Orthogonalization Strategies in a multi-disciplinary perspective

Thesis submitted in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy**

Annavarapu Ramesh Naidu

School of Physics

University of Hyderabad

Hyderabad 500 046

Andhra Pradesh, India

June 2010

To my parents & teachers

Declaration

I hereby declare that the research work embodied in this Ph. D. thesis entitled **Orthogonalization Strategies in a multi-disciplinary perspective** carried out by me under the supervision of Dr. Vipin Srivastava, Professor, School of Physics, University of Hyderabad, India has not been submitted for any other degree or diploma either in part or full to this or any other university or institution.

Place: School of Physics

Date: (A. Ramesh Naidu)

Certificate

This is to certify that the research work embodied in this Ph. D. thesis entitled **Orthogonalization Strategies in a multi-disciplinary perspective** has been carried out by Mr. Annavarapu Ramesh Naidu under my supervision in partial fulfilment of the requirements for the award of the degree of **Doctor of Philosophy** in **Physics** and the same has not been submitted for the award of the degree or diploma either in part or full to this or any other university or institution.

Place: School of Physics

Date:

(Dr. Vipin Srivastava)
Professor
Ph. D. Thesis Supervisor

Dean

School of Physics

University of Hyderabad

Acknowledgements

First and foremost, I would like to express my profound gratefulness to my supervisor Prof. Vipin Srivastava for his coherent advises, myriad patience and inspiring support throughout this research work and my carrier. I could not have completed this work without my supervisor's enduring support and constant co-operation. My supervisor's ebullience towards understanding the fundamental concepts and seeing their implications to the arena of inter-disciplinary areas such as mathematics, physics, biology, computer science and cognitive science etc., has been a continuous source of inspiration to me.

I would like to convey my richest gratefulness to Prof. Anil. K. Bhatnagar, *S. Chandrasekhar Chair Professor* at the University of Hyderabad, Hyderabad and former Vice-Chancellor of the Pondicherry University for all the help.

I am thankful to my doctoral committee members, Prof. S. N. Kaul, Prof. V. S. S. Sastry and Dr. P. K. Suresh for their valuable suggestions. My special thanks are to Prof. S. Chaturvedi, Prof. A. K. Kapoor, Prof. A. P. Pathak, Prof. C. S. Sunandana, Prof. Bindu A. Bambah, Prof. P. Anantha Lakshmi and all other faculty members at the School of Physics for their valuable words.

I am grateful to Prof. K. P. N. Murthy for his help in accessing the CMSD facility at the University of Hyderabad.

I am thankful to the Dean, School of Physics for providing the necessary facilities at the School of Physics.

I am thankful to Prof. G. Govindaraj, Head, Department of Physics, Pondicherry University for his help to finish my work. My thanks are to my colleagues at the Department of Physics, Pondicherry University for their generous support.

I am indebted to my parents for their enduring efforts to impart education as an asset to me and my sisters, without which this day of mine would have remained a dream. I would like to express my appreciation on to my wife for all the support to complete this work. My thanks are to my sisters and brothers-in-law for their support during my work. I am grateful my uncle and aunt at whose residence I have spent good amount of my time carrying out my research. I am grateful to my father-in-law, mother-in-law and their family members for their ample support.

I am thankful to my friend Dr. R. Seshagiri Rao for his constant help and hospitality at the University of Hyderabad. My thanks are to all the research scholars at the School of Physics for their valuable help.

I would like to thank all the non-teaching staff at the School of Physics, University of Hyderabad and at the Department of Physics, Pondicherry University for their help.

Summary

In this thesis we have studied the orthogonalization procedures with the aim of applying them to certain areas in mathematical physics, biology and cognitive science. Although some of these procedures were developed 50 - 100 years ago, it has been discovered recently that some of them possess novel properties, which make them particularly suitable for applications to a variety of areas.

In broad terms the orthogonalization schemes come in two classes – *sequential*, where a give set of vectors is orthogonalized one at a time in an arbitrary sequence; and *democratic*, where the given set of vectors is used in one go to generate an orthonormal basis set. These procedures suffer from a constraint in that the given set of vectors must me linearly-independent, which means that the number of vectors must be less than the dimensionality of the vectors. We have made two proposals to circumvent this problem. This makes the orthogonalization procedures so much more versatile. In the process of this generalization, we understood the equivalence of the democratic orthogonalizations with several other schemes in the literature.

We have used Löwdin orthogonalizations to invent two novel sets of orthogonal polynomials from the set of monomials in appropriate limits and appropriate weight factors.

Löwdin orthogonalizations have proven particularly useful in analysing large volumes of data on gene expression profiles. We can categorise genes on the basis of similarities in their expression. The expression of a gene at a particular time can be expressed in terms of certain characteristic modes which have some simple but novel interpretation.

Application of the Gram-Schmidt orthogonalization to understand some intricate problems in cognitive learning and memory has recently given some useful insight. The approach has been extended to use Löwdin's orthogonalizations. Besides helping in avoiding the serious problem of *memory catastrophe*, orthogonalization enables us to understand those phenomena in memory, which involve sequential and collective processes.

