Eigenproblem Solvers

Issue w/ prior methods: Slow convergence

Introduce Inverse Shifts

Focus on symmetric $\underline{A}$,

Let $\underline{Q}^{(k)} = \underline{Q}_1 \underline{Q}_2 \cdots \underline{Q}_k$

$$\underline{R}^{(k)} = \underline{R}_k \underline{R}_{k-1} \cdots \underline{R}_1$$

from the QR method for eigenproblems

You can show that

$$\underline{A}^k = \underline{A}\,\underline{Q}^{(k-1)}\underline{R}^{(k-1)} = \underline{Q}^{(k)}\underline{R}_k \underline{R}^{(k-1)} = \underline{Q}^{(k)}\underline{R}^{(k)}$$

Since $\underline{A} = \underline{A}^T$, then:

$$(\underline{A}^{-1})^k = \underline{A}^{-k} = (\underline{A}^k)^{-1} = (\underline{Q}^{(k)}\underline{R}^{(k)})^{-1}$$

$$= ((\underline{R}^{(k)})^T \underline{Q}^{(k)T})^{-1} = \underline{Q}^{(k)}(\underline{R}^{(k)})^{-T}$$

$$= (\underline{A}^{-k})^T$$

Now, let $\underline{P} = \begin{bmatrix} 0 & & \cdot^{\cdot^{\cdot}} & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & 0 \end{bmatrix}$

$\underline{P}$ flips rows & Columns

$$\underline{P}\,\underline{A} = \underline{A}^T$$

Note that $\underline{P}^2 = \underline{I}$ $\qquad (\underline{A}^T)^T = \underline{A}$

$$\Rightarrow (\underline{A}^{-k})^T = \underline{A}^{-k}\,\underline{P} = \underline{Q}^{(k)}\,\underline{P}^2\,(\underline{R}^{(k)})^{-T}\,\underline{P}$$

$$= \underbrace{\left[\underline{Q}^{(k)}\,\underline{P}\right]}_{\text{Orthogonal}}\underbrace{\left[\underline{P}\,(\underline{R}^{(k)})^{-T}\,\underline{P}\right]}_{\text{Upper triangular}}$$

$\Rightarrow$ This is the QR factorization

$(\underline{A}^{-k})^T$ has a $\underline{Q}\,\underline{R}$ factorization

$\Rightarrow$ You can use the QR method for eigen problems on $\underline{A}^{-1}$

Algorithm: Shifted QR for Eigen problems

Let $\underline{A}_0$ be given by $\underline{Q}_0^T \underline{A}_0 \underline{Q}_0 = \underline{A}$
($\underline{A}_0$ comes from the Upper Hessenberg algorithm)

for $k = 1, 2, \ldots$

Pick a shift μ_k

$\underline{Q}_k \underline{R}_k = \underline{A}_{k-1} - \mu_k \underline{I}$ QR of shifted matrix

$\underline{A}_k = \underline{R}_k \underline{Q}_k + \mu_k \underline{I}$.

How do you pick μ_k?

Typically, you want the smallest eigenvalue you are looking for.

Try Rayleigh Quotient on the last column of $\underline{Q}_k$

$\mu_k = q_k^{(m)T} \underline{A}\, q_k^{(m)}$ $\quad q_k^{(m)}$ = last column of $\underline{Q}_k$,

It turns out that the value at the (m, m) location of $\underline{A}_k$

is

$$A_k(m,m) = q_k^{(m)T} \underline{A}\, q_k^{(m)}$$

$\Rightarrow$ Just set $\mu_k = A_k(m,m)$

Sometimes this is not stable

You could try Wilkinson Shift

$\Rightarrow$ Lecture 29 of Trefethen

If μ_k is chosen correctly, convergence will be 3^{rd}-order.

Singular Value Decomposition (SVD)

SVD is an extension of eigen systems to singular & rectangular matrices

Eigenproblems require that $\underline{A}$ be square & defective eigenvalues cause issues for eigen decomposition.

