Does block size matter in randomized block Krylov low-rank approximation?

Raphael A. Meyer
UC Berkeley

With: Tyler Chen, Ethan Epperly, Chris Musco, and Akash Rao
JPM Chase, UC Berkeley, NYU, WSU

Low-Rank Approximation

Given $\overset{\text{PSD}}{A} \in \mathbb{R}^{n \times n}$, k, $\varepsilon > 0$ find ortho $Q \in \mathbb{R}^{n \times k}$ with

$$\|A - QQ^T A\|_{F,2} \leq (1+\varepsilon)\|A - A_k\|_{F,2}$$

Ideally, $Q =$ top k eigenvectors of A, so use Krylov

[Rokhlin et al. '09], [Halko et al. '11], [Drineas Ipsen '19], [Tropp '22], ...

Block Krylov

1. Pick a start block Usually Gaussian

$B \in \mathbb{R}^{n \times b}$ b = block size

2. Build Krylov subspace

$$Z = \mathrm{orth}(K) = \mathrm{orth}([B \;\; AB \; \cdots \; A^t B])$$ Compute Ortho. Basis for cols of K

3. Return a solution

$Q = Z^T U_k$ where U_k = top k eigvecs of $Z^T A A^T Z$

Runtime is $O(n^2 b t)$. How should we pick b?

How should we pick b?

1. Theory: Use either $b \approx 1$ or $b \approx K$

Rich line of work [Tropp, Halko, Martinson, Gu, Drineas, Ipsen, Woodruff, ...]

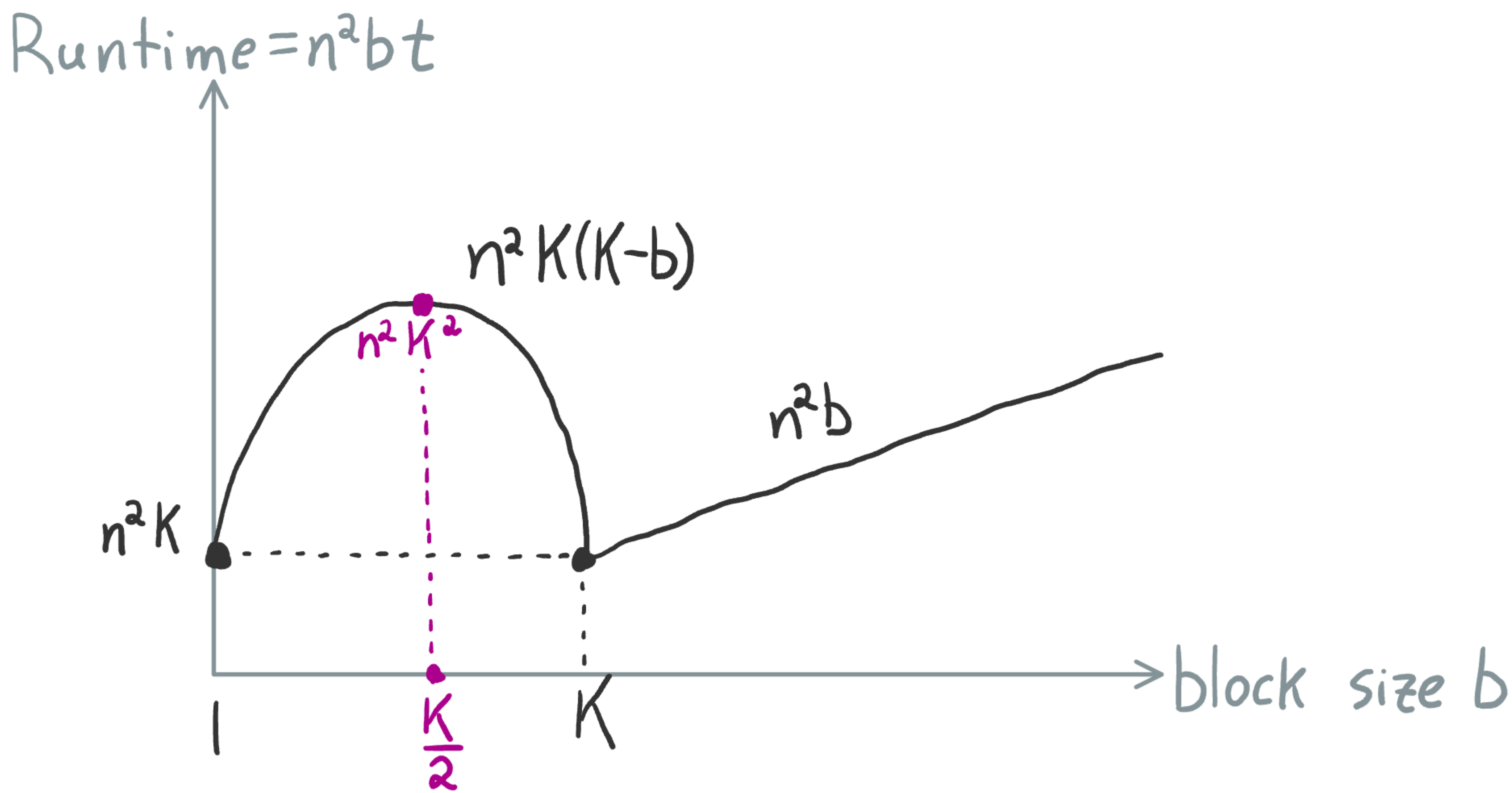
Large Block Convergence [Musco Musco '15]

