Matrix Decomposition

Ming Yang
Electrical and Computer Engineering
Northwestern University
Evanston, IL 60208
mya671@ece.northwestern.edu

Contents

1. Overview **2**

2 Matrix Multiplication and Definitions **2**
2.1 Matrix Multiplication 2
2.2 Special Matrix Definition 3

3 Singular Value Decomposition **4**
3.1 SVD decomposition 4
3.2 Corollary of SVD 5

4 LU and Cholesky Decomposition **5**
4.1 Elementary Operation and Gaussian Transform 5
4.2 LU decomposition 6
4.3 Cholesky decomposition 7

5 QR Decomposition **7**
5.1 Householder Reflections and Givens Rotations 8
5.2 Gram-Schmidt orthonormalization 9
5.3 QR Decomposition 9
5.4 Least Square Fitting 10

6 Schur Decomposition and Eigenvalue Decomposition **11**
6.1 Schur Decomposition 11
6.2 Eigenvalue Decomposition 13
6.3 Hessenberg Decomposition 14

7 Biconjugate Decomposition **15**
7.1 Biconjugate Decomposition 15
7.2 Summary 16

1. Overview

“Matrix decomposition refers to the transformation of a given matrix into a given canonical form.” [1], when the given matrix is transformed to a right-hand-side product of canonical matrices the process of producing this decomposition is also called “matrix factorization”. Matrix decomposition is a fundamental theme in linear algebra and applied statistics which has both scientific and engineering significance. The purposes of matrix decomposition typically involve two aspects: computational convenience and analytic simplicity. In the real world, it is not feasible for most of the matrix computations to be calculated in an optimal explicit way, such as matrix inversion, matrix determinant, solving linear system and least square fitting, thus to convert a difficult matrix computation problem into several easier tasks such as solving triangular or diagonal system will greatly facilitate the calculations. Data matrices representing some numerical observations such as proximity matrix or correlation matrix are often huge and hard to analyze, therefore to decompose the data matrices into some lower-order or lower-rank canonical forms will reveal the inherent characteristic and structure of the matrices and help to interpret their meaning readily.

This tutorial is primarily a summary of important matrix decomposition methods, we will first present some basic concepts in Section 2 and then introduce several fundamental matrix decomposition methods in the successive sections, e.g. SVD, LU, QR and Eigen decomposition. A unified view of matrix factorization derived from the Wedderburn rank-one reduction theorem is briefly discussed in the summary Section 7.

2 Matrix Multiplication and Definitions

2.1 Matrix Multiplication

Suppose $\mathbf{A} \in \mathbb{R}^{m\times r}$ and $\mathbf{B} \in \mathbb{R}^{r\times n}$, the matrix multiplication $\mathbf{C} = \mathbf{AB}$ can be viewed from three different perspectives as follows:

Dot Product Matrix Multiply. Every element c_{ij} of $\mathbf{C}$ is the dot product of row vector a_i^T and column vector b_j.

$$
\begin{aligned}
\mathbf{A} &= \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} \qquad a_k \in \mathbb{R}^r \\
\mathbf{B} &= (b_1, \ldots, b_n) \qquad b_k \in \mathbb{R}^r \\
\mathbf{C} &= (c_{ij}) \quad c_{ij} = a_i^T b_j.
\end{aligned}
$$

Column Combination Matrix Multiply. Every column c_j of $\mathbf{C}$ is a linear combination of column vector a_k of $\mathbf{A}$ with columns b_{kj} as the weight coefficients.

$$\mathbf{A} = (a_1, \dots, a_r) \quad a_i \in \mathbb{R}^m$$
$$\mathbf{B} = (b_1, \dots, b_n) \quad b_j \in \mathbb{R}^r$$
$$\mathbf{C} = (c_1, \dots, c_n) \quad c_j \in \mathbb{R}^m$$
$$c_j = \sum_{k=1}^{r} b_{kj} a_k \quad j = 1 : n.$$

Outer Product Matrix Multiply. $\mathbf{C}$ is the sum of r matrices, every matrix is an outer product of $\mathbf{A}$'s column vector and $\mathbf{B}$'s row vector, which is a rank-one matrix.

$$\mathbf{A} = (a_1, \dots, a_r) \quad a_i \in \mathbb{R}^m$$
$$\mathbf{B} = \begin{pmatrix} b_1^T \\ \vdots \\ b_r^T \end{pmatrix} \quad b_k \in \mathbb{R}^n$$
$$\mathbf{C} = \sum_{k=1}^{r} a_k b_k^T.$$

2.2 Special Matrix Definition

Before further discussion, we first present definitions of some special matrices, here we follow the terms in [2].

Definition 1 *A real matrix* $\mathbf{A}$ *is a symmetric matrix if it equals to its own transpose, that is* $\mathbf{A} = \mathbf{A}^T$.

Definition 2 *A complex matrix* $\mathbf{A}$ *is a hermitian matrix if it equals to its own complex conjugate transpose, that is* $\mathbf{A} = \mathbf{A}^H$.

Definition 3 *A real matrix* $\mathbf{Q}$ *is an orthogonal matrix if the inverse of* $\mathbf{Q}$ *equals to the transpose of* $\mathbf{Q}$, $\mathbf{Q}^{-1} = \mathbf{Q}^T$, *that is* $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$.

Definition 4 *A complex matrix* $\mathbf{U}$ *is a unitary matrix if the inverse of* $\mathbf{U}$ *equals the complex conjugate transpose of* $\mathbf{U}$, $\mathbf{U}^{-1} = \mathbf{U}^H$, *that is* $\mathbf{U}\mathbf{U}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}$.

Definition 5 *A matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is positive definite if* $x^T\mathbf{A}x \geq 0$ *for all nonzero* $x \in \mathbb{R}^n$. *Positive definite matrices have positive definite principle sub-matrices and all the diagonal entries are positive.*

Definition 6 *Suppose* $\mathbf{S} \subseteq \mathbb{R}^n$ *be a subspace with orthonormal basis* $\mathbf{V} = (v_1, \dots, v_k)$, $\mathbf{P} = \mathbf{V}^T\mathbf{V} \in \mathbb{R}^{n \times n}$ *is the* orthogonal projection matrix *onto* $\mathbf{S}$ *such that* $range(\mathbf{P}) = \mathbf{S}$, $\mathbf{P}^2 = \mathbf{P}$, *and* $\mathbf{P}^T = \mathbf{P}$. $\mathbf{P}$ *is unique for subspace* $\mathbf{S}$.

Hermitian matrix and unitary matrix are the counterparts of symmetric and orthogonal matrix in $\mathbb{R}$, the following theorems in $\mathbb{R}$ can be readily transformed to the corresponding forms in $\mathbb{C}$ by substituting the transpose by conjugate transpose and orthogonal matrix by unitary matrix. Therefore, for simplicity, we present most of the matrix decomposition results in $\mathbb{R}$.

3 Singular Value Decomposition

Suppose matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the column vectors of $\mathbf{A}$, namely $range(\mathbf{A})$, represent a subspace in $\mathbb{R}^m$, similarly $range(\mathbf{A}^T)$ is a subspace in $\mathbb{R}^n$, apparently the two subspaces have the same dimension equals to the rank of $\mathbf{A}$. SVD decomposition is able to reveal the orthonormal basis of the $range(\mathbf{A})$ and $range(\mathbf{A}^T)$ and the respective scale factors σ_i simultaneously.