In sum the thesis attempts to explore a range of interdisciplinary areas where the age old orthogonalization methods can be applied to extract new information into intricate phenomena. In the process new directions of further research have opened up.

Contents

Chapter 1

Introduction

1.1 Motivation

"Orthogonalization" is a mathematical transformation whereby a set of vectors is converted into a set of mutually orthogonal vectors. If, in addition, the newly obtained orthogonal vectors are also normalized, then the process is referred to as "orthonormalization". There exist a number of orthogonalization methods to get different orthogonal sets of vectors from a given set of linearly-independent vectors. The oldest and the well-known orthogonalization technique is the Gram-Schmidt method. The Gram-Schmidt procedure takes an arbitrary basis and generates an orthonormal one *sequentially*, processing the list of vectors in an arbitrary sequence and generating vectors that are perpendicular to the previously orthogonalized vectors.

But there are other orthogonalization methods that treat all the vectors *democratically* and form an orthonormal set. These methods due to Löwdin treat all the given vectors simultaneously and on equal footing without any reference to a sequence. They are called *symmetric* and *canonical* orthogonalization methods. These orthogonalization procedures possess some curious geometric properties in terms of the projections of the given vectors on to the orthonormal basis vectors generated from them.

Orthogonalization methods find applications in mathematics, physics, statistics, chemistry, computer science and even to understand certain biological and cognitive phenomena. The Gram-Schmidt procedure has been used to understand certain finer aspects of cognitive memory and the brain's ability to discriminate and categorize. It has also been used to address certain basic questions on acquisition of language in the mental lexicon. The symmetric orthogonalization is useful in the population analysis. The canonical orthogonalization is useful in organizing large volumes of data, e.g., on gene-expression profiles and protein folding etc. The Löwdin orthogonalization procedures can apparently be used for classification and categorization of large volumes of microarray gene expression data. They are useful for data reduction and data fitting too.

1.2 Thesis Outline

This Ph. D. thesis is organized in eight chapters and an appendix. The chapters contain the research work done, the new results, explanation of possible applications and discussion of future directions. In the last chapter, we have given the essence of our research work with conclusions and their possible relevance to further studies. Supplementary material is provided in an appendix which is essential to understand the results obtained in our work. A comprehensive list of references and list of publications is given at the end for all the chapters.

The **first chapter** explains the motivation to the problems undertaken for research and gives the outline of the thesis.

The prelude to the second chapter is given in Appendix A where we have given the gist of basic concepts of vector spaces, linear dependence and linear independence, basis of a space, inner product, orthogonalization, orthonormalization, finite dimensional vector spaces and linear transformation and matrix algebra. This has been done to make the thesis self-contained and to facilitate understanding of the deeper aspects of various orthogonalization methods in terms of projections and projection squares. The appendix also introduces some of the standard orthogonalization methods such as the Gram-Schmidt procedure, the Spectral Decomposition, the Principal Component Analysis (PCA) and the Singular Value Decomposition (SVD).

In the **second chapter**, we have discussed two different orthogonalization procedures due to Löwdin namely the *symmetric* and *canonical*. Their geometric properties are shown which give a deep insight into the techniques, which, it is contended, even the brain might employ to discriminate and categorize information.

Gram-Schmidt orthogonalization method and Löwdin's orthogonalization procedures are applicable only to a set of linearly-independent vectors. But one does not always handle sets of linearly-independent vectors. This restriction can be overcome with the help of another orthogonalization method called the *principal component analysis*. PCA is a standard mathematical tool used to detect correlations in large data sets. Using PCA, one can get uncorrelated vectors which are linearly-independent. This has special properties in terms of the variances. However, PCA lacks the interesting and useful properties that are possessed by, say, the *symmetric* orthogonalization. But the latter is constrained by the requirement of linear-independence. In the **third chapter**, we have proposed two ways of getting round the restriction of linear independence, so that Löwdin methods can be applied to a set of linearly-dependent vectors. We have found that this procedure is more efficient and economical than the PCA. We have extended this work to the *symmetric* orthogonalization and tested its limitations from the computational point of view.

One of the ways of generating the usual orthogonal polynomials such as the Legendre, Hermite and Laguerre etc. is by applying the Gram-Schmidt orthogonalization to the monomials $\{1, x, x^2, x^3, \cdots, x^N\}$, $N = 0, 1, 2, \cdots, \infty$. In the **fourth chapter**, we have worked out two new families of orthogonal polynomials which have some curious geometric properties in terms of their projections and projection squares. We have applied the Löwdin's symmetric and canonical orthogonalizations to these monomials in appropriate limits and with the appropriate weight factors to get entirely new sets of orthogonal polynomials. We have extended this work to the other orthogonal polynomials such as the Chebyshev I, Chebyshev II, Gegenbauer, Jacobi and Bessel. We named the newly obtained polynomials with Löwdin's name as prefix to the standard polynomials. We have found the asymptotic limits and zeros of the new classes of orthogonal polynomials and their graphs are also drawn. A unified view of all the polynomials with our new results are presented in a single graph. We have also compared these graphs with the graphs of classical orthogonal polynomials obtained through the Gram-

Schmidt process. The uniqueness of newly obtained orthogonal polynomials is also discussed in the background of Schweinler and Wigner parameter. The newly generated orthogonal polynomials may stimulate research to develop new approaches and algorithms in the numerical analysis.

