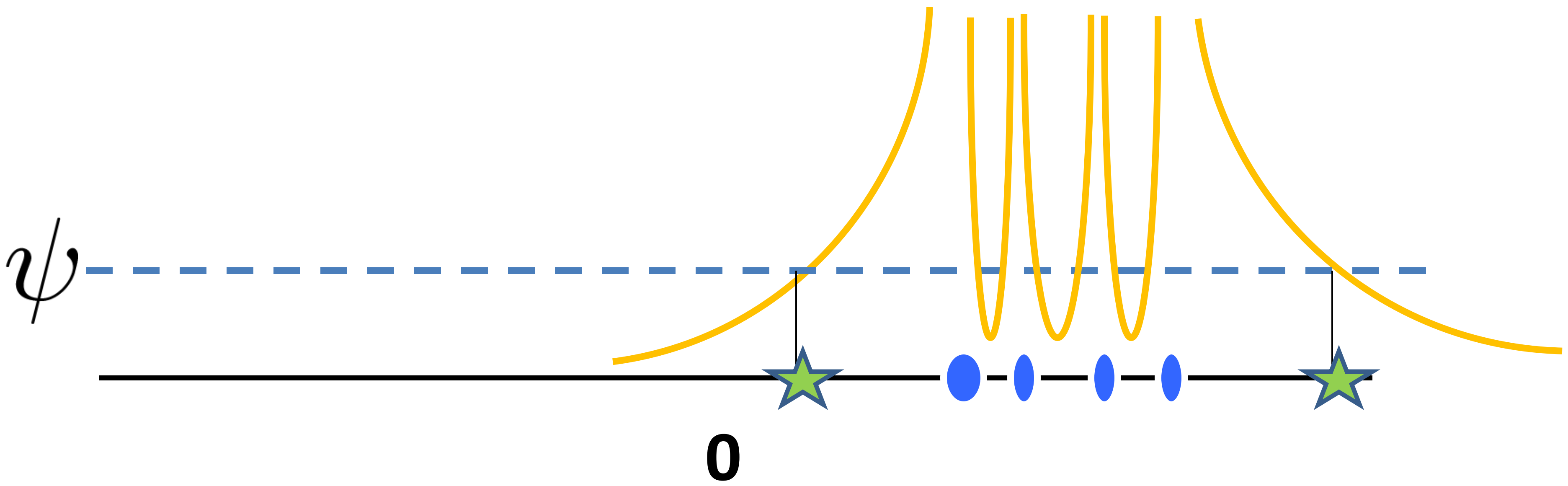


$$A_k = A_{k-1} + X_k X_k^T$$

Lemma.

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + k(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + k(1+\epsilon).$$

