Research Notes

Ruoyang Xu

Last Edited: January 19, 2021

Contents

1 Notations

Red texts: follow up notes.
Blue texts: commentary notes.

2 Todo List

- Householder reflection matrix
- Eigen decomposition matrix implications.
- Optimization

3 Linear Algebra

3.1 Definitions

3.1.1 Hermitian Matrix

Introduction Otherwise known as a self-adjoint matrix, a Hermitian matrix is a *complex square* matrix that is equal to its own *conjugate transpose*. In otherwords

$$A \text{ Hermitian } \Leftrightarrow A = \overline{A^T}$$

Concisely written as $A = A^H$

Properties Entries on the main diagonal of any hermitian matrix are real
forgot why I included this definition, need investigation
Perhaps the definition of a more general "Symmetric matrix"

3.1.2 Definiteness

Introduction Symmetri $n \times n$ real matrix is said to be positive-definite if scalar $z^T M z$ is strictly positive for every non-zero column vector $z \in \mathcal{R}^n$. Semi-definite is defined similarly, except above scalar must be non-negative.

Properties Squre root and Cholesky decomposition might be two relevant properties here.

A matrix is positive semidefinite iff. there is a positive semidefinite matrix B satisfying $M = BB$. B is unique, and is called the non-negative square root of M, denoted with $B = M^{\frac{1}{2}}$.

B is hermitian, so $B* = B$, complex conjugate of itself.

Some use square root and $\sqrt{M}$ for any such dcomposition, or specifically for the Cholesky decomposition, or any decomposition, other only use for the non-negative square root.

Related to Cholesky decomposition as $M = LL*$, where L is lower triangular with non-negative diagonal. See more for section 3.2.2 on Cholesky decomposition.

3.2 Methods

3.2.1 Whitening Transformation

Introduction Otherwise known as **Sphering Transformation**, is a linear transformation that transforms a vector of random variables with a known covariance matrix into a set of new RVs whose covariance is the identity matrix, **meaning uncorrelated**[1] **and each have variance 1**. Transformation is called whitening because it changes the input vector to a "white noise vector", zero mean and finite variance, and statistically independent.

$$Cov(X, Y) = 0 \Rightarrow \text{ Uncorrelated vs Independent } P(X, Y) = P(X)P(Y)$$

For column vector X with non-singular covariance matrix Σ and mean $\mathbf{0}$, then

$$Y = WX, W^T W = \Sigma^{-1}$$

gives whitened random vector Y with unit diagonal covariance. There are infinitely many possible whitening matrices W. Common coices are

1. $W = \Sigma^{-1/2}$ (Mahalanobis or ZCA whitening)
2. Cholesky decomposition of Σ^{-1}
3. Eigen-system of Σ

Optimal whitening transform can be singled out via cross-covariance and cross-correlation of X and Y. The unique optimal whitening transformation achieving maximal component-wise correlation between original X and whitened Y is produced by the whitening matrix

$$W = P^{-1/2} V^{-1/2}$$

where P is the correlation matrix and V is the variance matrix. This seems more of a verification, rather than a method to obtain the result, need to check the original citation [1].

Related transformation

- Decorrelation transform: removes only the correlations but leaves variances intact
- Standardization transform sets variances to 1 but leaves correlations intact
- Coloring transformation transforms a vector of white random variables into a random vector with a specified covariance matrix

3.2.2 Cholesky decomposition

Introduction Otherwise known as Cholesky factorization, is a decomposition of a Hermitian positive-definite matrix to the product of a lower triangular and its conjugate transpose, which is useful for efficient numerical solutions, e.g. Monte Carlo simulations [2].

Statement The cholesky decomposition of a Hermitian positive-definite matrix

$$A = LL^*$$

L is lower triangular matrix with real and positive diagonal entries, and L^* denotes the conjugate tranpose of L. Every A has a unique Cholesky decomposition. Converse holds trivially. Positive semi-definite matrix holds similarly, except the diagonal are allowed to be zero.

[1] Not necessarily independent

[2] When applicable, roughly twice as efficient compared to LU decomposition for solving linear systems [2]

Relevant Methods: LDL

$$A = LDL^* = LD^{1/2}(D^{1/2})^* L^* = \left(LD^{1/2}\right)\left(LD^{1/2}\right)^*$$

L is a lower triangular matrix, and D is a diagonal matrix. Main advantage being that LDL decomposition can be computed and used with essentially the same algorith, but **avoids extracting square roots**. Therefore often called *square-root-free* Cholesky decomposition.

