ASSORTED ALGORITHMS AMENABLE TO GPU COMPUTING: BACKGROUND FOR THE R PACKAGE, `gputools`

Linear algebraic algorithms are particularly amenable to GPU computing because they involve a high volume of simple arithmetic.

I will review:

1. the QR decomposition
2. the singular value decomposition
3. the Cholesky decomposition

I will also touch on the following algorithms found in the `gputools` R package:

4. Hierarchical Clustering for Vectors
5. FastICA algorithm of Aapo Hyvarinen et al.
6. Granger Causality Tests for Vectors

1. THE QR DECOMPOSITION

Theorem: Let A be an $m \times n$ matrix with linearly independent columns. Then:

$$A = QR$$

where:

- Q is an $m \times n$ matrix whose columns form an orthonormal basis for the column space of A.
- R is an $n \times n$ upper triangular (and therefore invertible) matrix with all positive entries on the diagonal.

EXAMPLE

$$A = \begin{bmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{bmatrix}$$

The columns of A are linearly independent. Therefore, A has a QR decomposition:

$$A = QR$$

The following choices for Q and R work:

$$Q = \frac{1}{6} \cdot \begin{bmatrix} 5 & -1 \\ 1 & 5 \\ -3 & 1 \\ 1 & 3 \end{bmatrix} \qquad R = \begin{bmatrix} 6 & 12 \\ 0 & 6 \end{bmatrix}$$

FINDING THE QR DECOMPOSITION USING ORTHOGONALITY AND THE GRAHAM SCHMIDT PROCESS

Orthogonal: two vectors u and v are orthogonal if $u \bullet v = 0$.

Note: If u and v are two component vectors in $\mathbb{R}^n$, $u \bullet v = 0$ iff they form a right angle. In Euclidian space, orthogonality is the same as perpendicularity.

Here "$\bullet$" denotes the dot product (or inner product) of two component vectors:

If $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$, then:

$$a \bullet b = (a_1 \cdot b_1,\ a_2 \cdot b_2,\ \ldots,\ a_n \cdot b_n)$$

THE GRAHAM SCHMIDT PROCESS

To find the QR decomposition of $A = [a_1, \ldots, a_n]$ (where $a_1, \ldots, a_n$ are linearly independent), you first want to find an orthogonal basis for the column space of A.

That's where the Graham Schmidt process comes in. The Graham Schmidt process is the construction one such basis, $\{v_1, v_2, \ldots, v_n\}$, by:

$$
\begin{aligned}
&v_1 = a_1 \\
&v_2 = a_2 - \frac{a_2 \bullet v_1}{v_1 \bullet v_1} \cdot v_1 \\
&v_3 = a_3 - \frac{a_3 \bullet v_1}{v_1 \bullet v_1} \cdot v_1 - \frac{a_3 \bullet v_2}{v_2 \bullet v_2} \cdot v_2 \\
&\vdots \\
&v_n = a_n - \frac{a_n \bullet v_1}{v_1 \bullet v_1} \cdot v_1 - \frac{a_n \bullet v_2}{v_2 \bullet v_2} \cdot v_2 - \cdots - \frac{a_n \bullet v_{n-1}}{v_{n-1} \bullet v_{n-1}} \cdot v_{n-1}
\end{aligned}
$$

Even better, you can get an ortho**normal** basis, $\{u_1,\ u_2, \ldots,\ u_n\}$, by:

$$
\left\{ u_1 = \frac{v_1}{v_1 \bullet v_1},\ u_2 = \frac{v_2}{v_2 \bullet v_2},\ \ldots,\ u_n = \frac{v_n}{v_n \bullet v_n} \right\}
$$

REARRANGING THAT LONG LIST OF EQUATIONS...

$$a_1 = (a_1 \bullet u_1) \cdot u_1$$

$$a_2 = (a_2 \bullet u_1) \cdot u_1 + (a_2 \bullet u_2) \cdot u_2$$

$$\begin{aligned}
&a_3 = (a_3 \bullet u_1) \cdot u_1 + (a_3 \bullet u_2) \cdot u_2 + (a_3 \bullet u_3) \cdot u_3 \\
&\vdots \\
&a_n = (a_n \bullet u_1) \cdot u_1 + (a_n \bullet u_2) \cdot u_2 + (a_3 \bullet u_3) \cdot u_3 + \cdots + (a_n \bullet u_n) \cdot u_n
\end{aligned}$$