End of the Proof

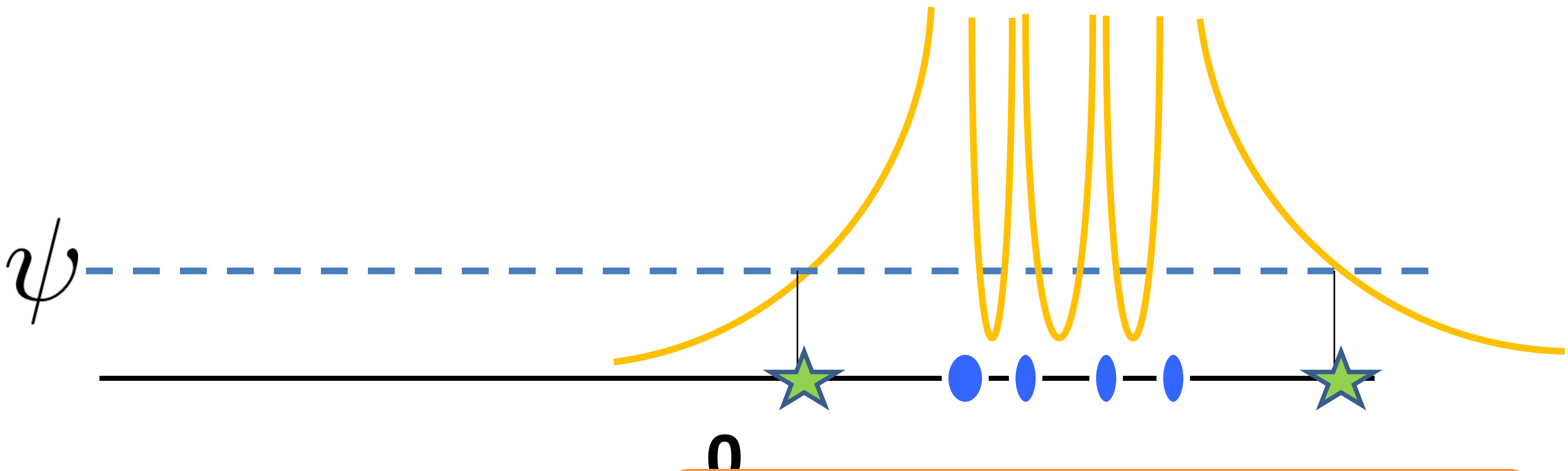


$$A_q = X_1X_1^T + \ldots, X_qX_q^T$$

After **q** steps

$$\mathbf{E}s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E}s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

End of the Proof



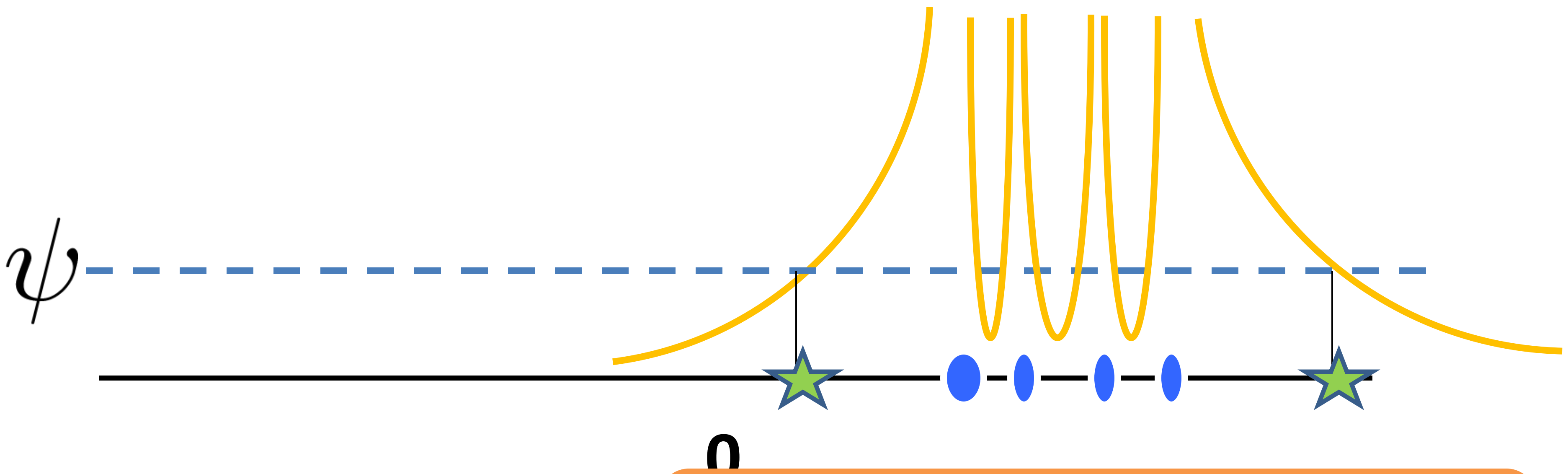
$$A_q = X_1 \ldots$$

Set $q > \frac{n}{\psi \epsilon} = c(\epsilon) n.$

After **q** steps

$$\mathbf{E} s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E} s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

End of the Proof



$A_q = X_1$ …

Set $q > \frac{n}{\psi\epsilon} = c(\epsilon)n.$

After **q** steps

$$\mathbf{E}s_{min}(A_k) \leq -n/\psi + q(1-\epsilon)$$
$$\mathbf{E}s_{max}(A_k) \geq n/\psi + q(1+\epsilon).$$

$\square$

Open Questions

Easily verifiable regularity assumption

Cf. [Vershynin'11] O(n(loglogn)^2)

Preserving higher marginals [RG'07, V'10]

High probability guarantees