An implementation of parallel solution of symmetric positive definite systems with hyperbolic rotations

Tse-Yu Lin

Data Science Degree Program

National Taiwan University & Academia Sinica

January 29, 2018

Outline

- In this project, we aim to solve the following symmetric positive definite (spd) system using different approach.

$$Ax = b,$$

where $A \in M_{n\times n}(\mathbb{R})$ is a spd matrix, and $b \in \mathbb{R}^n$ be a given column vector.

Definition

A symmetric n-by-n real matrix A is called is said to be positive definite if the $x^T A x \geq 0$ for any $x \in \mathbb{R}^n$.

- In this project, we aim to solve the following symmetric positive definite (spd) system using different approach.

$$Ax = b,$$

where $A \in M_{n \times n}(\mathbb{R})$ is a spd matrix, and $b \in \mathbb{R}^n$ be a given column vector.

Definition

A symmetric n-by-n real matrix A is called is said to be positive definite if the $x^T Ax \geq 0$ for any $x \in \mathbb{R}^n$.

Some characterizations of spd matrix A are shown below:

Proposition

- *All eigenvalues of A are positive.*
- *It associated bilinear form defines an inner prodect*
- *All eading principal minors of A are all positive.*
- A **has a unique Cholesky decomposition.**
  i.e., $A = LL^T$, where L is a lower triangular matrix.

Some characterizations of spd matrix A are shown below:

Proposition

- *All eigenvalues of A are positive.*
- *It associated bilinear form defines an inner prodect*
- *All eading principal minors of A are all positive.*
- A **has a unique Cholesky decomposition.**

  i.e., $A = LL^T$, where L is a lower triangular matrix.

Some characterizations of spd matrix A are shown below:

Proposition

- *All eigenvalues of A are positive.*
- *It associated bilinear form defines an inner prodect*
- *All eading principal minors of A are all positive.*
- **A has a unique Cholesky decomposition.**

  i.e., $A = LL^T$, where L is a lower triangular matrix.

Some characterizations of spd matrix A are shown below:

Proposition

- *All eigenvalues of A are positive.*
- *It associated bilinear form defines an inner prodect*
- *All eading principal minors of A are all positive.*
- A **has a unique Cholesky decomposition.**
  i.e., $A = LL^T$, where L is a lower triangular matrix.

Let

$$Ax = b,$$

where $A \in M_{n\times n}(\mathbb{R})$ is a spd matrix, and b is given.

An usual way to solve above system is as follows:

- Let $A = LL^T$ be its Cholesky decomposition.
- Using forward substitution to solve $Ly = b$.
- Using back substitution to solve $L^T x = y$.

Let

$$Ax = b,$$

where $A \in M_{n\times n}(\mathbb{R})$ is a spd matrix, and b is given.

An usual way to solve above system is as follows:

- Let $A = LL^T$ be its Cholesky decomposition.
- Using forward substitution to solve $Ly = b$.
- Using back substitution to solve $L^T x = y$.

Let

$$Ax = b,$$

where $A \in M_{n\times n}(\mathbb{R})$ is a spd matrix, and b is given.

An usual way to solve above system is as follows:

- Let $A = LL^T$ be its Cholesky decomposition.
- Using forward substitution to solve $Ly = b$.
- Using back substitution to solve $L^T x = y$.

Now, we display another method in [1] based on hyperbolic rotations.

Definition

A 2-by-2 real matrix A is called a hyperbolic rotations if A is of the form

$$\begin{bmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{bmatrix},$$

for some $\phi \in \mathbb{R}$, where $\cosh\phi = \dfrac{e^{\phi}+e^{-\phi}}{2}$ and $\sinh\phi = \dfrac{e^{\phi}-e^{-\phi}}{2}$.

Note that $\cosh\phi^2 - \sinh\phi^2 = 1$ for any $\phi \in \mathbb{R}$.

Idea to solve the Cholesky decomposition based on the hyperbolic rotations.

- Let R_0 and S_0 be upper and strict upper triangular so that $A = R_0^T R_0 - S_0^T S_0$.

  Find Q_k be a (2n)-by-(2n) matrix and define R_{k+1}, S_{k+1} by

$$\begin{bmatrix} \tilde{R}_{k+1} \\ \tilde{S}_{k+1} \end{bmatrix} = Q_k \begin{bmatrix} R_k \\ S_k \end{bmatrix}$$

  so that $\tilde{R}_{k+1}, \tilde{S}_{k+1}$ are upper and strict upper triangular, $A = \tilde{R}_{k+1}^T \tilde{R}_{k+1} - \tilde{S}_{k+1}^T \tilde{S}_{k+1}$, and first k rows of S are zeros.