In the **fifth chapter**, we have discussed the special properties of Löwdin orthogonalizations and how these methods are distinct from other orthogonalization schemes such as the Polar Decomposition, the Principal Component Analysis and the Singular Value Decomposition. We found a relationship between the symmetric and canonical orthogonalizations. We have analytically derived the relationship to find the symmetric orthonormal basis from the canonical orthonormal basis and vice-versa. The relations between different approaches become apparent and the theoretical basis for new methods is given.

In the **sixth chapter**, we have discussed the computational analysis of DNA microarray gene expression data using the Löwdin's symmetric and canonical orthogonalization methods. We have analyzed the large volumes of microarray gene expression data and found some genes which are co-expressing. These results were compared with the results obtained with the help of other orthogonalization methods such as the Singular Value Decomposition and the Principal Component Analysis. We have investigated the patterns underlying large volumes of gene expression data and divided them into clusters of similar expressions. It turns out that there are simple rules that nature follows in enabling a class of genes to express similarly. At the outset, the data appear extremely complex and hardly any pattern seems to be there in the gene expression profiles and protein structures etc. However, now that we have developed methods for using canonical orthogonalization for an arbitrary number of vectors in a space of given dimensions, it has become possible to detect underlying patterns and categorize the gene expression profiles. We have found that Löwdin's orthogonalization methods are more efficient in interpreting the results in terms of the characteristic modes. These characteristic modes, discussed by others also, are shown by us as the vectors projected onto the orthonormal bases in the case of Löwdin orthogonalizations.

In the **seventh chapter**, we have discussed the application of Löwdin methods to the Hopfield model of neural networks. We have tested the retrieval percentage of stored information in Hopfield networks using Gram-Schmidt and Löwdin methods as storage strategies. We have studied the Hopfield model in neural networks to understand how the brain stores, retrieves and processes the information. We have studied the Hopfield model for large number of neurons with large number of patterns of firing/not firing neurons. We have simulated the memory using Hopfield model with Gram-Schmidt and Löwdin's symmetric and canonical orthogonalization methods as storage strategies for huge data sets. We have seen how the stored synaptic patterns can be retrieved. From this data analysis, we have obtained some valuable information on how the brain categorizes and discriminates between different tasks simultaneously. We have got better understanding of the relationship between the number of patterns presented to the network and the number of patterns that can be retrieved with a maximum overlap of greater than or equal to 97%.

In the **eighth chapter**, we have summarised the conclusions of our research work based on applications of orthogonalization techniques to a variety of problems from a range of disciplines.

Chapter 2

Löwdin Orthogonalization Strategies

2.1 Introduction

In 1947, Per Olov Löwdin [1], a Swedish Chemist, introduced two new orthogonalization methods called *symmetric* and *canonical* orthogonalizations. These orthogonalization methods are widely used in quantum chemistry to orthogonalize the orbitals of electrons. It was found by V. Srivastava [5] that the Löwdin's orthogonalization methods have curious geometric properties in terms of the projections of the initial basis on to the orthonormal basis.

In this chapter, the Löwdin orthogonalization schemes are introduced and some new findings are discussed. The curious novel properties that these methods possess are discussed with the help of the projection-square matrix namely the Schweinler-Wigner matrix. We discuss their possible applications as well.

2.2 Löwdin Orthogonalization Procedures

Let $\boldsymbol{V} = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \cdots, \vec{\boldsymbol{v}}_m\}$ represent a given set of linearly-independent vectors in an n-dimensional space which can in general be a complex vector space $\mathbb{C}^n$.

A *non-linear singular transformation* $\boldsymbol{A}$ can transform the basis $\boldsymbol{V}$ into a new basis $\boldsymbol{Z}$ as follows,

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{A} \tag{2.1}$$

The new basis will be orthonormal if

$$\langle \boldsymbol{Z} \mid \boldsymbol{Z} \rangle = \langle \boldsymbol{V}\boldsymbol{A} \mid \boldsymbol{A}\boldsymbol{V} \rangle = \boldsymbol{A}^{\dagger}\langle \boldsymbol{V} \mid \boldsymbol{V} \rangle \boldsymbol{A} = \boldsymbol{A}^{\dagger}\boldsymbol{M}\boldsymbol{A} = \boldsymbol{I}. \tag{2.2}$$

Here $\boldsymbol{M}$ is the Hermitian metric matrix or the Gram matrix of the given set $\boldsymbol{V}$ and $\dagger$ on $\boldsymbol{A}$ represents the adjoint of matrix $\boldsymbol{A}$.

The substitution

$$\boldsymbol{A} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{B}, \tag{2.3}$$

where $\boldsymbol{B}$ is an arbitrary unitary matrix, gives the general solution of the orthogonalization problem.