$b \geq K \implies t = \tilde{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ suffices $\tilde{O}\left(n^2 b \frac{1}{\sqrt{\varepsilon}}\right)$ time

Small Block Convergence [Meyer Musco Musco '24]

$1 \leq b < K \implies t = \tilde{O}\left(\frac{K-b}{\sqrt{\varepsilon}}\right)$ suffices $\tilde{O}\left(n^2 b (K-b) \frac{1}{\sqrt{\varepsilon}}\right)$ time

The complexity landscape



$b = \frac{K}{2}$ is super expensive!?!

How should we pick b?

2. Practitioner: Completely ignore K

Bottleneck is computing matrix-matrix products

$$K = [B \quad AB \quad A^2B \quad \cdots \quad A^{t-1}B \quad A^tB]$$

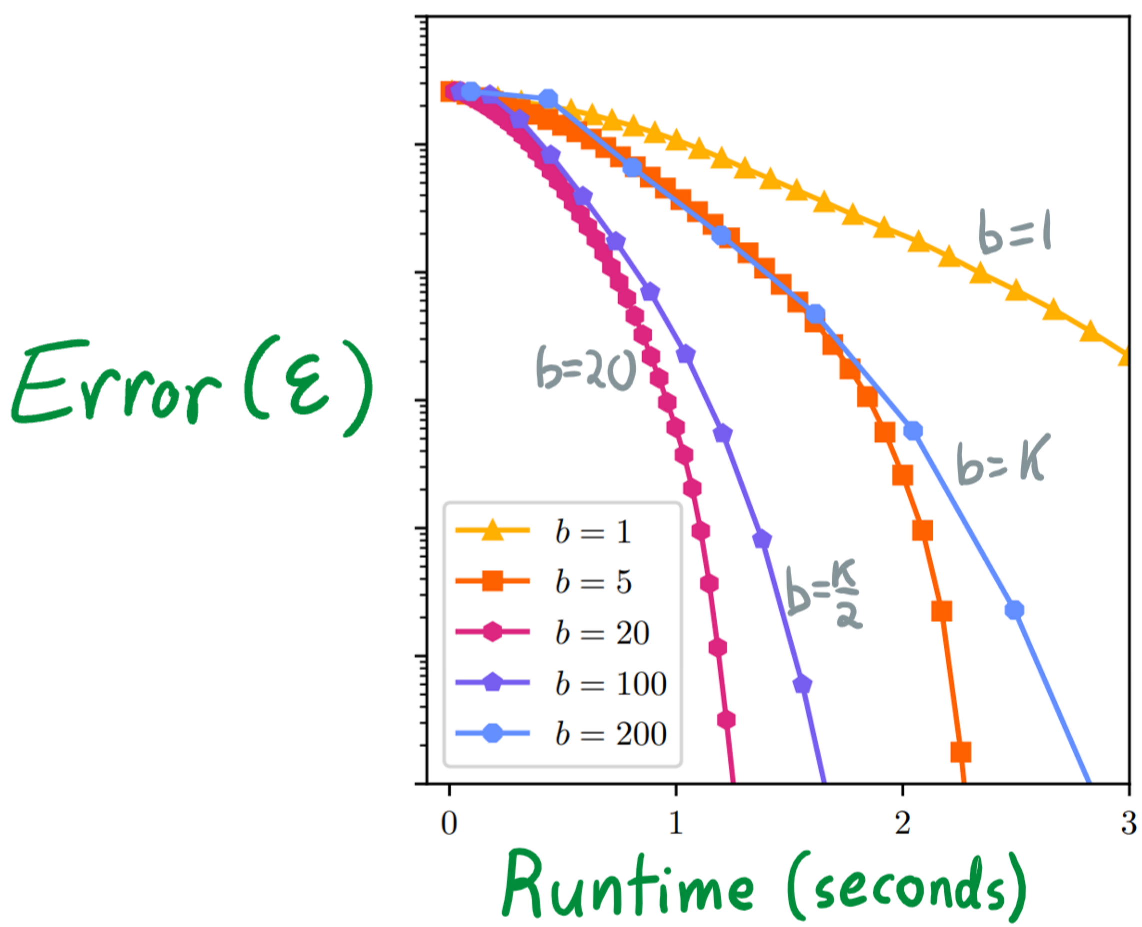
Product of $n \times n$ and $n \times b$ matrices

Unless your fast memory is saturated, you can increase b without almost any slow down!

Conclusion: use largest b that fits in fast memory!

Using this value of b works extremely well!

(K=200)



A theory/practice gap!

Should we pick block size b as a function of the target rank k in Block Krylov low-rank approximation?