3.1 SVD decomposition

Theorem 1 Singular Value Decomposition(SVD) *If matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, then there exist orthogonal matrices $\mathbf{U} = (u_1, \ldots, u_m) \in \mathbb{R}^{m\times m}$, $\mathbf{V} = (v_1, \ldots, v_n) \in \mathbb{R}^{n\times n}$ and diagonal matrix $\Sigma = diag(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n} \quad p = \min(m, n)$, such that*

$$\mathbf{A} = \mathbf{U\Sigma V}^T, \quad \textit{where} \quad \sigma_1 \geq \sigma_2 \ldots \geq \sigma_p \geq 0.$$

Proof 1 *Let $\sigma_1 = \|\mathbf{A}\|_2 = max_{\|v\|_2=1}\|\mathbf{A}v\|_2$. Then there exist unit 2-norm vectors $u_1 \in \mathbb{R}^m$ and $v_1 \in \mathbb{R}^n$, such that*

$$\|\mathbf{A}v_1\| = \sigma_1, u_1 = \frac{\mathbf{A}v_1}{\sigma_1}, \quad \textit{therefore} \quad \mathbf{A}v_1 = \sigma_1 u_1.$$

Any orthonormal set can be extended to form an orthonormal basis for the whole space, so we can find $\mathbf{V}_1 \in \mathbb{R}^{n\times(n-1)}$ and $\mathbf{U}_1 \in \mathbb{R}^{m\times(m-1)}$, such that $\mathbf{V} = (v_1\mathbf{V}_1) \in \mathbb{R}^{n\times n}$ and $\mathbf{U} = (u_1\mathbf{U}_1) \in \mathbb{R}^{nm\times m}$ are orthonormal basis, thus

$$\mathbf{A}_1 \doteq \begin{pmatrix} u_1^T \\ \mathbf{U_1^T} \end{pmatrix} (\mathbf{A}v_1 \ \mathbf{AV}_1) = \begin{pmatrix} u_1^T\mathbf{A}v_1 & u_1^T\mathbf{AV}_1 \\ \mathbf{U}_1^T\mathbf{A}v_1 & \mathbf{U}_1^T\mathbf{AU}_1 \end{pmatrix} = \begin{pmatrix} \sigma_1\|u_1\|_2^2 & u_1^T\mathbf{AV}_1 \\ \sigma_1\mathbf{U}_1^T u_1 & \mathbf{U}_1^T\mathbf{AU}_1 \end{pmatrix} = \begin{pmatrix} \sigma_1 & u_1^T\mathbf{AV}_1 \\ 0 & \mathbf{U}_1^T\mathbf{AU}_1 \end{pmatrix}$$

Let $\begin{pmatrix}\sigma_1 & u_1^T\mathbf{AV}_1\end{pmatrix}_T = \begin{pmatrix}\sigma_1 & \omega^T\end{pmatrix} \in \mathbb{R}^n$, the 2-norm of the product with $\mathbf{A}_1$ gives:

$$\|\mathbf{A}_1 \begin{pmatrix} \sigma_1 \\ \omega \end{pmatrix}\|_2^2 = \| \begin{matrix} \sigma_1^2 + \omega^T\omega \\ \ldots \end{matrix} \|_2^2 \geq (\sigma_1^2 + \omega^T\omega)^2$$

So the 2-norm of matrix $\mathbf{A}_1$ is

$$\|\mathbf{A}_1\|_2 = \sup_{x\in\mathbb{R}^n} \frac{\|\mathbf{A}_1 x\|}{\|x\|} \geq \frac{(\sigma_1^2 + \omega^T\omega)}{\sqrt{(\sigma_1^2 + \omega^T\omega)}} = \sqrt{(\sigma_1^2 + \omega^T\omega)},$$

while $\mathbf{U}$ and $\mathbf{V}$ are both orthonomal basis and $\|\mathbf{A}_1\|_2 = \|\mathbf{A}\|_2 = \sigma_1$, so $\omega = 0$. An induction on arguments completes the proof.

The σ_i are the *singular values* of $\mathbf{A}$ and the vector u_i and v_i are the *left singular vector* and *right singular vector*, which satisfy that

$$\mathbf{A}v_i = \sigma_i u_i \quad \text{and} \quad \mathbf{A}^T u_i = \sigma_i v_i.$$

3.2 Corollary of SVD

SVD decomposition reveals many intrinsic properties of matrix $\mathbf{A}$ and is numerical stable for calculations.

Corollary 1 *If $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is a SVD of $\mathbf{A}$ with $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r \geq \sigma_{r+1} = \ldots = \sigma_p = 0$, we have the following statements:*

1. $rank(\mathbf{A}) = r$.
2. $null(\mathbf{A}) = span\{v_{r+1}, \ldots, v_n\}$.
3. $range(\mathbf{A}) = span\{u_1, \ldots, u_r\}$.
4. $\mathbf{A} = \sum_{j=1}^{r} \sigma_j u_j v_j^T = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r$, *where* $\mathbf{U}_r = (u_1, \ldots, u_r)$,$\mathbf{V}_r = (v_1, \ldots, v_r)$,$\mathbf{\Sigma}_r = (\sigma_1, \ldots, \sigma_r)$.
5. $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \ldots + \sigma_p^2}$.
6. $\|\mathbf{A}\|_2 = \sigma_1$.
7. $\sigma_j = \sqrt{\lambda_j(\mathbf{A}^T\mathbf{A})}, j = 1, \ldots, p$, *where* $\lambda_j(\mathbf{A}^T\mathbf{A})$ *is the* j*th largest eigenvalue of* $\mathbf{A}^T\mathbf{A}$.
8. v_i *are orthonormalized eigenvectors of* $\mathbf{A}^T\mathbf{A}$ *and* u_i *are orthonormalized eigenvectors of* $\mathbf{A}\mathbf{A}^T$.

SVD is generalized to simultaneously diagonalize two matrices [3] or decomposition of a matrix that employs different metrics in the normalizations [4].

4 LU and Cholesky Decomposition

Solution to the linear system equation $\mathbf{A}x = b$ is the basic problem in linear algebra. Theoretically when $\mathbf{A}$ is a non-singular square matrix there exists a unique solution $x = \mathbf{A}^{-1}b$, however the inverse of a matrix is typically not easy to compute. So we hope to transform $\mathbf{A}$ to some triangular systems which are much easier to solve by forward or backward substitution, this process is referred to as *Gaussian elimination* [5]. This process can be summarized in matrix form as LU decompostion and a series of evolutions when matrix $\mathbf{A}$ has extra properties.

4.1 Elementary Operation and Gaussian Transform

For square matrix $\mathbf{A}$, the following three operations are referred to as *elementary row (column) operations of type 1, 2, and 3* respectively:

1. interchanging two rows (or columns) in $\mathbf{A}$.
2. multiplying all elements of a row (or column) of $\mathbf{A}$ by some nonzero number.
3. adding to any row (or column) of $\mathbf{A}$ any other row (or column) of $\mathbf{A}$ multiplied by a non zero number.