- Define $\begin{bmatrix} R_{k+1} \\ S_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} \tilde{R}_{k+1} \\ \tilde{S}_{k+1} \end{bmatrix}$, where P is a permutaion matrix of the form

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

  i.e,, $R_{k+1} = \tilde{R}_{k+1}$ and S_{k+1} shift down all rows of $\tilde{S}_{k+1}$.

- Then $R_n = L^T$ is the desired result.

- Define $\begin{bmatrix} R_{k+1} \\ S_{k+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} \tilde{R}_{k+1} \\ \tilde{S}_{k+1} \end{bmatrix}$, where P is a permutaion matrix of the form

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

i.e,, $R_{k+1} = \tilde{R}_{k+1}$ and S_{k+1} shift down all rows of $\tilde{S}_{k+1}$.

- Then $R_n = L^T$ is the desired result.

What is R_0 and S_0?

Let $A = \sum_{i=1}^{n} A^{(k)}$, where $a_{ij}^{(k)} = a_{ij}$ if $i = k, j \geq i$ or $j = k, i \geq j$.

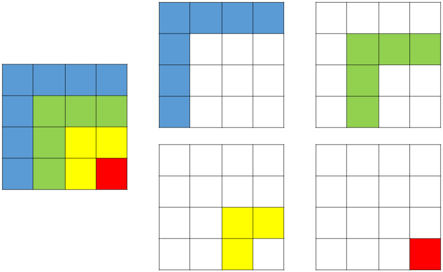

Figure: Stick-wise decomposition

What is R_0 and S_0?

Then $A^{(k)} = v_k^T v_k - w_k^T w_k$, where

$$v_{kj} = \begin{cases} a_k k^{-1/2} a_{kj} & \text{, if } j \geq k \\ 0 & \text{, otherwise} \end{cases}, w_{kj} = \begin{cases} v_{kj} & \text{, if } j! = k \\ 0 & \text{, otherwise} \end{cases}$$

What is R_0 and S_0?

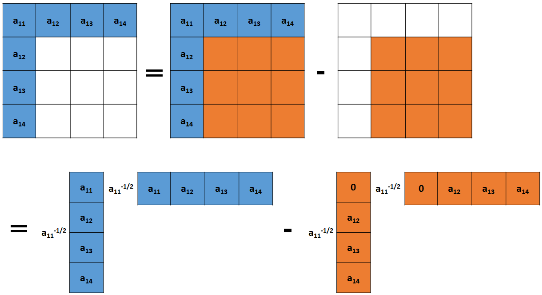

Figure: Write $A^{(1)}$ into a difference of outer product.

What is R_0 and S_0?

Let $R_0 = \begin{bmatrix} v_1 \\ v_2 \\ \vdots v_n \end{bmatrix}$ and $S_0 = \begin{bmatrix} w_1 \\ w_2 \\ \vdots w_n \end{bmatrix}$, then $A = R_0^T R_0 - S_0^T S_0$.

What is Q_k?

Let $\rho_i^{(k)} = s_{ii}/r_{ii}$, and $\cosh\phi_i = \dfrac{1}{\sqrt{1-{\rho_i^{(k)}}^2}}$, $\sinh\phi_i = \dfrac{-\rho_i^{(k)}}{\sqrt{1-{\rho_i^{(k)}}^2}}$.

Then Q_k is of the form

Then Q_k is of the form

$$\left[\begin{array}{cccc:cccc}
\cosh\phi_1 & 0 & \cdots & 0 & \sinh\phi_1 & 0 & \cdots & 0 \\
0 & \cosh\phi_2 & \cdots & 0 & 0 & \sinh\phi_2 & \cdots & 0 \\
0 & 0 & \ddots & \vdots & 0 & 0 & \ddots & \cdots \\
0 & 0 & \cdots & \cosh\phi_n & 0 & 0 & \cdots & \sinh\phi_n \\
\hdashline
\sinh\phi_1 & 0 & \cdots & 0 & \cosh\phi_1 & 0 & \cdots & 0 \\
0 & \sinh\phi_2 & \cdots & 0 & 0 & \cosh\phi_2 & \cdots & 0 \\
0 & 0 & \ddots & \vdots & 0 & 0 & \ddots & \cdots \\
0 & 0 & \cdots & \sinh\phi_n & 0 & 0 & \cdots & \cosh\phi_n
\end{array}\right]$$