2.3 The Hermitian Metric Matrix

The matrix $\boldsymbol{M}$ is the Hermitian metric matrix or the overlap matrix or the Gram matrix which is formed from the inner product of the given set of linearly independent vectors. Let $\boldsymbol{V}$ be a matrix formed by the set of linearly independent vectors as its columns,

$$\boldsymbol{V} = [\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \cdots, \vec{\boldsymbol{v}}_m] = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{pmatrix} \tag{2.4}$$

Then the Hermitian matrix can be explicitly constructed as follows:

$$\boldsymbol{M} = \boldsymbol{V}^T\boldsymbol{V} = \begin{pmatrix} \vec{v}_1.\vec{v}_1 & \vec{v}_1.\vec{v}_2 & \cdots & \vec{v}_1.\vec{v}_m \\ \vec{v}_2.\vec{v}_1 & \vec{v}_2.\vec{v}_2 & \cdots & \vec{v}_2.\vec{v}_m \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_m.\vec{v}_1 & \vec{v}_m.\vec{v}_2 & \cdots & \vec{v}_m.\vec{v}_m \end{pmatrix}. \tag{2.5}$$

The Hermitian matrix is symmetric and positive semi-definite.

2.4 The $M^{-1/2}$ Transformation

The symmetric orthogonalization involves finding the inverse square root of the overlap matric $\boldsymbol{M}$. This can be achieved by first diagonalizing the Gram matrix with the help a linear transformation $\boldsymbol{B}$.

If $\boldsymbol{B} = \boldsymbol{U}$, where $\boldsymbol{U}$ diagonalizes $\boldsymbol{M}$, we will have

$$\boldsymbol{d} = \boldsymbol{U}\boldsymbol{M}\boldsymbol{U}^\dagger = \boldsymbol{U}\boldsymbol{M}^{1/2}\boldsymbol{M}^{1/2}\boldsymbol{U}^\dagger = \boldsymbol{U}\boldsymbol{M}^{1/2}\underbrace{\boldsymbol{U}^\dagger\boldsymbol{U}}\boldsymbol{M}^{1/2}\boldsymbol{U}^\dagger$$

$$\boldsymbol{d} = \underbrace{\boldsymbol{U}\boldsymbol{M}^{1/2}\boldsymbol{U}^\dagger}\ \underbrace{\boldsymbol{U}\boldsymbol{M}^{1/2}\boldsymbol{U}^\dagger} = \boldsymbol{d}^{1/2}\ \boldsymbol{d}^{1/2}$$

$$\Longrightarrow \boldsymbol{d}^{1/2} = \boldsymbol{U}\boldsymbol{M}^{1/2}\boldsymbol{U}^\dagger \tag{2.6}$$

Multiplying both sides from right with $\boldsymbol{U}$, we get

$$\boldsymbol{d}^{1/2}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{M}^{1/2}\underbrace{\boldsymbol{U}^\dagger\boldsymbol{U}} \tag{2.7}$$

$$\boldsymbol{d^{1/2}U} = \boldsymbol{UM^{1/2}} \tag{2.8}$$

Further, multiplying both sides from left with $\boldsymbol{U}^{\dagger}$, we get

$$\boldsymbol{U^{\dagger}d^{1/2}U} = \underbrace{\boldsymbol{U^{\dagger}U}}\boldsymbol{M^{1/2}} \tag{2.9}$$

$$\Longrightarrow \boldsymbol{M^{1/2}} = \boldsymbol{U^{\dagger}d^{1/2}U.} \tag{2.10}$$

$$\text{Or, } \boldsymbol{M^{-1/2}} = \boldsymbol{U^{\dagger}d^{-1/2}U.} \tag{2.11}$$

2.5 Symmetric Orthogonalization

The specific choice $\boldsymbol{B} = \boldsymbol{I}$ gives Löwdin symmetric orthogonalization,

$$\boldsymbol{Z} \equiv \boldsymbol{\Phi} = \boldsymbol{VM^{-1/2}}, \tag{2.12}$$

where the symmetric orthonormal basis is,

$$\boldsymbol{\Phi} = [\vec{\boldsymbol{\phi}}_{\boldsymbol{1}}, \vec{\boldsymbol{\phi}}_{\boldsymbol{2}}, \cdots, \vec{\boldsymbol{\phi}}_{\boldsymbol{m}}]. \tag{2.13}$$

The symmetric orthonormal basis formed from the set of linearly independent vectors spans the basis set $\boldsymbol{V}$. The symmetric orthonormal basis can be written as follows where the columns of $\boldsymbol{\Phi}$ spans the columns of $\boldsymbol{V}$.

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{n1} & \phi_{n2} & \cdots & \phi_{nm} \end{pmatrix}. \tag{2.14}$$

For complete orthonormality, the symmetric bases should satisfy the identity

$$< \boldsymbol{\Phi} \mid \boldsymbol{\Phi} > = \boldsymbol{I}. \tag{2.15}$$

If the initial basis set is not completely linearly independent or they are close to linear dependence then the symmetric orthonormal set is not "completely orthonormal" but is "near orthonormal". It can be written as

$$< \boldsymbol{\Phi} \mid \boldsymbol{\Phi} > \neq \boldsymbol{I} \approx \boldsymbol{I}. \tag{2.16}$$

If the given basis set is close to linear dependence, then the Hermitian metric matrix $\boldsymbol{M}$ becomes *positive semi-definite* and it is said to be *skew-Hermitian*. This results in *zero* or *negative* eigenvalues. They disturb the completeness of orthogonality. So, the completeness of orthonormal set of vectors obtained through the symmetric orthogonalization depends upon whether the given vectors are completely linearly independent or dependent. If the set is linearly dependent, the symmetric orthonormal set is said to be *near orthonormal* or *semi-orthonormal*.