These operations can be implemented by pre- or post-multiplying an appropriate matrices called *elementary matrices*, the type 3 row elementary matrices have the following forms:

$$\mathbf{E}^{(3)} = \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & \dots & \tau & & \\ & & & 1 & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{E}^{(3)} = \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & 1 & & & \\ & & \tau & \dots & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix}$$

Gaussian elimination process can be described as matrix multiplications of type 3 lower triangle elementary matrices. For $x \in \mathbb{R}^n$ with $x_k \neq 0$, *Gaussian Transformation* is defined as matrix $\mathbf{M}_k = \mathbf{I} - \tau e_k^T$, where *Gauss vector* τ is

$$\tau^T = \left(\underbrace{0, \dots, 0}_{k}, \tau_{k+1}, \dots, \tau_n\right) \quad \tau_i = \frac{x_i}{x_k} \quad i = k+1 : n \quad \text{and} \quad e_k^T = \left(0, \dots, 0, \underbrace{1}_{k\text{th}}, 0, \dots, 0\right)$$

Pre-multiply x with $\mathbf{M}_k$ then the last $k+1$ to n elements of x are zeroed.

$$\mathbf{M}_k x = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & & 1 & 0 & & 0 \\ 0 & & -\tau_{k+1} & 1 & & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & -\tau_n & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \dots \\ x_k \\ x_{k+1} \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \dots \\ x_k \\ 0 \\ \dots \\ 0 \end{pmatrix}$$

It is easy to verify that Gaussian transform matrix is the product of lower triangular type 3 elementary matrices with $\det(\mathbf{M}_k) = 1$. So by multiplying a series of Gaussian transform matrix, the lower part of $\mathbf{A}$ can be gradually zeroed given that the pivots $x_{kk} \neq 0$ during the process. This process can be summarized as LU decomposition.

4.2 LU decomposition

Theorem 2 LU Decomposition *Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *and all the leading principal minors* $\det(\mathbf{A}(1:k, 1:k)) \neq 0, k = 1, \dots, n-1$. *Then there exist a unique unit lower triangular* $\mathbf{L}$ *with diagonal elements all equal to one and a unique upper triangular matrix* $\mathbf{U}$ *such that* $\mathbf{A} = \mathbf{LU}$*, and* $\det(\mathbf{A}) = u_{11}u_{22}\dots u_{nn}$.

Proof 2 *Given* $a_{11} \neq 0$ *in* $\mathbf{A}$*, we can find Gaussian transform* $\mathbf{M}_1$ *to zero the* $a_{21}, \dots, a_{n1}$. *Suppose at* $k-1$ *step* $\mathbf{M}_{k-1} \dots \mathbf{M}_1\mathbf{A} = \mathbf{A}^{(k-1)}$*, consider the* $k \times k$ *portion of this equation, since Gaussian transforms are unit lower triangular with determinants equal to one,* $\det(\mathbf{A}(1:k, 1:k) = a_{11}^{(k-1)} \dots a_{kk}^{(k-1)} \neq 0$. *Therefore the* k*th pivot* $a_{kk}^{(k-1)} \neq 0$*, we can proceed to find Gaussian transform* $\mathbf{M}_k$.

If $\mathbf{A} = \mathbf{L}_1\mathbf{U}_1 = \mathbf{L}_2\mathbf{U}_2$ *are two LU decompositions of a non-singular* $\mathbf{A}$*, then* $\mathbf{L}_2^{-1}\mathbf{L}_1 = \mathbf{U}_2\mathbf{U}_1^{-1}$*, since the left part of the equation is unit lower triangular while the right side is upper triangular, both of the matrices must be the identity to satisfy the equation. Hence,* $\mathbf{L}_1 = \mathbf{L}_2$ *and* $\mathbf{U}_1 = \mathbf{U}_2$*.*

If $\mathbf{A} = \mathbf{LU}$ *then* $\det(\mathbf{A}) = \det(\mathbf{LU}) = \det(\mathbf{L})\det(\mathbf{U}) = u_{11}u_{22}\dots u_{nn}$*.*

For linear system $\mathbf{A}x = b$ if we pre-compute the LU decomposition of $\mathbf{A} = \mathbf{LU}$, the problem reduces to solve two triangle systems $\mathbf{L}y = b$ and $\mathbf{U}x = y$ which can be calculated much more readily. Moreover when the system has to be solved with respect to many different b, such as the solution of certain circuit under different excitations, the LU decomposition method is very efficient.

4.3 Cholesky decomposition

If the matrix $\mathbf{A}$ has additional properties, the LU decomposition will have particular forms. In this section we will present the specilized LU decomposition for symmetric and positive definite matrices. First we express the LU decomposition in an equavalent way.

Theorem 3 $\mathbf{LDM}^T$ **Decomposition**. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *and all the leading principal minors* $\det(\mathbf{A}(1:k, 1:k)) \neq 0, k = 1, \dots, n-1$*. Then there exist unique unit lower triangular matrices* $\mathbf{L}$ *and* $\mathbf{M}$ *and a unique diagonal matrix* $\mathbf{D} = diag(d_1, \dots, d_n)$*, such that* $\mathbf{A} = \mathbf{LDM}^T$*.*

If $\mathbf{A}$ has a LU decomposition $\mathbf{A} = \mathbf{LU}$ and let $\mathbf{D} = diag(u_{11}, \dots, u_{nn})$, observe that $\mathbf{M}^T = \mathbf{D}^{-1}\mathbf{U}$ is unit upper triangular. Thus $\mathbf{A} = \mathbf{LU} = \mathbf{LD}(\mathbf{D}^{-1}\mathbf{U}) = \mathbf{LDM}^T$. Uniqueness follows from the uniqueness of LU decomposition.

Further if $\mathbf{A}$ is symmetric, $\mathbf{A} = \mathbf{LDM}^T = \mathbf{A}^T = \mathbf{MDL}^T$, $\mathbf{MDL}^T$ is also the $\mathbf{LDM}^T$ Decomposition of $\mathbf{A}$. From the uniqueness we have $\mathbf{L} = \mathbf{M}$.

Theorem 4 $\mathbf{LDL}^T$ **Decomposition**. *If* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is a non-singular symmetric matrix and has* $\mathbf{LDM}^T$ *Decomposition* $\mathbf{A} = \mathbf{LDM}^T$*, then* $\mathbf{L} = \mathbf{M}$ *and* $\mathbf{A} = \mathbf{LDL}^T$*.*

For a positive definite matrix $\mathbf{A}$, the $\mathbf{D} = diag(d_1, \dots, d_n)$ in $\mathbf{LDM}^T$ decomposition has positive diagonal entries. So we can further specilize the LU decomposition for symmetric positive definite matrices.

Theorem 5 Cholesky Decomposition. *If* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *is symmetric positive definite, then there exists a unique lower triangular* $\mathbf{G} \in \mathbb{R}^{n\times n}$ *with positive diagonal entries such that* $\mathbf{A} = \mathbf{GG}^T$*, and* $\mathbf{G}$ *is referred to the* Cholesky triangle.

In the $\mathbf{LDL}^T$ decomposition of symmetric $\mathbf{A}$, the entries of the diagonal matrix $\mathbf{D} = diag(d_1, \dots, d_n)$ are all positive, so let $\mathbf{G} = \mathbf{L}diag(\sqrt{d_1}, \dots, \sqrt{d_n})$ is a lower triangular with positive diagonal entries and $\mathbf{A} = \mathbf{GG}^T$, the uniqueness follows from the uniqueness of the $\mathbf{LDL}^T$ decomposition.