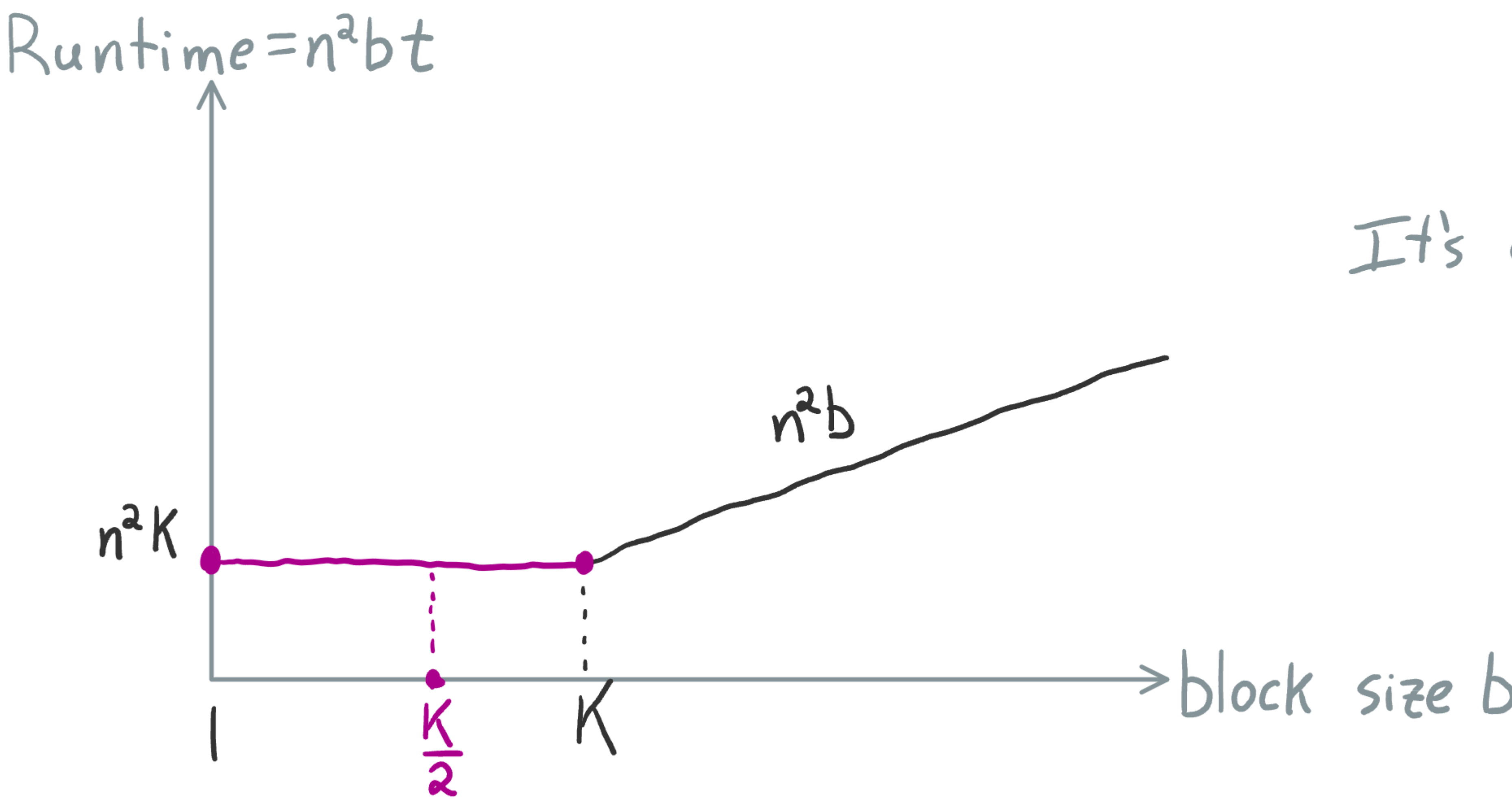
Caveat: Infinite Precision, Matrix-vector Complexity