Instead, look for the singular values, σ, and the vectors $\underline{u}$ & $\underline{v}$, such that

$$\underline{A}\,\underline{v} = \sigma\,\underline{u} \quad , \quad \underline{A} \in \mathbb{R}^{m \times n} \quad , m \neq n$$

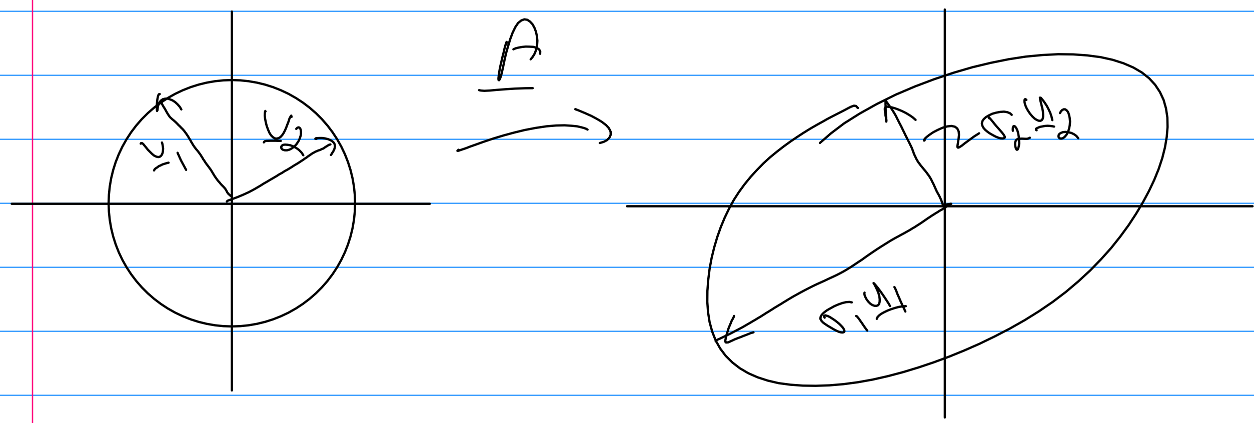
$\underline{v}$ is in the row space of $\underline{A}$

$\underline{u}$ is in the column space of $\underline{A}$.

$\Rightarrow$ We can expect to have $r = \text{rank}(\underline{A})$ of σ, $\underline{u}$ & $\underline{v}$

What does σ, $\underline{u}$ & $\underline{v}$ do?

Look at the application of $\underline{A}$ to the unit circle



The singular decomposition gives the principle directions of the hyperellipes of $\underline{A}$ applied to the unit circle.

Let $\hat{\underline{V}} = [\underline{v}_1\ \underline{v}_2 \cdots \underline{v}_r]$ $\quad \hat{\underline{U}} = [\underline{u}_1\ \underline{u}_2 \cdots \underline{u}_r]$

$$\hat{\underline{\Sigma}} = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}$$

$\Rightarrow \underline{A}\,\hat{\underline{V}} = \hat{\underline{U}}\,\hat{\underline{\Sigma}}$

$\hat{\underline{V}}$ & $\hat{\underline{U}}$ are both unitary

$\Rightarrow \hat{\underline{V}}^T \hat{\underline{V}} = \underline{I} \qquad \hat{\underline{U}}^T \hat{\underline{U}} = \underline{I}$

$\Rightarrow \underline{A} = \hat{\underline{U}}\,\hat{\underline{\Sigma}}\,\hat{\underline{V}}^T$ ← The reduced SVD

If $r < \min(m,n)$

$\Rightarrow$ Non-zero null space

$\Rightarrow$ we have the set of vectors that correspond to singular values of 0.

$$\underline{A}\,\underline{u} = \sigma\,\underline{v} = 0\,\underline{v} = \underline{0}$$

$\Rightarrow$ $\underline{u}$ = null space of $\underline{A}$.