We can put it all in matrix form:

$$A = [a_1, \ldots, a_n] \qquad Q = [u_1, \ldots, u_n] \qquad R = \begin{bmatrix} (a_1 \bullet u_1) & (a_2 \bullet u_1) & (a_3 \bullet u_1) & \cdots \\ 0 & (a_2 \bullet u_2) & (a_3 \bullet u_2) & \cdots \\ 0 & 0 & (a_3 \bullet u_3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

2. SINGULAR VALUE DECOMPOSITION

Theorem: Let A be an $m \times n$ matrix with rank r. Then, there exist:

- An $r \times r$ diagonal matrix D with the largest r nonzero **singular values** of A on the diagonal.
- An $m \times n$ matrix, $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$
- An $m \times m$ orthonormal matrix, U
- An $n \times n$ orthonormal matrix, V

Such that:

$$A = U\Sigma V^T$$

HOW DO I FIND THE SVD OF A MATRIX?

To answer these questions, I first need to review:

a. eigenvalues and eigenvectors

b. eigenvector bases

c. diagonalization

d. orthogonal diagonalizations

e. singular values

a. EIGENVALUES AND EIGENVECTORS

Let A be an $n \times n$ square matrix. Let λ be a scalar and x be a nonzero vector such that:

$$Ax = \lambda x$$

Then:

- λ is an eigenvalue of A.
- x is an eigenvector of A corresponding to λ.

FINDING EIGENDATA

1. *Finding eigenvalues:* The eigenvalues of A are exactly the solutions λ to the characteristic equation:

$$\det(A - \lambda I) = 0$$

   NOTE: Since the characteristic equation is a polynomial of degree n, every square matrix is guaranteed to have n complex eigenvalues and therefore n complex eigenvectors.

2. *Finding eigenvectors:* To find the space of eigenvectors corresponding to an eigenvalue λ, solve for x:

$$(A - \lambda I)x = 0$$

   And pick any basis you want for the solution space.

FUN FACTS ABOUT EIGENDATA

- Eigenvectors that correspond to different eigenvalues are linearly independent.
- Eigenvectors **of a symmetric matrix** that correspond to distinct eigenvalues are orthogonal.
- A is **diagonalizable** iff A has n linearly independent eigenvectors.

DIAGONALIZATION

Denote the eigenvectors of A by $x_1, \ldots, x_n$, and let them correspond to eigenvalues $\lambda_1, \ldots, \lambda_n$, respectively. We have the equations:

$$\begin{aligned} Ax_1 &= x_1\lambda_1 \\ Ax_2 &= x_2\lambda_2 \\ &\vdots \\ Ax_n &= x_n\lambda_n \end{aligned}$$

Which we can put in matrix form:

$$A \underbrace{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{P} = \underbrace{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{D}$$

$$AP = PD$$

And if P is invertible:

$$A = PDP^{-1}$$

Recap: if the eigenvector matrix $P = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ is invertible, then we can write:

$$A = PDP^{-1}$$

And we say that:

- A is **diagonalizable**.
- PDP^{-1} is the **diagonalization** of A.

P is invertible (and therefore is A diagonalizable) whenever:

- The columns of P are linearly independent.
- All the eigenvectors of A are linearly independent.
- A has n distinct eigenvalues.

d. ORTHOGONAL DIAGONALIZATIONS

An orthogonal diagonalization of A is a diagonalization, PDP^{-1} such that P is an orthogonal matrix (that is, all the columns of P are mutually orthogonal.)

Recall: for symmetric matrices, eigenvectors with distinct eigenvalues are orthogonal.

Hence, any $n \times n$ symmetric matrix with n distinct eigenvalues has an orthogonal diagonalization.

e. SINGULAR VALUES

Let A be an $m \times n$ matrix. Let:

$$\gamma_1, \gamma_2, \ldots, \gamma_n$$

be the eigenvalues of $A^T A$. Then the singular values A are the square roots, $\sigma_1, \sigma_2, \ldots, \sigma_n$, of the eigenvalues of A:

$$\sigma_1 = \sqrt{\gamma_1},\ \sigma_2 = \sqrt{\gamma_2},\ \ldots,\ \sigma_n = \sqrt{\gamma_n}$$

BACK TO THE SINGULAR VALUE DECOMPOSITION

Theorem: Let A be an $m \times n$ matrix with rank r. Then, there exist:

- An $r \times r$ diagonal matrix D with the largest r nonzero **singular values** of A on the diagonal.
- An $m \times n$ matrix, $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$
- An $m \times m$ orthonormal matrix, U
- An $n \times n$ orthonormal matrix, V