Main Result

All Block Sizes Converge Fast

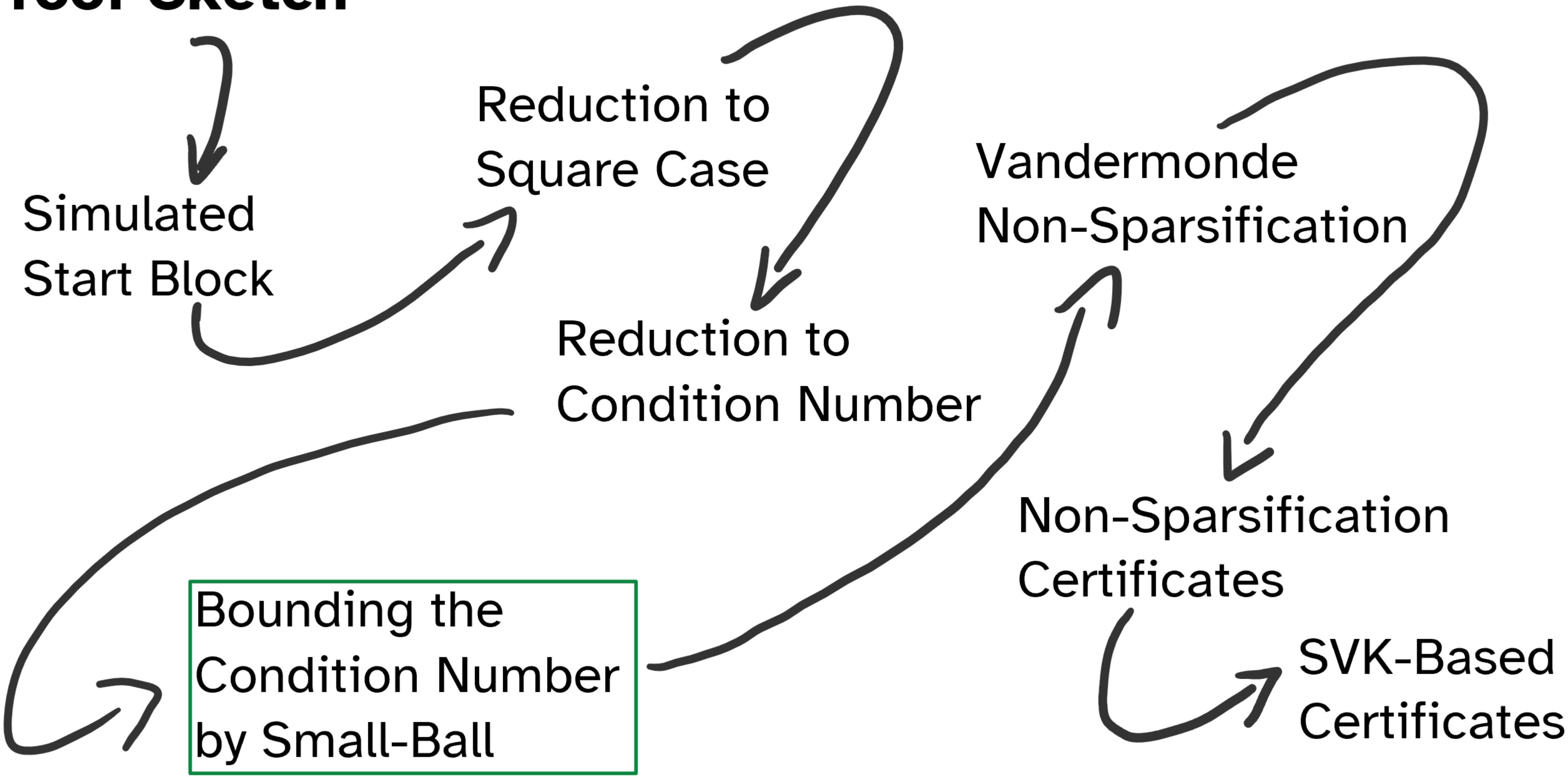
$1 \le b < K \implies t = \widetilde{O}\left(\frac{K}{b} \cdot \frac{1}{\sqrt{\varepsilon}}\right)$ suffices

$\widetilde{O}\left(n^2 K \frac{1}{\sqrt{\varepsilon}}\right)$ time



It's all cheap!

Proof Sketch



Intermediate Result: Krylov Matrix Conditioning

Let $C \in \mathbb{R}^{K \times K}$ with eigenvalues $\lambda_1, \cdots, \lambda_K$

$$\underset{K}{\overset{K}{\boxed{C}}}, \quad \underbrace{\lambda_i > 0}_{\text{PSD}}, \quad \underbrace{\mathrm{cond}(C) \leq \kappa}_{\text{well-conditioned}}, \quad \underbrace{|\lambda_i - \lambda_{i+1}| \geq \Delta \ \lambda_i}_{\text{bounded eigenvalue gaps}}$$

Suppose $K = bt$ exactly. Build Krylov matrix

$$\mathbb{K} = [B \quad CB \quad C^2B \cdots C^{t-1}B] \in \mathbb{R}^{K \times K} \qquad \underset{\text{Gaussian}}{B} \in \mathbb{R}^{K \times b}$$

Then,

$$\mathrm{cond}(\mathbb{K}) \leq \left(\frac{\Delta}{\kappa}\right)^t \mathrm{poly}(K) \quad \text{w.p. } \frac{2}{3}$$

[Peng Vempala 24] [Nie 22] had shown this for $b \geq \sqrt{K}$ only

In the paper: Grab bag of more implications

- **Beyond worst-case matrices**
- Smoothed Analysis
- Experiments
- A whole bunch of math

Any questions?