$\Rightarrow$ The full SVD of $\underline{A}$ is then:

$$\underline{A}\,[\underbrace{\underline{v}_1 \cdots \underline{v}_n}_{\substack{r \text{ vectors} \\ \text{in} \\ \text{Row Space}}} \;\; \underbrace{\underline{v}_{r+1} \cdots \underline{v}_n}_{\substack{n-r \\ \text{vectors} \\ \text{in null} \\ \text{space}}}] = [\underbrace{\underline{u}_1 \cdots \underline{u}_r}_{\substack{r \text{ vecs} \\ \text{in Col.} \\ \text{Space}}} \;\; \underbrace{u_{r+1} \cdots u_m}_{\substack{m-n \\ \text{vecs} \\ \text{in} \\ \text{Null}(A^T)}}]\,\underline{\Sigma}$$

$\Rightarrow$ $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$ contains the orthonormal basis for all 4 matrix subspaces.

Formal Definition

Let $\underline{A} \in \mathbb{R}^{m\times n}$, $m \geq n$ not required
$\underline{A}$ might not be full rank.

SVD of $\underline{A}$ is given by $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$

$\underline{U} \in \mathbb{R}^{m\times m}$ is unitary

$\underline{V} \in \mathbb{R}^{n\times n}$ is unitary

$\underline{\Sigma} \in \mathbb{R}^{m\times n}$ is diagonal

It is also assumed that all σ_j in Σ are real, non-negative & in non-increasing order!

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0 \qquad p = \min(m,n)$$

To show real & non-negative look at $\underline{A}^T\underline{A}$

$$\underline{A}^T\underline{A} = (\underline{U}\,\underline{\Sigma}\,\underline{V}^T)^T(\underline{U}\,\underline{\Sigma}\,\underline{V}^T)$$

$$= \underline{V}\,\underline{\Sigma}^T\underline{U}^T\,\underline{U}\,\underline{\Sigma}\,\underline{V}^T$$

$$= \underline{V}\,\underline{\Sigma}^T\underline{\Sigma}\,\underline{V}^T = \underline{V}\,\underline{\Sigma}^2\,\underline{V}^T$$

$\rightarrow$ looks like an eigen decomposition of $\underline{A}^T\underline{A}$.

Since $\underline{A}^T\underline{A}$ is normal $\Rightarrow \underline{V}$ is unitary

Since $\underline{A}^T\underline{A}$ is real & symmetric $\Rightarrow$ All eigenvalues are real

Now look at $\underline{x}^T(\underline{A}^T\underline{A})\underline{x}$ for any $\underline{x}$,

$$\underline{x}^T(\underline{A}^T\underline{A})\underline{x} = (\underline{A}\,\underline{x})^T(\underline{A}\,\underline{x}) = \underline{y}^T\underline{y} > 0$$

$\Rightarrow \underline{A}^T\underline{A}$ is positive definite

$\Rightarrow \underline{A}^T\underline{A}$ only has positive eigenvalues,

Since $\underline{\Sigma}^2$ is the matrix of eigenvalues of $\underline{A}^T\underline{A}$

$\Rightarrow \sigma_j = \sqrt{\lambda_j} \Rightarrow$ will be positive & real

Thm: Every matrix $\underline{A} \in \mathbb{R}^{m \times n}$ has a SVD, and the singular values $\{\sigma_j\}$ are all uniquely determined.

If $\underline{A}$ is square & all $\{\sigma_j\}$ are distinct, then $\{\underline{u}_j\}$ & $\{\underline{v}_j\}$ are uniquely determined up to a sign.

Properties

Let $\underline{A} \in \mathbb{R}^{m \times n}$, $p = \min(m,n)$
$r = \#$ of singular values $\leq p$

1. $\text{rank}(\underline{A}) = r$

2. $\text{range}(\underline{A}) = \text{span}(\underline{u}_1, \underline{u}_2, \dots, \underline{u}_r)$

$\text{null}(\underline{A}) = \text{span}(\underline{v}_{r+1}, \dots, \underline{v}_n)$

3. $\|\underline{A}\|_2 = \sigma_1$

$\|\underline{A}\|_F = (\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2)^{1/2}$

4. Non-zero singular values of $\underline{A}$ are the square roots of the eigenvalues of

$$\underline{A}^T \underline{A} \quad \text{or} \quad \underline{A}\,\underline{A}^T$$

5. If $\underline{A} = \underline{A}^T$, σ_j is $|\lambda_j|$ of $\underline{A}$,

6. If $\underline{A} \in \mathbb{R}^{m \times m}$, $|\det(\underline{A})| = \prod_{i=1}^{m} \sigma_i$

$\Rightarrow$ If $\underline{A}$ is square but one $\sigma_i = 0$, $\underline{A}^{-1}$ does not exist.