2.6 Remarkable Property

The orthonormal set obtained through symmetric orthogonalization possesses a remarkable property which the other orthogonalization procedures do not have. The novel property is that the orthonormal basis resembles the original set in the *nearest-neighbour* sense in the case of linearly independent vectors and in the *next-nearest neighbour* sense if the basis is close to linear dependence. This property enables us to estimate the closeness of the newly obtained orthonormal set with the initial set of vectors in the case of linearly independent vectors. It is also useful to estimate the closeness of the orthonormal set with the original set to the first order approximation in the case of linearly dependent set. The dominant overlaps are those between neighbours of the given set of vectors.

2.7 Canonical Orthogonalization

The specific choice $\boldsymbol{B} = \boldsymbol{U}$ gives the canonical orthogonalization

$$\boldsymbol{Z} \equiv \boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{U}\boldsymbol{d}^{-\frac{1}{2}}. \tag{2.17}$$

The canonical orthonormal bases are given as

$$\boldsymbol{\Lambda} = [\vec{\boldsymbol{\lambda}}_1, \vec{\boldsymbol{\lambda}}_2, \cdots, \vec{\boldsymbol{\lambda}}_m]. \tag{2.18}$$

The canonical orthonormal basis formed from the set of linearly independent vectors spans the basis set $\boldsymbol{V}$. The canonical orthonormal basis can be written as follows where the columns of $\boldsymbol{\Lambda}$ span the columns of $\boldsymbol{V}$.

$$\boldsymbol{\Lambda} = [\vec{\boldsymbol{\lambda}}_1, \vec{\boldsymbol{\lambda}}_2, \cdots, \vec{\boldsymbol{\lambda}}_m] = \begin{pmatrix} \boldsymbol{\lambda}_{11} & \boldsymbol{\lambda}_{12} & \cdots & \boldsymbol{\lambda}_{1m} \\ \boldsymbol{\lambda}_{21} & \boldsymbol{\lambda}_{22} & \cdots & \boldsymbol{\lambda}_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\lambda}_{n1} & \boldsymbol{\lambda}_{n2} & \cdots & \boldsymbol{\lambda}_{nm} \end{pmatrix}. \tag{2.19}$$

For complete orthonormality of the canonical bases, it should satisfy the identity

$$< \boldsymbol{\Lambda}|\boldsymbol{\Lambda} > = \boldsymbol{I}. \tag{2.20}$$

2.8 Schweinler-Wigner Matrix and Geometrical Properties

The Schweinler-Wigner matrix [11] is formed by taking the projection-squares of the given vectors $\{\vec{\boldsymbol{v}}_k\}$ onto an orthonormal basis set $\{\vec{\boldsymbol{z}}_\kappa\}$.

It can be constructed as follows,

$$\begin{pmatrix} |(\vec{v}_1,\vec{z}_1)|^2 & |(\vec{v}_1,\vec{z}_2)|^2 & \dots & |(\vec{v}_1,\vec{z}_N)|^2 \\ |(\vec{v}_2,\vec{z}_1)|^2 & |(\vec{v}_2,\vec{z}_2)|^2 & \dots & |(\vec{v}_2,\vec{z}_N)|^2 \\ \vdots & \vdots & & \vdots \\ |(\vec{v}_N,\vec{z}_1)|^2 & |(\vec{v}_N,\vec{z}_2)|^2 & \dots & |(\vec{v}_N,\vec{z}_N)|^2 \end{pmatrix}.$$

The elements in a row corresponding to a particular $\vec{v}_k$ add up to $|\vec{v}_k|^2$:

$$\sum_{\kappa} |(\vec{v}_k,\vec{z}_\kappa)|^2 = |\vec{v}_k|^2\,, \quad k = 1, ..., N\,, \tag{2.21}$$

and the elements in a column for a particular $\vec{z}_\kappa$ add up to a real positive number c_κ:

$$\sum_{k} |(\vec{v}_k,\vec{z}_\kappa)|^2 \; = (AMMA^\dagger)_{\kappa\kappa} = (BMB^\dagger)_{\kappa\kappa} = c_\kappa\,, \quad \kappa = 1, ..., N. \tag{2.22}$$

Note that we have the identity

$$\sum_{\kappa=1}^{N} c_\kappa = \sum_{k=1}^{N} |\vec{v}_k|^2, \tag{2.23}$$

a constant for a given set V. A basis set Z will satisfy the set of simultaneous equations (2.22) with the positive real numbers c_κ's obeying the identity (2.23). We can gain further insight in terms of Sheweinler-Wigner parameter

$$\sum_{\kappa} {c_\kappa}^2 = m. \tag{2.24}$$

Specific bases will satisfy additional conditions on the values of c_κ's either through (2.24) or otherwise. For instance, if $c_\kappa = |\vec{v}_k|^2$ with $\kappa = k$ (i.e. $B = I$ in (2.22)), we get the *Symmetric* basis $Z = \Phi$.