$$Q_k \left[\begin{array}{ccccc} r_{11} & & & & \\ & r_{22} & & * & \\ & & r_{33} & & \\ & & & \ddots & \\ & & & & r_{nn} \\ \hdashline & & 0_{k-1,n} & & \\ & & s_{kk} & & \\ & & & \ddots & * \\ & & & & s_{nn} \end{array}\right] = \left[\begin{array}{cccc} \tilde{r}_{11} & & & \\ & \ddots & * & \\ & & \ddots & \\ & & & \tilde{r}_{nn} \\ \hdashline & & 0_{k-1,n} & \\ & 0 & \tilde{s}_{k.k+1} & * \\ & & \ddots & \tilde{s}_{n-1.n} \\ & & & 0 \end{array}\right]$$

$$\begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_k \begin{bmatrix} R_k \\ S_k \end{bmatrix} = \left[\begin{array}{ccccc} \tilde{r}_{11} & & & & \\ & \ddots & & * & \\ & & \tilde{r}_{kk} & & \\ & & & \ddots & \\ & & & & \tilde{r}_{nn} \\ \hline & & 0_{k,n} & & \\ & & \tilde{s}_{k.k+1} & & \\ & & & \ddots & * \\ & & & & \tilde{s}_{n-1.n} \end{array}\right] = \begin{bmatrix} R_{k+1} \\ S_{k+1} \end{bmatrix}$$

- Apply above procedure n times, we have $S_n = 0$ and $A = R_n^T R_n$. Hence, $L = R_n^T$ is the desired result.
- Moreover, $Q = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_n \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_{n-1} \cdots \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_0$ satisfying
$$Q \begin{bmatrix} R_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} R_n \\ 0 \end{bmatrix}.$$
- Note that $\begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix}$ and each Q_k are sparse matrix.
- However, Q is dense with $2n(n-1)+2$ nonzeros

- Apply above procedure n times, we have $S_n = 0$ and $A = R_n^T R_n$. Hence, $L = R_n^T$ is the desired result.
- Moreover, $Q = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_n \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_{n-1} \cdots \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_0$ satisfying

$$Q \begin{bmatrix} R_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} R_n \\ 0 \end{bmatrix}.$$

- Note that $\begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix}$ and each Q_k are sparse matrix.
- However, Q is dense with $2n(n-1)+2$ nonzeros

- Apply above procedure n times, we have $S_n = 0$ and $A = R_n^T R_n$. Hence, $L = R_n^T$ is the desired result.
- Moreover, $Q = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_n \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_{n-1} \cdots \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_0$ satisfying
$$Q \begin{bmatrix} R_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} R_n \\ 0 \end{bmatrix}.$$
- Note that $\begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix}$ and each Q_k are sparse matrix.
- However, Q is dense with $2n(n-1) + 2$ nonzeros

- Apply above procedure n times, we have $S_n = 0$ and $A = R_n^T R_n$. Hence, $L = R_n^T$ is the desired result.
- Moreover, $Q = \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_n \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_{n-1} \cdots \begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix} Q_0$ satisfying

$$Q \begin{bmatrix} R_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} R_n \\ 0 \end{bmatrix}.$$

- Note that $\begin{bmatrix} I & 0 \\ 0 & P \end{bmatrix}$ and each Q_k are sparse matrix.
- However, Q is dense with $2n(n-1)+2$ nonzeros

- Let $\Delta = \begin{bmatrix} D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}$, where D is the square of diagonal element of A, then $Q\Delta \begin{bmatrix} b \\ b \end{bmatrix} = \begin{bmatrix} U^{-T}b \\ L^{-T}b \end{bmatrix}$
- Further, the solution of $Ax = b$ can be obtained by

$$x = \begin{bmatrix} I & I \end{bmatrix} \Delta Q^T \begin{bmatrix} \alpha U^{-T}b \\ (1-\alpha)L^{-T}b \end{bmatrix}$$

for any $\alpha \in \mathbb{R}$. (!!)

- Let $\Delta = \begin{bmatrix} D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}$, where D is the square of diagonal element of A, then $Q\Delta \begin{bmatrix} b \\ b \end{bmatrix} = \begin{bmatrix} U^{-T}b \\ L^{-T}b \end{bmatrix}$
- Further, the solution of $Ax = b$ can be obtained by

$$x = \begin{bmatrix} I & I \end{bmatrix} \Delta Q^T \begin{bmatrix} \alpha U^{-T} b \\ (1-\alpha) L^{-T} b \end{bmatrix}$$

for any $\alpha \in \mathbb{R}$. (!!)