Because of 2 above, the SVD says that <u>any</u> matrix can be made diagonal if one uses the proper row & column space basis.

Consider $\underline{A}\,\underline{x} = \underline{b}$ $\quad \underline{A} \in \mathbb{R}^{m \times n}$

$\underline{x} \in \mathbb{R}^{n} \quad \underline{b} \in \mathbb{R}^{m}$

$\underline{V}$ spans $\mathbb{R}^{n}$, while $\underline{U}$ spans $\mathbb{R}^{m}$

$\Rightarrow$ I can write $\underline{x}$ in terms of coordinates of $\underline{V}$:

$$\underline{x}' = \underline{V}^T \underline{x}$$

Similarly, $\underline{b}' = \underline{U}^T \underline{b}$

$\underline{A}\,\underline{x} = \underline{b} \Rightarrow \underline{U}^T \underline{A}\,\underline{x} = \underline{U}^T \underline{b}$

$\Rightarrow \underline{U}^T \underline{U}\,\underline{\Sigma}\,\underline{V}^T \underline{x} = \underline{U}^T \underline{b}$

$\Rightarrow \underline{\Sigma}\,\underline{V}^T \underline{x} = \underline{U}^T \underline{b} \Rightarrow \underline{\Sigma}\,\underline{x}' = \underline{b}'$

Coordinates in $\underline{U}$

Coord in $\underline{V}$

Diagonal matrix

Uses of SVD

① Psuedo- Inverse

All matrices have $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$

Define the psuedo-Inverse as

$\underline{A}^+ \underline{A} = \underline{I} = \underline{A}\,\underline{A}^+$ Note: $\underline{A}^{-1}$ might not exist

Let $\underline{A}^+ = \underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T$, $\underline{\Sigma}^{-1} = \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \sigma_2^{-1} & \\ 0 & & \ddots \end{bmatrix}$

$\underline{A}^+\underline{A} = (\underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T)(\underline{U}\,\underline{\Sigma}\,\underline{V}^T)$

$= \underline{V}\,\underline{\Sigma}^{-1}\underline{\Sigma}\,\underline{V}^T = \underline{V}\,\underline{V}^T = \underline{I}$

$\underline{A}\,\underline{A}^+ = (\underline{U}\,\underline{\Sigma}\,\underline{V}^T)(\underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T) = \underline{U}\,\underline{\Sigma}\,\underline{\Sigma}^{-1}\underline{U}^T$

$= \underline{U}\,\underline{U}^T = \underline{I}$

(2) Low Rank Approximations

Let $\Sigma_j = \begin{bmatrix} 0 & & & & \\ & \ddots & & 0 & \\ & & 0 & & \\ & & & \sigma_j & \\ & & & & 0 \\ & 0 & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$

Then $\underline{U}\,\underline{\Sigma}_j\,\underline{V}^T = \sigma_j\,\underline{u}_j\,\underline{v}_j^T$ $\quad \underline{u}_j = j^{th}$ Column of $\underline{U}$

$\underline{v}_j^T = j^{th}$ column of $\underline{V}$

$$\Rightarrow \underline{U}\,\underline{\Sigma}\,\underline{V}^T = \sum_{j=1}^{r} \underline{U}\,\underline{\Sigma}_j\,\underline{V}^T = \sum_{j=1}^{r} \sigma_j\,\underline{u}_j\,\underline{v}_j^T$$

$\Rightarrow$ Any matrix $\underline{A}$ can be written as the finite sum of rank-1 matrices.

