Singular Value Decomposition

Cholesky decomposition

regularized least-squares routine

Stochastic Average Gradient descent

conjugate gradient solver

**Inverse Matrices:**

For a square matrix A, the inverse is written *A*−1. The inverse of a square matrix is sometimes called a reciprocal matrix.  
→ *AA*−1 = I, where I is an identical matrix.

**Similar Matrices**

A square matrix A is similar to the square matrix B if there exists a nonsingular square matrix X such that, A = *X*−1 BX.

**Symmetric Matrices**

A symmetric matrix is a square matrix that is equal to its transpose. Let B be a symmetric matrix. Then B = BT

**Unitary Matrices**

A square matrix A is a unitary matrix if *A*∗=*A*−1, where A\* is the conjugate transpose and *A*−1 is the matrix inverse.

**Commuting Matrices**

Two matrices are said to be commutative under if AB = BA. But, matrix multiplication is not cummutative (ie AB ≠ BA). Matrices are commutative under addition, ie A + B = B + A.

**Diagonal Matrices**

A diagonal matrix is a square matrix in which the entries outside the main diagonal are all zero. For matrix A, A*ij* = 0 if i ≠ j.

**Positive Definite Matrices**

A symmetric matrix A is said to be positive definite if *XT* A X is positive, for all non-zero column vectors X of n real numbers, where XT denotes the transpose of X.

**Row Equivalent Matrices**

Two matrices are row equivalent if one can be changed to the other by a sequence of elementary row operations or two matrices having the same row space.

**Permutation Matrices**

A permutation matrix is a matrix obtained by permuting the rows of an m x m identity matrix according to some permutation of the numbers 1 to m. There are n! permutation matrices of size m.

The permutation matrices of order two are given below:  
  
[1001] and [0110]

**Singular Matrices**

A matrix is singular if and only if its determinant is 0 and does not have inverse.  
  
A square matrix A is singular, iff, det A = 0.

**Invertible Matrices**

A square matrix A of size m x m is invertible or non-singular, if there is a matrix B such that

AB = BA = *lm*

Here, matrix B is called the inverse of A and is denoted *A*−1

 .

**Elementary Matrices**

A square matrix A of order *n*×*n*

is called an elementary matrix if it is obtained by applying only one elementary row operation to the identity matrix.

**Orthogonal Matrices**

An orthogonal matrix is a square matrix with real entries whose columns and rows are orthogonal unit vectors. Matrix A is orthogonal if its transpose is equal to its inverse, ie. AT = *A*−1

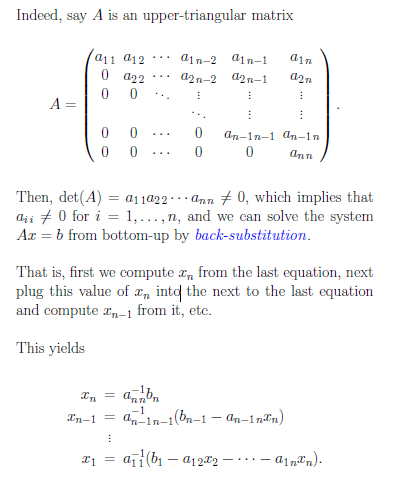
.   
  
Matrix A is orthogonal matrix if A AT = I.

Our goal is to solve the system Ax = b. Since A is assumed to be invertible, we know that this system has a unique solution, x = A−1b.

One should avoid computing the inverse, A−1, of A explicitly. This is because this would amount to solving the n linear systems

most direct methods are based is that if A is an upper-triangular matrix, which means that aij = 0 for 1 =< j < i =< n (resp. lower-triangular, which means that aij = 0 for 1 =< j < i =< n), then

computing the solution, x, is trivial.



If A was lower-triangular, we would solve the system from top-down by forward-substitution.

Thus, what we need is a method for transforming a matrix to an equivalent one in upper-triangular form.This can be done by elimination.

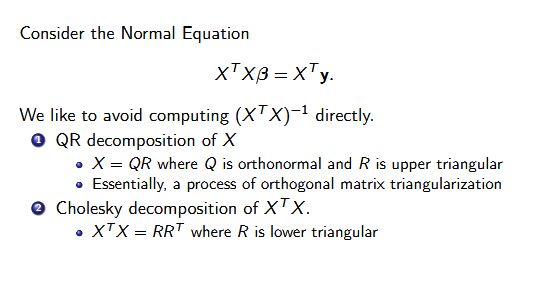
The general method is to iteratively eliminate variables using simple row operations (namely, adding or subtracting a multiple of a row to another row of the matrix) while simultaneously applying these operations to the vector b,to obtain a system, MAx = Mb, where MA is upper-triangular.

**Such a method is called Gaussian elimination.**

**Least Squares Solution from Normal Equations:** Also called normal equation method

Least Squares Solution x minimizes ‖r(x)‖2 , where r(x) =b−Ax for x∈Rn

AT Ax = AT b



**Cholesky decomposition:** If A has full column rank

Also called normal equation method. The Normal Equations Method is much quicker than other algorithms but is in general more unstable.

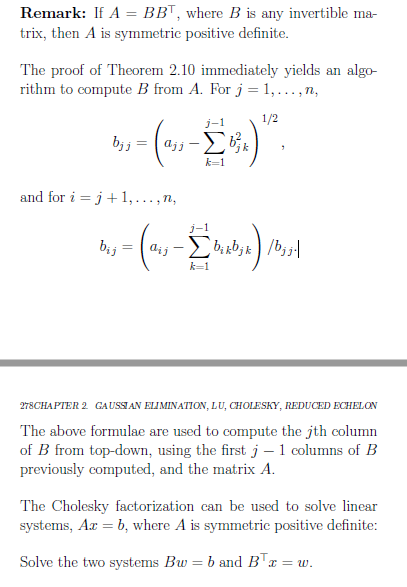
It is commonly used to solve the normal equations ATAX = AT b that characterize the least squares solution to the overdetermined linear system Ax = b.