5 QR Decomposition

If the linear system $\mathbf{A}x = b$ is overdetermined, namely, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m \geq n$ and $b \in \mathbb{R}^m$, the exact solution may not exist. So we can use the least square solution of the minimization $\|\mathbf{A}x - b\|_2$ as a substitution. In this section we will present several methods to construct the QR decomposition and how to compute the least square fitting by QR, LU and SVD decomposition.

5.1 Householder Reflections and Givens Rotations

Let $v \in \mathbb{R}^n$ be nonzero, a n-by-n matrix **P** of the form

$$\mathbf{P} = \mathbf{I} - 2vv^T/v^Tv,$$

is called a *Householder reflection* or *Householder matrix* or *Householder transformation*. The vector v is called a *Householder vector*. A geometric illustration: when a vector x is multiplied by $\mathbf{P}$, it is reflected with the hyperplane of v's orthogonal complement $span\{v\}^\perp$.

Householder reflections can be used to zero selected entries of a vector in similar way to Gauss transformations, while Gauss transformations are unit lower triangular and Householder matrices are orthogonal and symmetric:

$$\begin{aligned}\mathbf{PP}^T = \mathbf{PP} &= (\mathbf{I} - 2vv^T/v^Tv)(\mathbf{I} - 2vv^T/v^Tv) \\ &= \mathbf{I} - 4vv^T/v^Tv + 4vv^Tvv^T/(v^Tv)^2 = \mathbf{I}.\end{aligned}$$

Given a vector $0 \neq x \in \mathbb{R}^n$, we will show there exists a Householder reflection can zero the all but the first elements in x, such that $\mathbf{P}x \in span\{e_1\}$. Let $v = x + \alpha e_1$ and $\alpha = \|x\|_2$,

$$\begin{aligned}\mathbf{P}x &= x - \frac{2v^Tx}{v^Tv}v \\ &= x - \frac{2(x^Tx+\alpha x_1)}{x^Tx+2\alpha x_1+\alpha^2}(x + \alpha e_1) \\ &= x - \frac{2(\alpha^2+\alpha x_1)}{2\alpha^2+2\alpha x_1}(x + \alpha e_1) = \alpha e_1\end{aligned}$$

Furthermore, to zero all the elements except the first two entries of vector $x = (x_1, x_2, \ldots, x_n)^T$, we can obtain the Household vector $v' = x' + \|x'\|e_1$ where $x' = (x_2, \ldots, x_n)^T$ and extend $v = (0, v')^T$. So on so forth we can apply a series of Householder reflections to reduce matrix $\mathbf{A}$ to a upper triangular matrix.

Household reflections are capable of introducing zeros to all but the first element of a vector, *Givens rotations* are able to selectively zero one element. For vector $x = (x_1, \ldots, x_i, \ldots, x_k, \ldots, x_n)^T, x_k \neq 0$, the following Givens rotation can force x_k to be zero:

$$\mathbf{G}(i,k,\theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

where $c = \cos(\theta) = \frac{x_i}{\sqrt{x_i^2+x_k^2}}$ and $s = \sin(\theta) = \frac{-x_k}{\sqrt{x_i^2+x_k^2}}$. In a geometric view, Givens rotation amounts to a counterclockwise rotation θ in (i,k) coordinate plane. It is easy to check $\mathbf{G}(i,k,\theta)$ is also orthogonal, and by the pre-multiplication of a series of Givens rotations we can zero the lower part of a matrix.

5.2 Gram-Schmidt orthonormalization

If $\mathbf{A} = (a_1, \ldots, a_n) \in \mathbb{R}^{m\times n}$ is a linear independent set of vectors, by subtracting from the the projections of a_k onto a_i $(i < k)$ from a_k and adequate normalization, we can gradually orthonormalize $\mathbf{A}$ to an orthonomal set $\mathbf{Q} = (q_1, \ldots, q_m)$ as follows:

$$\begin{array}{lcll} q_1 & = & a_1/r_{11} & r_{11} = \|a_1\| \\ q_2 & = & (a_2 - r_{21}q_1)/r_{22} & r_{21} = x_2^T q_1, r_{22} = \|a_2 - r_{21}q_1\|_2 \\ q_3 & = & (a_3 - r_{31}q_1 - r_{32}q_2)/r_{33} & r_{31} = x_3^T q_1, r_{32} = x_3^T q_2, r_{33} = \|a_3 - r_{31}q_1 - r_{32}q_2\|_2 \\ \vdots & = & \vdots & \vdots \\ q_k & = & (a_k - \sum_{i=1}^{k-1} r_{ik}q_i)/r_{kk}) & r_{ik} = q_i^T a_k, r_{kk} = \|a_k - \sum_{i=1}^{k-1} r_{ik}q_i\|_2 \end{array}$$

Therefore, $a_k = \sum_{i=1}^{k} r_{ik}q_i$, $\mathbf{A}$ is the product of $\mathbf{Q}$ and an upper triangular $\mathbf{R} = (r_{ij})$, this process is called the *Gram-Schmidt orthonormalization process*.

This process is sensitive to roundoff errors. A modified version of Gram-Schmidt process subtracts the projections onto q_k of all the succeeding a_i from a_i instead of subtract from a_i all the previous q_k. When q_k is determined we first subtract the projection of a_i onto q_k from a_i for $i > k$ and then normalize the new a_{k+1} to get q_{k+1}:

$$\begin{array}{ll} \text{for k=1 to n} & \\ q_k = a_k^k/r_{kk} & , r_{kk} = \|a_k\|_2 \\ a_i^{k+1} = a_i^k - r_{ki}q_k & , r_{ki} = a_i^{kT} q_k, \quad \text{for} \quad i = k+1 : n \end{array}$$

After $\mathbf{Q}$ is calculated, by sequentially substituting a_k^k with the previous $a_k^i, i < k$ we can easily get the representation a_k^1 namely a_k with respect to $q_1, \ldots, q_k$:

$$\begin{array}{rcl} q_k & = & a_k^k/r_{kk} = \frac{1}{r_{kk}}(a_k^{k-1} - r_{k-1\ k}q_{k-1}) \\ & = & \frac{1}{r_{kk}}(a_k^{k-2} - r_{k-2\ k}q_{k-2} - r_{k-1\ k}q_{k-1}) \\ & = & \vdots \\ & = & \frac{1}{r_{kk}}(a_k^1 - \sum_{i=1}^{k-1} r_{ik}q_i). \end{array}$$

Thus $a_k = \sum_{i=1}^{k} r_{ik}q_i$, which implies $\mathbf{A}$ is the product of $\mathbf{Q}$ and an upper triangular $\mathbf{R} = (r_{ij})$.

5.3 QR Decomposition

Theorem 6 QR Decomposition. *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, there exist an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{m\times m}$ and an upper triangular matrix $\mathbf{R} \in \mathbb{R}^{m\times n}$, such that*

$$\mathbf{A} = \mathbf{QR}$$

All the methods in the previous sub-sections can be viewed as different constructive proofs of QR decomposition, including Householder reflection, Givens rotation and Gram-Schimdt orthogonalization process and its modification version.