For $m = m_{max}$, which will arise for the maximally lop-sided distribution of the c_κ's (satisfying (2.23)), we get the *canonical* orthonormal basis $Z = \Lambda$; and

$\boldsymbol{m} = \boldsymbol{m_{min}}$, which will correspond to an average distribution, $\boldsymbol{c_1} = \boldsymbol{c_2} = \boldsymbol{...} = \boldsymbol{c_N} = (\boldsymbol{c_1} + \boldsymbol{c_2} + \boldsymbol{...} + \boldsymbol{c_N})/\boldsymbol{N},$ will give the basis $\boldsymbol{Z} = \boldsymbol{\Phi}$ for normalized $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s.

Following useful information is embedded in the above identification: In the symmetric case, where $\boldsymbol{Z} = \boldsymbol{\Phi}$, since the sum of squared-projections of all the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s on a $\vec{\boldsymbol{\phi}}_{\boldsymbol{\kappa}}$, say $\vec{\boldsymbol{\phi}}_{\boldsymbol{l}}$, is equal to the sum of squared-projections of all the $\vec{\boldsymbol{\phi}}_{\boldsymbol{\kappa}}$'s on the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$ with $\boldsymbol{k} = \boldsymbol{l}$, the symmetry properties of $\boldsymbol{V}$, if any, are preserved in $\boldsymbol{\Phi}$. This feature also ensures that $\boldsymbol{\Phi}$ resembles the original set $\boldsymbol{V}$ in that Löwdin's resemblance measure [2], $\langle \boldsymbol{Z} - \boldsymbol{V} | \boldsymbol{Z} - \boldsymbol{V} \rangle$, has its smallest value when $\boldsymbol{B}$=$\boldsymbol{I}$ or $\boldsymbol{Z} = \boldsymbol{\Phi}$. If seen slightly differently, the above symmetry interestingly implies that the squared-projections of all the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s on a $\vec{\boldsymbol{\phi}}_{\boldsymbol{l}}$ add up to the squared length of $\vec{\boldsymbol{v}}_{\boldsymbol{l}}$.

The last property above turns into a stricter condition in the case of $\boldsymbol{m} = \boldsymbol{m_{min}}$: the basis vectors $\vec{\boldsymbol{\phi}}_{\boldsymbol{\kappa}}$ are arranged such that the sum of squared-projections of all the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s on each $\vec{\boldsymbol{\phi}}_{\boldsymbol{\kappa}}$ is the *same* and is equal to the average of $|\vec{\boldsymbol{v}}_{\boldsymbol{k}}|^{\boldsymbol{2}}$ irrespective of how the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s are arranged. In effect, the set $\boldsymbol{\Phi}$ is arranged so as to cancel the effects of inhomogeneity in the distribution of $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s. On the other hand in the $\boldsymbol{m} = \boldsymbol{m_{max}}$ case, with $\boldsymbol{Z} = \boldsymbol{\Lambda}$, the basis vectors $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$'s must be oriented such that they sample those directions in which bunches of $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s tend to be oriented. In order to attain $\boldsymbol{m} = \boldsymbol{m_{max}}$ the Canonical basis set is arranged in such an optimal fashion that the sum of squared-projections of all the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s on one of the $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$'s, say $\vec{\boldsymbol{\lambda}}_{\boldsymbol{1}}$ – *i.e.* $\boldsymbol{c_1}$ – is the largest. After the $\vec{\boldsymbol{\lambda}}_{\boldsymbol{1}}$, is fixed, the rest of the set, orthogonal to $\vec{\boldsymbol{\lambda}}_{\boldsymbol{1}}$, is oriented such that another $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$, say $\vec{\boldsymbol{\lambda}}_{\boldsymbol{2}}$, is able to maximize the *total* squared-projection of all the $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s on it. This will be $\boldsymbol{c_2}$ and will be smaller than $\boldsymbol{c_1}$. All the $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$'s are arranged according to this optimum-principle that ensures for the given set $\{\vec{\boldsymbol{v}}_{\boldsymbol{k}}\}$ the most lop-sided distribution of $\boldsymbol{c_\kappa}$'s with $\boldsymbol{c_1} > \boldsymbol{c_2} > \boldsymbol{...} > \boldsymbol{c_N}$. This particular set of $\boldsymbol{c_\kappa}$'s in fact comprises the eigenvalues of $\boldsymbol{M}$ in the basis $\boldsymbol{\Lambda}$ (take $\boldsymbol{B} = \boldsymbol{U}$ in equation (2.22)) which are generically non-degenerate. The arrangement of $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$'s that yields the S-W condition of $\boldsymbol{m} = \boldsymbol{m_{max}}$ manifests Löwdin's optimal property [2] of the canonical orthogonalization. Note that just as the $\boldsymbol{c_\kappa}$ gains its largest values $(\mathbf{d})_{\kappa\kappa}$ for $\boldsymbol{B}$=$\boldsymbol{U}$, the quantity