Such that:

$$A = U \Sigma V^T$$

HOW TO FIND THE SVD OF AN $m \times n$ RANK r MATRIX A

1. Find the n eigenvalues $\gamma_1, \gamma_2, \ldots, \gamma_n$ of $A^T A$ (indexed in decreasing order), along with corresponding unit-length orthogonal eigenvectors, $v_1, v_2, \ldots, v_n$.
2. Let $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$
3. Let $D = \begin{bmatrix} \sqrt{\gamma_1} & & & \\ & \sqrt{\gamma_2} & & \\ & & \ddots & \\ & & & \sqrt{\gamma_r} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$
4. Construct U:
   a. Let the first r columns of U be $\frac{1}{||Av_1||}Av_1,\ \frac{1}{||Av_2||}Av_2,\ \ldots,\ \frac{1}{||Av_r||}Av_r$
   b. If need be, form the last $m - r$ columns of U by extending the above r vectors to an orthonormal basis for R^m.

EXAMPLE

Let's find the SVD of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$

1. $A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$

   By inspection, two orthogonal eigenvectors of $A^T A$ are $x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, with eigenvalues $\gamma_1 = 18$ and $\gamma_2 = 0$, respectively.

   From x_1 and x_2, I construct orthonormal eigenvectors $v_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.

2. Let $V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

3. The only nonzero singular value is $\sqrt{18} = 3\sqrt{2}$. Hence, $D = [3\sqrt{2}]$ and $\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

4. Let's construct U:

a. $Av_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}$ and $Av_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. We take $\frac{1}{||Av_1||}Av_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$ to be the first column of U.

b. To find the other two columns, we extend Av_1 to an orthonormal basis for $\mathbb{R}^3$. If a vector $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is to be in this basis, we need:

$$Av_1 \bullet x = 0$$

and hence:

$$\frac{2}{\sqrt{2}}x_1 - \frac{4}{\sqrt{2}}x_2 + \frac{4}{\sqrt{2}}x_3 = 0$$

Solving for two linearly independent solutions to the above, applying the Graham Schmidt process, and normalizing, we get the last two columns of U:

$$\begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix} \qquad \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

Hence, $U = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix}$

To summarize:

$$\underbrace{\begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}}_{V^T}$$

3. THE CHOLESKY DECOMPOSITION

Let A be an $n \times n$ matrix. Then, a Cholesky decomposition of A is:

$$A = LL^T$$

where L is an $n \times n$ invertible lower triangular matrix whose diagonals are all positive.

Note: A has a Cholesky decomposition iff A is positive definite.

The following slides explaining how to get the Cholesky decomposition are taken directly from UCLA Professor Lieven Vandenberghe's EE103 class:

Courses
EE103
EE236A
EE236B
EE236C

Research
Group
Publications
Software

Lieven Vandenberghe

Professor of Electrical Engineering and Mathematics
University of California, Los Angeles

Contact information

UCLA Electrical Engineering Department
66-147L Engineering IV
Los Angeles, CA 90095-1594

Phone: (310) 206 1259
Fax: (310) 206 4685
E-mail: vandenbe@ee.ucla.edu

Courses

- EE103: Applied numerical computing
- EE236A: Linear programming
- EE236B: Convex optimization
- EE236C: Optimization methods for large-scale systems

Cholesky factorization algorithm

partition matrices in $A = LL^T$ as

$$\begin{bmatrix} a_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} l_{11} & L_{21}^T \\ 0 & L_{22}^T \end{bmatrix}$$

$$= \begin{bmatrix} l_{11}^2 & l_{11}L_{21}^T \\ l_{11}L_{21} & L_{21}L_{21}^T + L_{22}L_{22}^T \end{bmatrix}$$

algorithm

1. determine l_{11} and L_{21}:

$$l_{11} = \sqrt{a_{11}}, \qquad L_{21} = \frac{1}{l_{11}} A_{21}$$

2. compute L_{22} from

$$A_{22} - L_{21}L_{21}^T = L_{22}L_{22}^T$$

   this is a Cholesky factorization of order $n - 1$

Example

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

- first column of L

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & l_{22} & 0 \\ -1 & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

- second column of L

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} l_{22} & 0 \\ l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{22} & l_{32} \\ 0 & l_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & l_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & l_{33} \end{bmatrix}$$

- third column of L: $10 - 1 = l_{33}^2$, *i.e.*, $l_{33} = 3$

conclusion:

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

4. Hierarchical Clustering for Vectors