Corollary 2 *If $\mathbf{A} \in \mathbb{R}^{m\times n}$ has full column rank $m \geq n$ and $\mathbf{A} = \mathbf{QR}$ is a QR decomposition. $\mathbf{A} = (a_1, \ldots, a_n)$ and $\mathbf{Q} = (q_1, \ldots, q_m)$ are column partition forms, then*

$$\begin{array}{ll} span\{a_1, \ldots, a_k\} = span\{q_1, \ldots, q_k\} & k = 1 : n \\ range(A) = span\{q_1, \ldots, q_n\} & \\ range(A)^{\perp} = span\{q_{n+1}, \ldots, q_m\} & \end{array}.$$

Let $\mathbf{Q}_1 = (q_1, \ldots, q_n)$, $\mathbf{A} = \mathbf{Q}_1\mathbf{R}_1$ with $\mathbf{R}_1 \in \mathbb{R}^{n\times n}$, then $\mathbf{G} = \mathbf{R}_1^T$ is the lower triangular Cholesky factor of $\mathbf{A}^T\mathbf{A}$.

The first part of the corollary can be easily proved by the Gram-Schmidt process.
$\mathbf{A}^T\mathbf{A} = (\mathbf{Q}_1\mathbf{R}_1)^T\mathbf{Q}_1\mathbf{R}_1 = \mathbf{R}_1^T\mathbf{R}_1 = \mathbf{G}^T\mathbf{G}$, so $\mathbf{R}_1$ is unique upper triangular with positive diagonal entries.

5.4 Least Square Fitting

Let's consider the overdetermined system $\mathbf{A}x = b$, where the *data matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ and the *observation vector* $b \in \mathbb{R}^m$ with $m \geq n$, typically the system has no exact solution if b is not an element of $range(\mathbf{A})$. The goal of least square fitting problem is to find $x \in \mathbb{R}^n$ to minimize $J = \|\mathbf{A}x - b\|_p$, where $p = 2$ the optimization function is analytic.

If $\mathbf{A}$ doesn't have full column rank the solution to the LS fitting is not unique, if x minimizes the J then $x + z, z \in null(\mathbf{A})$ is also a solution. Assume $\mathbf{A}$ has full column rank the unique LS solution x_{LS} is give by the pseudo inverse $x_{LS} = \mathbf{A}^{\dagger}b = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^Tb$. As we mentioned before the inverse of a matrix is typically not easy to compute, the normal equation $\mathbf{A}^T\mathbf{A}x_{LS} = \mathbf{A}^Tb$ is more practical to be solved by QR decomposition, LU decomposition or SVD decomposition.

Note that the 2-norm of a vector is invariant under orthogonal transformation. Suppose $\mathbf{A} = \mathbf{QR}$ is the QR decomposition, we can get

$$\mathbf{Q}^T\mathbf{A} = \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ 0 \end{pmatrix} \begin{array}{c} n \\ m-n \end{array},$$

where $\mathbf{R}_1$ is square upper triangular, and

$$\mathbf{Q}^Tb = \begin{pmatrix} c \\ d \end{pmatrix} \begin{array}{c} n \\ m-n \end{array},$$

then

$$\|\mathbf{A}x - b\|_2^2 = \|\mathbf{Q}^T\mathbf{A}x - \mathbf{Q}^Tb\|_2^2 = \|\mathbf{R}_1x - c\|_2^2 + \|d\|_2^2.$$

Thus $x_{LS} \in \mathbb{R}^n$ can be readily solved with back substitution of the upper triangular system $\mathbf{R}_1x = c$. Once the QR decomposition of $\mathbf{A}$ is computed by Householder reflection or any other methods, the full rank LS problem can be solved by the above procedure.

An alternative method is to solve $\mathbf{A}^T\mathbf{A}x_{LS} = \mathbf{A}^Tb$ by LU decomposition of $\mathbf{A}^T\mathbf{A}$. $\mathbf{C} = \mathbf{A}^T\mathbf{A}$ is a symmetric positive definite matrix, there exists the Cholesky decomposition $\mathbf{C} = \mathbf{G}\mathbf{G}^T$, so solving the two triangular system gives the LS solution:

$$d = \mathbf{A}^Tb, \mathbf{G}y = d, \mathbf{G}^Tx_{LS} = y.$$

If $\mathbf{A}$ doesn't have full column rank, $\mathbf{A}^T\mathbf{A}$ may not be invertible. we can use SVD decomposition of $\mathbf{A}$ to solve the LS problem.

Theorem 7 *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{A} = \mathbf{U\Sigma V}^T$ is its SVD decomposition with $rank(\mathbf{A}) = r$. If $\mathbf{U} = (u_1, \ldots, u_m)$ and $\mathbf{V} = (v_1, \ldots, v_n)$ are column partitions and $b \in \mathbb{R}^m$, then the LS solution to $\mathbf{A}x = b$ is:*

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

Proof 3

$$\begin{aligned} \|\mathbf{A}x - b\|_2^2 &= \|(\mathbf{U}^T\mathbf{A}\mathbf{V})(\mathbf{V}^T x) - \mathbf{U}^T b\|_2^2 = \|\mathbf{\Sigma}\alpha - \mathbf{U}^T b\|_2^2 \\ &= \sum_{i=1}^{r}(\sigma_i\alpha_i - u_i^T b)^2 + \sum_{i=r+1}^{m}(u_i^T b)^2 \end{aligned}$$

where $\alpha = \mathbf{V}^T x$. Clearly, only the first part related to x, so $\alpha = \left(u_i^T b/\sigma_i, \ldots, u_r^T b/\sigma_r, 0, \ldots, 0\right)^T$ minimizes the fitting, thus

$$x_{LS} = \mathbf{V}\alpha = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i.$$

In addition given the SVD decomposition the pseduo inverse of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ and $\mathbf{A}^\dagger = \mathbf{V\Sigma}^\dagger\mathbf{U}^T$, where

$$\mathbf{\Sigma}^\dagger = diag\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0\right) \in \mathbb{R}^{n\times m}.$$

6 Schur Decomposition and Eigenvalue Decomposition

Given a square matrix $\mathbf{A}$, we have interests about what is the simplest form $\mathbf{B}$ in $\mathbb{C}$ or $\mathbb{R}$ under unitary(orthogonal) similarity transform $\mathbf{A} = \mathbf{QBQ}^H$ or similarity transform $\mathbf{A} = \mathbf{XBX}^{-1}$. Matrix $\mathbf{B}$ reveals the intrinsic information of $\mathbf{A}$ in that many attributes and structure of matrices are invariant under similarity transform.

Definition 7 *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$, if there exists a non-zero vector $x \in \mathbb{C}^n$ that satisfies $\mathbf{A}x = \lambda x, \lambda \in \mathbb{C}$, λ is called the* eigenvalue *of matrix $\mathbf{A}$ and x is referred to as* eigenvector.