$$\sum_{\alpha} |\boldsymbol{A_{\alpha\kappa}}|^{2} = (\mathbf{A}^{\dagger}\mathbf{A})_{\kappa\kappa} = (\mathbf{B}^{\dagger}\mathbf{M}^{-1}\mathbf{B})_{\kappa\kappa}\,, \tag{2.25}$$

has its smallest value for $\boldsymbol{B} = \boldsymbol{U}$, in which case it is $(\mathbf{d}^{-1})_{\kappa\kappa}$. If $\boldsymbol{d_1^{-1}}$ is the smallest value and the associated basis vector is $\vec{\boldsymbol{\lambda}}_{\mathbf{1}}$, Löwdin [2] showed that for all vectors orthogonal to $\vec{\boldsymbol{\lambda}}_{\mathbf{1}}$ the sum $\sum_\alpha |\boldsymbol{A_{\alpha\kappa}}|^2$ has the smallest value $\boldsymbol{d_2^{-1}} (> \boldsymbol{d_1^{-1}})$ associated with $\vec{\boldsymbol{\lambda}}_{\mathbf{2}}$, and that one can go on in this manner to find the smallest $\boldsymbol{d_\kappa^{-1}}$'s associated with the respective $\vec{\boldsymbol{\lambda}}_{\boldsymbol{\kappa}}$'s.

It should be noted that the value of $\boldsymbol{m_{max}}$ depends on the distribution of orientations of $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s relative to each other, whereas the value of $\boldsymbol{m_{min}}$ is independent of this. There can be any number of $\boldsymbol{V}$'s satisfying a particular identity (2.23) but differing in distributions of orientations of the vectors $\vec{\boldsymbol{v}}_{\boldsymbol{k}}$'s. Each of these will have a different $\boldsymbol{m_{max}}$ but the same $\boldsymbol{m_{min}}$. The inhomogeneity in the distribution of directions of a given set of vectors will decide the value of $\boldsymbol{m_{max}}$ – the larger the inhomogeneity, the larger will be the value of $\boldsymbol{m_{max}}$.

Chapter 3

Getting Round the Restriction of Linear-Independence in the Löwdin Orthogonalizations

3.1 Introduction

The problem of conversion of a given set of linearly-independent vectors into a set of mutually orthogonal vectors has been studied for over a century. It finds applications in mathematics, physics, chemistry, and even in understanding certain biological and cognitive phenomena. The methods come in two categories: 'sequential', where the given vectors are picked up one by one and orthogonalized; and 'democratic', where all the given vectors are handled simultaneously and an orthonormal basis is constructed to conform to certain chosen conditions. Two prominent orthogonalization procedures of the latter kind, widely used by quantum chemists, are due to Löwdin [1, 3, 4].

Löwdin's canonical orthogonalization procedure is applicable only to a set of linearly-independent vectors. This puts a serious constraint for there can be at most $\boldsymbol{n}$ linearly-independent vectors in an $\boldsymbol{n}$-dimensional space. We propose two ways of getting round this restriction so that Löwdin's procedure can be used to find the vector along which all the given vectors – any number of them in a space of arbitrary dimensionality – project maximally. Under these conditions this orthogonalization procedure is equivalent to the PCA [12].

Let $\boldsymbol{V} = \{\vec{\boldsymbol{v}}^1, \vec{\boldsymbol{v}}^2, \cdots, \vec{\boldsymbol{v}}^N\}$ possess $\boldsymbol{N}$ vectors with $\boldsymbol{n}$, the dimensionality of the vector space, $\boldsymbol{N} > \boldsymbol{n}$, the dimensionality of each vector. Consequently the vectors will not be linearly-independent and we will not be able to use canonical orthogonalization to sample those directions in which the given vectors have a hierarchy of maximal projections.

3.2 Methods of Getting Round the Restriction of Linear-Independence

We propose two ways of getting round the restriction of the condition of linear independence to apply the Löwdin methods to any arbitrary set of vectors. The two methods are discussed in the following two subsections.

3.2.1 General Method

To overcome the above problem we have proposed to view $\boldsymbol{N}$ vectors in $\boldsymbol{n}$-dimension as $\boldsymbol{n}$ vectors in $\boldsymbol{N}$-dimension, i.e., the dimensionality of the vector space and the dimensionality of each vector are interchanged.

Consider n vectors, $\underset{\rightarrow}{v_1}, \underset{\rightarrow}{v_2}, \cdots, \underset{\rightarrow}{v_n}$, such that $\underset{\rightarrow}{v_1}$ comprises the first components of $\vec{v}^1$ through $\vec{v}^N$, (i.e. $\underset{\rightarrow}{v_1} = (v_1^1, v_1^2, \cdots, v_1^N)$), $\underset{\rightarrow}{v_2}$ comprises the second components of $\vec{v}^1$ through $\vec{v}^N$, and so on. Now we can construct the $n \times n$ Hermitian metric matrix M from the N-dimensional vectors $\underset{\rightarrow}{v_1}, \underset{\rightarrow}{v_2}, \cdots, \underset{\rightarrow}{v_n}$.