Thm: Let $\underline{A}_\nu = \sum_{j=1}^{\nu} \sigma_j \underline{u}_j \underline{v}_j^T$ be a low rank approximation of $\underline{A}$, where $\nu \leq \text{rank}(\underline{A})$

Then, it can be shown that

$$\|\underline{A} - \underline{A}_\nu\|_2 = \inf_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank}(B) \leq \nu}} \|\underline{A} - \underline{B}\|_2 = \sigma_{\nu+1}$$

where $\sigma_{\nu+1} = 0$ if $\nu = p = \min(m, n)$

$\Rightarrow$ $\underline{A}_\nu$ minimizes the error

You can also show that $\underline{A}_\nu$ minimizes

$\|\underline{A} - \underline{A}_\nu\|_F$ error,

To show this in use, look at compression.

ex.) Image compression: Focus on gray-scale

An image is just a matrix w/ values between 0 & 255

0 = black 255 = white

Look at a 256 x 512 pixel image

Storing the full image takes

$256 * 512 = 131072$ pixels / data points

Instead, store only the 5 largest singular values.

$$\underline{A} = \text{Image} \approx \sigma_1 \underline{u}_1 \underline{v}_1^T + \sigma_2 \underline{u}_2 \underline{v}_2^T + \cdots \sigma_5 \underline{u}_5 \underline{v}_5^T$$

=> Size of compressed image is

$5 + 5(256 + 512) = 3845$ data points

Compression Ratio: $\dfrac{131072}{3845} \approx \dfrac{34}{1}$

Compute SVD

Let $\underline{A} \in \mathbb{R}^{m \times n}$

Naive Method:

① Compute eigendecomposition of $\underline{A}^T \underline{A}$

$$\underline{A}^T \underline{A} = \underline{V} \underline{\Lambda} \underline{V}^T$$

② $\underline{\Sigma} = \underline{\Lambda}^{1/2}$

③ Solve $\underline{U} \underline{\Sigma} = \underline{A} \underline{V}$ for $\underline{U}$

$\Rightarrow$ Condition # of $\underline{A}^T \underline{A}$ is large,

$\Rightarrow$ loss of accuracy if $\sigma_k \ll \|\underline{A}\|$

Use iterative methods

2-Step Process

① Convert $\underline{A}$ to bi-Diagonal

② Convert bi-Diagonal to diagonal

$$\overset{A}{\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}} \rightarrow \begin{bmatrix} x & x & 0 \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \\ 0 & 0 & 0 \end{bmatrix}$$

Golub-Kahan (G-K) Bidiagonalization

Use Householder on both left & right:

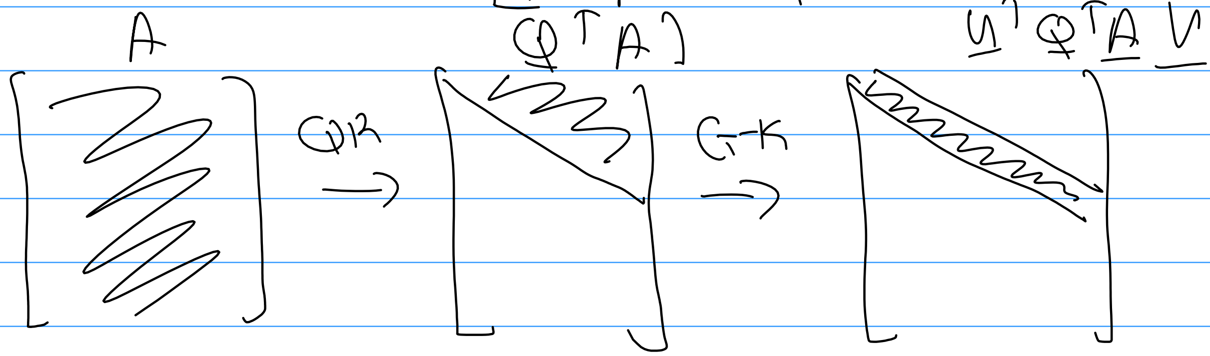
$$\underset{A}{\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}} \xrightarrow{U_1^T} \underset{U_1^T A}{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix}} \xrightarrow{V_1} \underset{U_1^T A V_1}{\begin{bmatrix} x & x & 0 \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix}} \xrightarrow{U_2^T} \cdots$$

Cost $\sim 4mn^2 - \frac{4}{3}n^3$

$\sim 2x$ as QR decomposition, but result is pretty close to the final B.