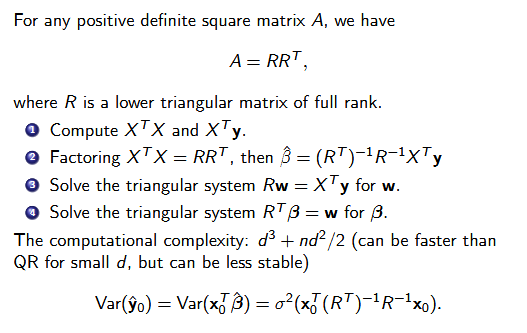
We now consider the special case of symmetric positive definite matrices (SPD matrices).

Recall that an n x n symmetric matrix, A, is positive definite iff XTAX >0

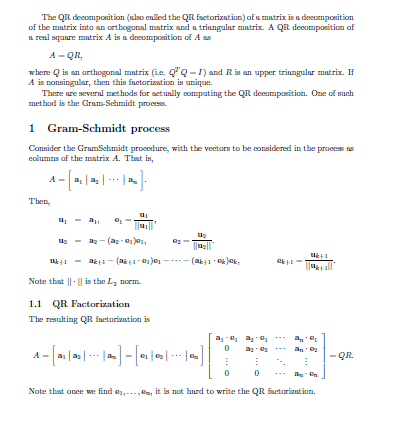
we prove that a symmetric positive definite matrix has a special LU-factorization of the form A = BBT where B is a lower-triangular matrix whose diagonal elements are strictly positive.

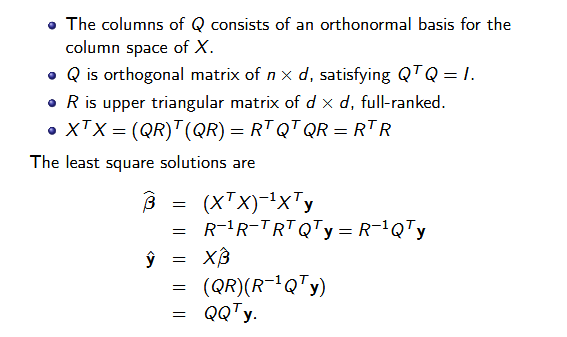
This is the Cholesky factorization

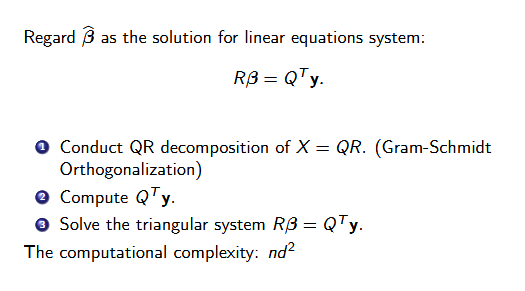




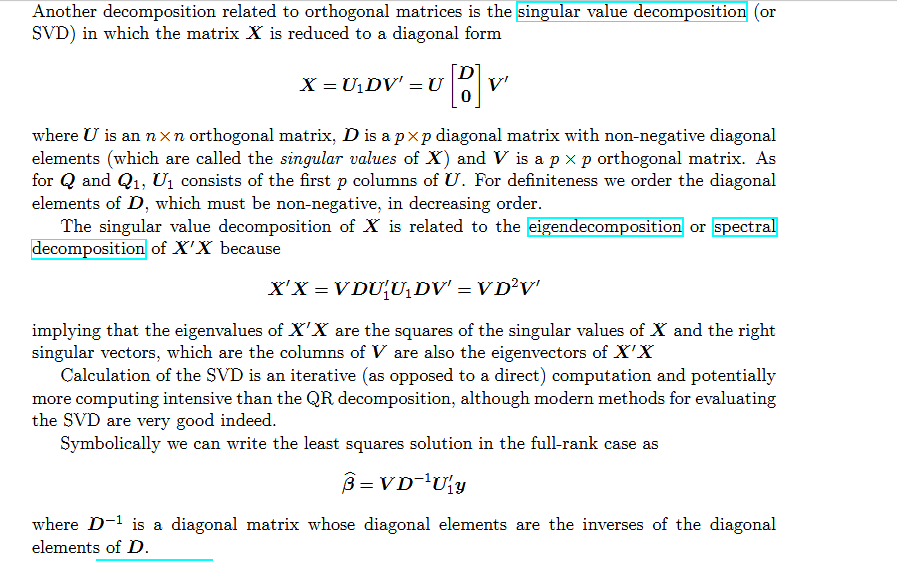
**QR Factorization:**

****





**SVD:** Another decomposition related to orthogonal matrices is the singular value decomposition (or SVD) in which the matrix X is reduced to a diagonal form



**Comparison of Methods:**

So far, we have seen three algorithms for solving the Least Squares Problem:

(1) Normal equations by Cholesky

It is the fastest method but at the same time numerically unstable.

(2) QR Factorization costs

It is more accurate and broadly applicable, but may fail when A is nearly rank-deficient

(3) SVD It is expensive to compute, but is numerically stable and can handle rank

Deficiency

Generally the QR decomposition is preferred to the Cholesky decomposition for least squares problems because there is a certain loss of precision when forming X′X. However, when n is very large you may want to build up X′X using locks of rows. Also, if X is sparse it is an advantage to use sparse matrix techniques to evaluate and store the Cholesky decomposition.

The Matrix package for R provides even more capabilities related to the Cholesky decom- position, especially for sparse matrices