Eigenvalues are the n roots of matrix $\mathbf{A}$'s *characteristic polynomial* $\det(\lambda\mathbf{I} - \mathbf{A})$, the set of eigenvalues is also called the *spectrum* of $\mathbf{A}$. The sum of the diagonal elements of $\mathbf{A}$ is referred to as *trace* of $\mathbf{A}$,

$$trace(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i.$$

6.1 Schur Decomposition

Theorem 8 **Schur Decomposition**. *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$, then there exists a unitary $\mathbf{Q} \in \mathbb{C}^{n\times n}$ such that*

$$\mathbf{Q}^H\mathbf{A}\mathbf{Q} = \mathbf{T} = \mathbf{D} + \mathbf{N}$$

where $\mathbf{D} = diag(\lambda_1, \dots, \lambda_n)$ *and* $\mathbf{N}$ *is strictly upper triangular.* $\mathbf{Q} = (q_1, \dots, q_n)$ *is a column partitioning of the unitary matrix* $\mathbf{Q}$ *where* q_i *is referred to as* Schur vectors *and from* $\mathbf{AQ} = \mathbf{QT}$ *Schur vector satisfy*

$$\mathbf{A}q_k = \lambda_k q_k + \sum_{i=1}^{k-1} n_{ik} q_i, \quad k = 1 : n.$$

Proof 4 *The theorem obviously holds when* $n = 1$*. Suppose* λ *is an eigenvalue of matrix* $\mathbf{A}$ *and* $\mathbf{A}x = \lambda x$ *with* $x \in \mathbb{C}^n$ *is a unit vector. Then* x *can be extended to a unitary matrix* $\mathbf{U} = (x, u_2, \dots, u_n)$*,*

$$\begin{aligned} \mathbf{A}U &= (\mathbf{A}x, \mathbf{A}u_2, \dots, \mathbf{A}u_n) \\ &= (\lambda x, \mathbf{A}u_2, \dots, \mathbf{A}u_n) \\ &= \mathbf{U} \begin{pmatrix} \lambda & \omega^T \\ 0 & \mathbf{C} \end{pmatrix} \end{aligned}$$

Suppose the theorem holds for matrices of order $n-1$*, there is a unitary* $\tilde{\mathbf{U}}$ *such that* $\tilde{\mathbf{U}}^H \mathbf{C} \tilde{\mathbf{U}}$ *is upper triangular. Thus, lets* $\mathbf{Q} = \mathbf{U} diag(1, \tilde{\mathbf{U}})$ *it is easy to verify the theorem holds for order* n*.*

For a real matrix $\mathbf{A}$, the eigenvalues are either real or conjugate complex in pairs. In order to operate all with real numbers, $\mathbf{T}$ changes to block upper triangular with either 1-by1 or 2-by-2 diagonal blocks which is called as real Schur decomposition.

Theorem 9 **Real Schur Decomposition.***Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$*, then there exists an orthogonal* $\mathbf{Q} \in \mathbb{R}^{n \times n}$ *such that*

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1m} \\ 0 & \mathbf{R}_{22} & \dots & \mathbf{R}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{R}_{mm} \end{pmatrix}$$

where each $\mathbf{R}_{ii}$ *is either a 1-by-1 matrix a 2-by-2 matrix having complex conjugate eigenvalues.*

Proof 5 *The theorem obviously holds for* $n = 1$*. Let* $\mathbf{A} \in \mathbb{R}^{n \times n}$*, if* A *has a real eigenvalue* λ *then* A *can be block diagonalized and reduced to order* $n-1$ *as shown in the proof of Schur decomposition. If* A *has a couple of conjugate complex eigenvalue* $\lambda_{1,2} = \alpha \pm i\beta$*, it is easily to see the corresponding eigenvectors are also complex conjugate* $x_{1,2} = y \pm iz$*, where* y *and* z *are real vectors.*

$$\mathbf{A}(y + iz) = (\alpha + i\beta)(y + iz) \Rightarrow \mathbf{A} \begin{pmatrix} y & z \end{pmatrix} = \begin{pmatrix} y & z \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}.$$

$\beta \neq 0$ *implies that* y *and* z *are independent, thus by Gram-Schimt process we can extend* y *and* z *to an orthogonal* $\mathbf{Q} = (y, (y - r_{12}z)/r_{22}, q_3, \dots, q_n)$*, such that*

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ 0 & \mathbf{R}_{22} \end{pmatrix}$$

where $\mathbf{R}_{11}$ *is a 2-by-2 matrix with eigenvalues* $\lambda_{1,2} = \alpha + i\beta$*. By induction the theorem holds.*

Corollary 3 $\mathbf{A}$ *is normal, namely* $\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H$*, if and only if there exists a unitary* $\mathbf{Q} \in \mathbb{C}^{n\times n}$ *such that* $\mathbf{Q}^H\mathbf{A}\mathbf{Q} = diag(\lambda_1, \ldots, \lambda_n)$*.*

Corollary 4 $\mathbf{A}$ *is real symmetric matrix, there exists an orthogonal* $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *such that* $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = diag(\lambda_1, \ldots, \lambda_n)$*.*

Consider the real Schur decomposition of symmetric $\mathbf{A}$, so $\mathbf{R}$ is also symmetric. And the eigenvalues of 2-by-2 symmetric matrices are real, thus $\mathbf{A}$ can be diagonalized.

6.2 Eigenvalue Decomposition

Theorem 10 Block Diagonal Decomposition. *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *and a* Schur decomposition *as follows:*

$$\mathbf{Q}^H\mathbf{A}\mathbf{Q} = \mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \ldots & \mathbf{T}_{1q} \\ 0 & \mathbf{T}_{22} & \ldots & \mathbf{T}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbf{T}_{qq} \end{pmatrix}$$

assume that the $\mathbf{T}_{ii}$ *are square and the eigenvalues of* $\mathbf{T}_{ii}$ *and* $\mathbf{T}_{jj}$ *are different whenever* $i \neq j$*, then there exists a nonsingular matrix* $\mathbf{Y} \in \mathbb{C}^{n\times n}$*, such that*

$$(\mathbf{Y}^{-1}\mathbf{Q}^H)\mathbf{A}(\mathbf{Q}\mathbf{Y}) = diag(\mathbf{T}_{11}^{'}, \ldots, \mathbf{T}_{qq}^{'}).$$

For matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, the order of eigenvalue λ_i in the characteristic polynomial is referred to as *algebraic multiplicity* of λ_i, the dimensions of $null(\lambda_i\mathbf{I} - \mathbf{A})$ is called *geometric multiplicity* of λ_i which implies the number of independent eigenvectors associated with λ_i.

Corollary 5 Diagonal Decomposition. *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, there exists a non-singular* $\mathbf{X} \in \mathbb{C}^{n\times n}$ *which can diagonalize* $\mathbf{A}$

$$\mathbf{X}^{-1}\mathbf{A}\mathbf{X} = diag(\lambda_1, \ldots, \lambda_n),$$

if and only if the geometric multiplicities of all eigenvalue λ_i *equal to their algebraic multiplicities.*

Theorem 11 Jordan Decomposition. *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$*, then there exists a non-singular* $\mathbf{X} \in \mathbb{C}^{n\times n}$ *such that* $\mathbf{X}^{-1}\mathbf{A}\mathbf{X} = diag(\mathbf{J}_1, \ldots, \mathbf{J}_t)$*,where*

$$\mathbf{J_i} = \begin{pmatrix} \lambda_i & 1 & & \ldots & 0 \\ 0 & \lambda_i & \ddots & & \ldots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \ldots & & 0 & \lambda_i \end{pmatrix}$$

is $m_i - by - m_i$ *square matrix and* $m_1 + \ldots + m_t$*=n,* $\mathbf{J}_i$ *is referred to as* Jordan blocks.