Otherwise called as LDLT decomposition for real matrices. Closely related to the eigen decomposition of real symmetric matrices, $\mathbf{A} = \mathbf{Q\Lambda Q^T}$, see section 3.2.4.

Applications Mainly used for the numerical solution of linear equations $\mathbf{Ax} = \mathbf{b}$. If $\mathbf{A}$ is symmetric and positive definite, then we can solve it by computing Cholesky decomposition $\mathbf{A} = \mathbf{LL}^*$, then $\mathbf{Ly} = \mathbf{b}$ by forward substitution, and finally $\mathbf{L}^*\mathbf{x} = \mathbf{y}$ by back substitution.

Alternative way to eliminate taking square roots in the $\mathbf{LL}^*$ is via LDLT decomposition.

Computation The cholesky algorithm is a modified version of Gaussian elimination. It recursively computes matrix A in the form that

$$A^{(i)} = LA^{(i+1)}L^*$$

Where $A^{(1)} = A$, and at each step, $A^{(i)}$ gives meaning in the following form

$$A^{(i)} = \begin{bmatrix} I & 0 & 0 \\ 0 & a_{i,i} & \mathbf{b}^* \\ 0 & \mathbf{b} & \mathbf{B} \end{bmatrix}, L = \begin{bmatrix} I & 0 & 0 \\ 0 & \sqrt{a_{i,i}} & 0 \\ 0 & \frac{1}{\sqrt{a_{i,i}}}\mathbf{b} & I \end{bmatrix}$$

We compute recursively in this manner until the following $\mathbf{A}^{i+1}$ matrix becomes identity matrix.

$$A^{i+1} = \begin{bmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mathbf{B} - \frac{1}{a_{i,i}}\mathbf{bb}^* \end{bmatrix}$$

This operation in total operates in about $O(n^3)$, roughly about $n^3/3$ FLOPs.

3.2.3 QR Decomposition

Introduction Decomposition of a matrix A into a product $A = QR$ of orthogonal matrix Q and upper triangular matrix R.

Square Matrix For real square matrices, $A = QR$, if A is invertible, then factorization is unique if we require the diagonal elements of R to be positive. For n linearly independent columns in A, first n columns of Q form an orthonormal basis for the column space of A.

Rectangular Matrix More generally, we factor a complex $m \times n, (m \geq n)$ matrix A. Product would be $m \times mQ$ and $m \times n$ uppper triangular matrix R.

$$A = QR = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

Computation Gram-Schmidt method, Householder reflection. numerical instability and add others, explain how to use householder reflection

Application Least squares problem

$$A^*Ax = A^*b$$
$$R^*Rx = (QR)^*QRx = R^*Q^*b \Leftrightarrow Rx = Q^*b$$

There are a couple of advantages to this method

1. Numerical stability (A^TA is worse than A, since the ratio of eigen vector is bigger)
2. Faster

3.2.4 Eigendecomposition of a Matrix

Introduction Factorization of a matrix into a canonical form, whereby matrix is represented in terms of its eigenvalues and eigenvectors. Only diagonalizable[3] matrices can be factorized in this way.

Statement Let A be square matrix $n \times n$ with n linearly independent eigenvectors q_i, then A can be factorized as

$$A = Q\Lambda Q^{-1}$$

where Q is the square matrix $n \times n$ whoce ith column is the ith eigenvector of A and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. The decomposition can be derived from the fundamental property of eigenvectors:

$$Av = \lambda v \Rightarrow AQ = Q\Lambda \Rightarrow A = Q\Lambda Q^{-1}$$

Other characteristics.

4 Optimization

4.1 Methods

4.1.1 Steepest Descent

4.1.2 Gauss-Newton

4.1.3 Levenberg-Maquardt

4.1.4 Dogleg

References

[1] Agnan Kessy, Alex Lewin, and Korbinian Strimmer. "Optimal Whitening and Decorrelation". In: *The American Statistician* 72.4 (Jan. 2018), pp. 309–314. ISSN: 1537-2731. DOI: 10.1080/00031305.2016.1277159. URL: https://arxiv.org/abs/1512.00809.

[2] William H.; Saul A. Teukolsky; William T. Vetterling; Brian P. Flannery Press. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University England EPress, 1992, p. 994. ISBN: 0-521-43108-5.

[3] there exists an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$