- With hyperbolic Cholesky algorithm, we have the following flow:

$$
\begin{array}{ccccccc}
 & & & b & & & \\
 & & & \downarrow & & & \\
A & \rightarrow & R_0, S_0 & \rightarrow & \left\{\rho^{(k)}, Q_k, R_{k+1}, S_{k+1}\right\}_{0\leq k\leq n-1} & \rightarrow & L = R_n \\
 & & & \downarrow & & & \\
 & & & Q & & & \\
 & & & \downarrow & & & \\
 & & & x = A^{-1}b & & &
\end{array}
$$

- We solve the Cholesky decomposition of a spd matrix A of size n by both standard and hyperbolic way.
- n is ranged from 2 to 128.

- We solve the Cholesky decomposition of a spd matrix A of size n by both standard and hyperbolic way.
- n is ranged from 2 to 128.

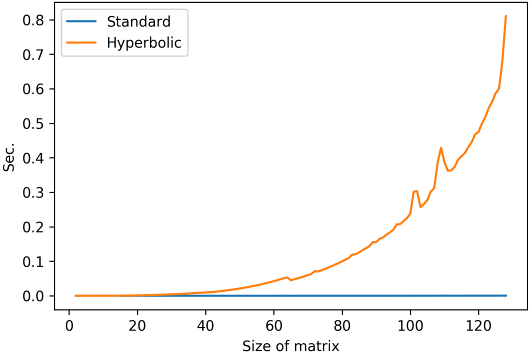

Figure: Time Comparison

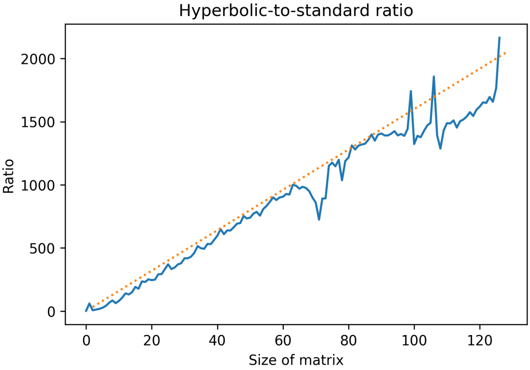

Figure: Hyperbolic-to-standard ratio (dashed line y = 16x)

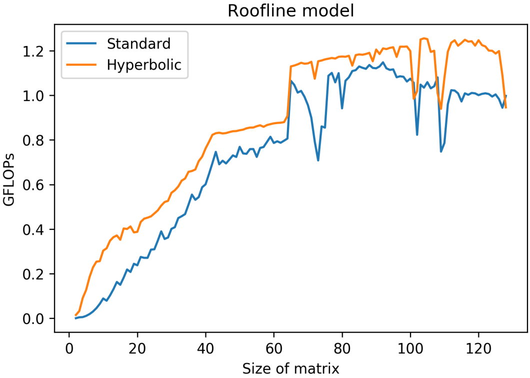

Figure: Roofline model

- In conclusion, computational speed basically is a **DISASTER**.
- Time complexity of Cholesky decomposition is $O(n^3)$ while hyperbolic version is around $O(n^4)$ before optimized.
- Luckily, there are some computation can be reduced by either parallel computing or SpMV/SpMM calculation.
- ... however, this part is not fully finished yet, still working on.

- In conclusion, computational speed basically is a **DISASTER**.
- Time complexity of Cholesky decomposition is $O(n^3)$ while hyperbolic version is around $O(n^4)$ before optimized.
- Luckily, there are some computation can be reduced by either parallel computing or SpMV/SpMM calculation.
- ... however, this part is not fully finished yet, still working on.

- In conclusion, computational speed basically is a **DISASTER**.
- Time complexity of Cholesky decomposition is $O(n^3)$ while hyperbolic version is around $O(n^4)$ before optimized.
- Luckily, there are some computation can be reduced by either parallel computing or SpMV/SpMM calculation.
- ... however, this part is not fully finished yet, still working on.

- In conclusion, computational speed basically is a **DISASTER**.
- Time complexity of Cholesky decomposition is $O(n^3)$ while hyperbolic version is around $O(n^4)$ before optimized.
- Luckily, there are some computation can be reduced by either parallel computing or SpMV/SpMM calculation.
- ... however, this part is not fully finished yet, still working on.

Thanks for your attention

$BC\ddot{E}$

Reference

[1] Delosme, J. M., & Ipsen, I. C. (1986). Parallel solution of symmetric positive definite systems with hyperbolic rotations. *Linear Algebra and its applications*, 77, 75-111.

[2] van der Vorst, H., & Van Dooren, P. (Eds.). (2014). *Parallel algorithms for numerical linear algebra* (Vol. 1). Elsevier. ISO 690