Alternative: Lawson-Hanson-Chan (LHC)

Do QR on A first, then G-K on the B matrix.



Cost of LHC is

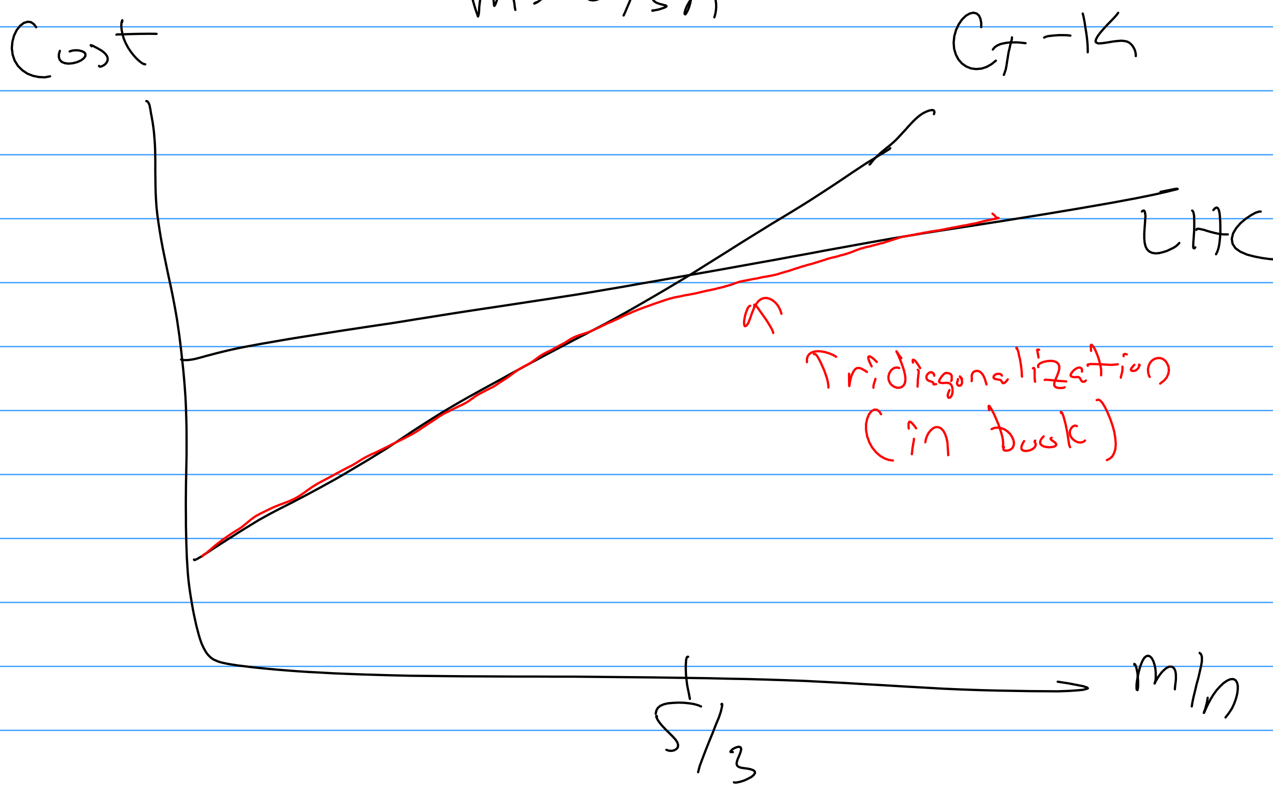
$$QR \sim 2m^2 - \frac{2}{3}n^3$$

$$\text{Gr-K on } R \sim \frac{8}{3}n^3$$

$$\text{Total} \sim 2mn^2 + 2n^3 \leftarrow$$

LHC is cheaper than Gr-K if

$$m > 5/3\, n$$



We do <u>not</u> talk about how to

Diagonlize the result.