6.3 Hessenberg Decomposition

Theorem 12 Hessenberg Decomposition. *Let* $\mathbf{A} \in \mathbb{R}^{n\times n}$*, then there exists an orthogonal matrix* $\mathbf{Q} \in \mathbb{R}^{n\times n}$*, such that*

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{H}$$

where $\mathbf{H}$ *is a* Hessenberg matrix *which means the elements below the sub-diagonal are zero.*

Proof 6 *We claim* $\mathbf{Q}$ *is a product of* $n-2$ *Householder matrices* $\mathbf{P}_1, \ldots, \mathbf{P}_{n_2}$*. We can find* $n-1$ *order Householder reflection* $\overline{\mathbf{P}}_1$ *to zero the first column of* $\mathbf{A}$ *except the first two entries. Let* $\alpha = (a_{21}, \ldots, a_{n1})^T$ *and* $\overline{\mathbf{P}}_1\alpha = (\overline{a}_{21}, 0, \ldots, 0)^T$*. Let* $\mathbf{P}_1 = diag(1, \overline{\mathbf{P}})$*, note Householder matrices are symmetric and* $\mathbf{P}_1$ *is symmetric, then*

$$\mathbf{P}_1^T\mathbf{A}\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & \overline{\mathbf{P}}_1 \end{pmatrix} \begin{pmatrix} a_{11} & \omega \\ \alpha & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \overline{\mathbf{P}}_1 \end{pmatrix} = \begin{pmatrix} a_{11} & \omega^T\overline{\mathbf{P}}_1 \\ \overline{\mathbf{P}}_1\alpha & \overline{\mathbf{P}}_1\mathbf{A}_{22}\overline{\mathbf{P}}_1 \end{pmatrix}$$

Now suppose the $k-1$ *step has been done we find* $k-1$ *Householder matrices* $\mathbf{P}_1, \ldots, \mathbf{P}_{k-1}$ *such that*

$$(\mathbf{P}_1 \ldots \mathbf{P}_{k-1})^T \mathbf{A} (\mathbf{P}_1 \ldots \mathbf{P}_{k-1}) = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{11} & b_{22} & \mathbf{B}_{23} \\ 0 & \mathbf{B}_{32} & \mathbf{B}_{33} \end{pmatrix}$$

is upper Hessenberg through its first $k-1$ *columns.* $\mathbf{B}_{32}$ *is a vector with* $n-k$ *elements, we can find* $n-k$ *order Householder matrix* $\overline{\mathbf{P}}_k$ *to zero* $\mathbf{B}_{32}$*'s elements except the first entry, Let* $\mathbf{P}_k = diag(\mathbf{I}_{n-k}, \overline{\mathbf{P}}_k)$*, then*

$$(\mathbf{P}_1 \ldots \mathbf{P}_k)^T \mathbf{A} (\mathbf{P}_1 \ldots \mathbf{P}_k) = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13}^T\overline{\mathbf{P}}_k \\ \mathbf{B}_{11} & b_{22} & \mathbf{B}_{23}^T\overline{\mathbf{P}}_k \\ 0 & \overline{\mathbf{P}}_k\mathbf{B}_{32} & \overline{\mathbf{P}}_k\mathbf{B}_{33}\overline{\mathbf{P}}_k \end{pmatrix}$$

is upper Hessenberg through its first k *columns. By induction, the theorem holds.*

If matrix $\mathbf{A}$ is symmetric, the Hessenberg decomposition leads to a tri-diagonal form of $\mathbf{A}$. This claim can be easily verified by setting $\omega = \alpha^T and \mathbf{B}_{23} = \mathbf{B}_{32}^T$ in the above proof.

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & 0 & \ldots & 0 \\ h_{21} & h_{22} & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & h_{n-1\ n} \\ 0 & \ldots & 0 & h_{n\ n-1} & h_{nn} \end{pmatrix}.$$

Companion matrix decomposition is a non-orthogonal(non-unitary in complex domain) analog of the Hessenberg decompositon, just like the relation of Schur decomposition and Jordan decomposition. *Companion matrix* indicates the matrices have the following forms and their transpose, which can be easily derived from the characteristic polynomial $\det(\lambda\mathbf{I} - \mathbf{C}) = c_0 + c_1\lambda + \ldots + c_{n-1}\lambda^{n-1} + \lambda^n$:

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -c_{n-1} \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} -c_{n-1} & \ldots & -c_2 & -c_1 & -c_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Schur Decomposition is an important means to compute eigenvalues. A practical iteration scheme based on Hessenberg decomposition and QR decomposition is called QR iteration as follows:

Hessenberg decomposition
$\mathbf{H}_0 = \mathbf{U}_0^T \mathbf{A} \mathbf{U}_0$
for $k = 1, 2, \ldots$
    QR decomposition
    $\mathbf{H}_{k-1} = \mathbf{U}_k \mathbf{R}_k$
    $\mathbf{H}_k = \mathbf{R}_k \mathbf{U}_k$

The QR iteration converges to the Schur decomposition of matrix $\mathbf{A}$. Please refer to [3] for details.

7 Biconjugate Decomposition

7.1 Biconjugate Decomposition

A variety of matrix decomposition processes can be unified with the Wedderburn rank-one reduction theorem [6], such as Gram-Schmidt orthogonalization process, LU, QR, SVD decomposition.

Theorem 13 *If $\mathbf{A} \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are vectors such that $\omega = y^T \mathbf{A} x \neq 0$, then the matrix $\mathbf{B} \doteq \mathbf{A} - \omega^{-1}\mathbf{A} x y^T \mathbf{A}$ has rank exactly one less than the rank of $\mathbf{A}$.*

Proof 7 *We will show the order of $\mathbf{B}$'s null space is one larger than that of $\mathbf{A}$. $\forall z \in null(\mathbf{A})$,e.g. $\mathbf{A}z = 0$ we get $\mathbf{B}z = 0$, so $null(\mathbf{A}) \subseteq null(\mathbf{B})$. $\forall z \in null(\mathbf{B})$,*

$$0 = \mathbf{B}z = \mathbf{A}z - \omega^{-1}\mathbf{A}x(y^T\mathbf{A}z).$$

Let $k = \omega^{-1} y^T \mathbf{A} z$, which is a scalar, thus

$$\mathbf{A}(z - kx) = 0,$$

$(z - kx) \in null(\mathbf{A})$, note $\mathbf{A}x \neq 0$, the null space of $\mathbf{B}$ is therefore obtained from that of $\mathbf{A}$ by adding x to its basis, which increase the order of this space by 1. Thus, the rank of $\mathbf{B}$ is one less then $\mathbf{A}$.

Suppose $rank(\mathbf{A}) = r$, we can define a *rank reducing process* to generate a sequence of Wedderburn matrices $\{\mathbf{A}_k\}$ by using

$$\mathbf{A}_1 \doteq \mathbf{A}, \mathbf{A}_{k+1} \doteq \mathbf{A}_k - \omega_k^{-1}\mathbf{A}_k x_k y_k^T \mathbf{A}_k$$

for any vector $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ satisfying $\omega_k = y_k^T \mathbf{A}_k x_k \neq 0$. The sequence will terminate in r steps since $\{rank(\mathbf{A}_k)\}$ decreases by exactly one at each step. This process can be summarized in matrix outer-product factorization form:

$$\mathbf{A} = \mathbf{\Phi}\mathbf{\Omega}^{-1}\mathbf{\Psi}^T \tag{1}$$

where $\mathbf{\Omega} \doteq diag\{\omega_1, \ldots, \omega_r\}$, $\mathbf{\Phi} \doteq (\phi_1, \ldots, \phi_r) \in \mathbb{R}^{m\times r}$ and $\mathbf{\Psi} \doteq (\psi_1, \ldots, \psi_r) \in \mathbb{R}^{n\times r}$ with

$$\phi_k \doteq \mathbf{A}_k x_k, \quad \psi_k \doteq \mathbf{A}_k^T y_k$$

Further equ. 1 can be written:

$$\mathbf{A} = (\mathbf{A}_1 x_1, \ldots, \mathbf{A}_r x_r)\, \mathbf{\Omega}^{-1} \left(y_1^T \mathbf{A}_1, \ldots, y_r^T \mathbf{A}_r\right) \tag{2}$$

Note every $\mathbf{A}_k$ can be expressed with $\mathbf{A}$, we can find $\mathbf{U} = (u_1, \ldots, u_r) \in \mathbb{R}^{n\times r}$ and $\mathbf{V} = (v_1, \ldots, v_r) \in \mathbb{R}^{m\times r}$, where $\mathbf{A}u_k = \mathbf{A}_k x_k$ and $v_k^T \mathbf{A} = y_k^T \mathbf{A}_k$.

$$u_k \doteq x_k - \sum_{i=1}^{k-1} \left(\frac{v_i^T \mathbf{A} x_k}{v_i^T \mathbf{A} u_i}\right) u_i, \quad v_k \doteq y_k - \sum_{i=1}^{k-1} \left(\frac{y_k^T \mathbf{A} u_i}{v_i^T \mathbf{A} u_i}\right) v_i$$

Thus equ. 1 can be rewritten as

$$\mathbf{V}^T \mathbf{A} \mathbf{U} = \mathbf{\Omega} \tag{3}$$
$$\mathbf{A} = \mathbf{A} \mathbf{U} \mathbf{\Omega}^{-1} \mathbf{V}^T \mathbf{A}. \tag{4}$$

This matrix decomposition process in equ. 1,3,4 is referred to as *biconjugate decomposition* in [6], which can be easily verified by substitution Wedderburn matrix $\mathbf{A}_{r+1} = 0$ with $\{\mathbf{A}_k\}$.$(\mathbf{U}, \mathbf{V})$ is called $\mathbf{A}$-*biconjugate pair* and $(\mathbf{X}, \mathbf{Y})$ is called $\mathbf{A}$-*biconjugatable*.

Depending on the initial matrix $\mathbf{A}$ and the choice of the vector sets $(\mathbf{X}, \mathbf{Y})$, a variety of factorizations can be derived from biconjugate decomposition. Here we list the results for some well-known matrix decompositions, please refer to [6] for details.

Gram-Schmidt let $\mathbf{A}$ be the identity matrix and $(\mathbf{X}, \mathbf{Y})$ are identical and contain the vectors for which an orthogonal basis is desired, $(\mathbf{U} = \mathbf{V})$ give the resultant orthogonal basis.

LDM For $\mathbf{A} \in \mathbb{R}^{n\times n}$ of rank n, if the A-biconjugatable $(\mathbf{X}, \mathbf{Y})$ are both the identity matrix $(\mathbf{I}, \mathbf{I})$, then equ. 3 provides the unique $\mathbf{LDM}^T$ decomposition of $\mathbf{A}$, where $\mathbf{A} = \mathbf{V}^{-T} \mathbf{\Omega} \mathbf{U}^{-1}$ for $\mathbf{V}^{-T}$ and $\mathbf{U}^{-T}$ unit lower triangular matrices.

QR For $\mathbf{A} \in \mathbb{R}^{n\times n}$ of rank n, if the A-biconjugatable $(\mathbf{X}, \mathbf{Y})$ is $(\mathbf{I}, \mathbf{A})$ and gives the the biconjugate pair $(\mathbf{U}, \mathbf{V}) = (\mathbf{R}_1^{-1}, \mathbf{Q}\mathbf{\Psi})$ and $\mathbf{\Omega} = \mathbf{\Psi}^2$ in equ. 3, where $\mathbf{\Psi}$ is a diagonal matrix and $\mathbf{R}_1^{-1}$ is the unit upper triangular matrix, $\mathbf{R} = \mathbf{\Psi}\mathbf{R}_1$ and $\mathbf{Q}$ give the QR decomposition $\mathbf{A} = \mathbf{QR}$.

SVD For $\mathbf{A} \in \mathbb{R}^{m\times n}$ of rank, the SVD of $\mathbf{A}$ given as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the A-biconjugatable $(\mathbf{X}, \mathbf{Y})$ is $(\mathbf{V}, \mathbf{U})$.

7.2 Summary

$$\mathbf{A} \in \mathbb{C}^{m\times n} \begin{cases} \text{SVD} : \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \\ \text{QR} : \mathbf{A} = \mathbf{QR} \\ \text{(square)Schur} : \mathbf{A} = \mathbf{QTQ}^H \text{(non-unitary)Jordan} : \mathbf{A} = \mathbf{XJX}^{-1} (\cdots)\text{Eigen} : \mathbf{A} = \mathbf{X\Sigma X}^{-1} \\ \text{(real square)Real Schur} : \mathbf{A} = \mathbf{QRQ}^T \text{(symmetric)Eigen} : \mathbf{A} = \mathbf{X\Sigma X}^T \\ \text{(real square)Hessenberg} : \mathbf{A} = \mathbf{QHQ}^T \text{(non-orthogonal)Companion(symmetric)Tri-Diagonal} \\ \text{(non-singular square} \cdots \text{)LU(LDM)} : \mathbf{A} = \mathbf{LU} = \mathbf{LDM}^T \text{(symmetric)LDL} : \mathbf{A} = \mathbf{LDL}^T \\ \text{(symmetric positive definite square)Cholesky} : \mathbf{A} = \mathbf{GG}^T \end{cases}$$

References

[1] E. W. Weisstein. Matrix decomposition. MathWorld–A Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/MatrixDecomposition.html

[2] R. Bronson, *Matrix Methods An Introduction*, 2nd ed. Academic Press, 1991.

[3] G. H.Golub and C. F. Loan, *Matrix Computations*, 2nd ed. Johns Hopkins Press, 1996.

[4] L. Hubert, J. Meulman, and W. Heiser, "Two purposes for matrix factorization: A historical appraisal," *SIAM Review*, vol. 42, no. 1, pp. 68–82, 2000.

[5] P. Lancaster and M. Tismenestsky, *The Theory of Matrices*, 2nd ed., W. Rheinboldt, Ed. Academic Press, 1985.

[6] M. T. Chu, R. E. Funderlic, and G. H. Golub, "A rank-one reduction formula and its applications to matrix factorizations," *SIAM Review*, vol. 37, no. 4, pp. 512–530, 1995.