In other words, while under normal circumstances of $N \leq n$ one constructs $\vec{M}$ by taking $V^T V$ (V, is an $n \times N$ matrix) as follows,

$$\vec{M} = V^T V = \begin{pmatrix} \vec{v}^1.\vec{v}^1 & \vec{v}^1.\vec{v}^2 & \cdots & \vec{v}^1.\vec{v}^N \\ \vec{v}^2.\vec{v}^1 & \vec{v}^2.\vec{v}^2 & \cdots & \vec{v}^2.\vec{v}^N \\ \vdots & \vdots & \vdots & \vdots \\ \vec{v}^N.\vec{v}^1 & \vec{v}^N.\vec{v}^2 & \cdots & \vec{v}^N.\vec{v}^N \end{pmatrix}, \tag{3.1}$$

for $N > n$, we have proposed to construct $\underset{\rightarrow}{M}$ by taking $V V^T$,which can be represented in terms of $\underset{\rightarrow}{v_i}$'s as,

$$\underset{\rightarrow}{M} = V V^T = \begin{pmatrix} \underset{\rightarrow}{v_1}.\underset{\rightarrow}{v_1} & \underset{\rightarrow}{v_1}.\underset{\rightarrow}{v_2} & \cdots & \underset{\rightarrow}{v_1}.\underset{\rightarrow}{v_n} \\ \underset{\rightarrow}{v_2}.\underset{\rightarrow}{v_1} & \underset{\rightarrow}{v_2}.\underset{\rightarrow}{v_2} & \cdots & \underset{\rightarrow}{v_2}.\underset{\rightarrow}{v_n} \\ \vdots & \vdots & & \vdots \\ \underset{\rightarrow}{v_n}.\underset{\rightarrow}{v_1} & \underset{\rightarrow}{v_n}.\underset{\rightarrow}{v_2} & \cdots & \underset{\rightarrow}{v_n}.\underset{\rightarrow}{v_n} \end{pmatrix}. \tag{3.2}$$

The matrix $\underset{\rightarrow}{M}$ is some times referred to as 'sum of squares and cross products' (SSCP) matrix in the literature [6]. It is in fact the so called Covariance matrix [7] without subtracting the mean. In Covariance matrix, one constructs $\underset{\rightarrow}{v_m}$ as $((v_m^1 - \overline{v}_m), (v_m^2 - \overline{v}_m), \cdots, (v_m^N - \overline{v}_m))$, where $\overline{v}_m$ is the average of the m^{th} component of all the N vectors, and is ignored in constructing the SSCP matrix. The PCA [9] is typically done using $\underset{\rightarrow}{M}$. While the i-j element of $\vec{M}$ gives the overlap of i^{th} and j^{th} vectors, i.e. overlap of all the features of vectors, $\vec{v}^i$ and $\vec{v}^j$, the $\boldsymbol{p}$-$\boldsymbol{q}$ element in $\underset{\rightarrow}{\boldsymbol{M}}$ represents the overlap of $\boldsymbol{p}^{\boldsymbol{th}}$ and $\boldsymbol{q}^{\boldsymbol{th}}$ features (or components) of all the given $\boldsymbol{N}$ vectors.

When $\boldsymbol{N} > \boldsymbol{n}$, $\vec{\boldsymbol{M}}$ can be constructed but canonical orthogonalization (CO) cannot be applied as the given vectors will not be linearly-independent. However, we can have the condition of (no. of vectors) $\leq$ (dimensionality) satisfied by interchanging $\boldsymbol{N}$ and $\boldsymbol{n}$ and can then apply CO. This amounts to applying canonical orthogonalization to $\underset{\rightarrow}{\boldsymbol{M}}$ [8]. We found numerically that the CO applied through $\underset{\rightarrow}{\boldsymbol{M}}$ for $\boldsymbol{n}$ *normalized* $\boldsymbol{N}$-dimensional vectors is equivalent to the PCA applied through $\underset{\rightarrow}{\boldsymbol{M}}$ constructed from $\boldsymbol{N}$ *normalized* $\boldsymbol{n}$-dimensional vectors.

3.2.2 Hierarchical Method

We have also attempted the problem of canonical orthogonalization for $\boldsymbol{N} > \boldsymbol{n}$ vectors through the usual $\vec{\boldsymbol{M}}$-route. We found that we can proceed in a hierarchical manner. We have divided $\boldsymbol{N}$ vectors into $\boldsymbol{N}/\boldsymbol{n}$ (+1, if $\boldsymbol{N}$ is not a multiple of $\boldsymbol{n}$) groups. We have applied canonical orthogonalization to each group and isolated the canonical orthogonalization basis vectors which get the maximum projection in each group. We make groups from these vectors with $\leq \boldsymbol{n}$ vectors in each group and apply canonical orthogonalization to these groups and continue until we are left with one group of $\leq \boldsymbol{n}$ vectors. The canonical orthogonalization applied to this group will yield the final vector that captures the maximum projection of the given $\boldsymbol{N}$ vectors. Computationally this method is more economical than the CO or PCA applied through $\underset{\rightarrow}{\boldsymbol{M}}$.

3.3 Computations, Results and Discussion

Before coming to the situation of main interest, namely $\boldsymbol{N > n}$, we have studied the case of $\boldsymbol{N = n}$ where the canonical orthogonalization holds in normal course. We have considered a number of random examples of 3 vectors in 3-dimensions $\boldsymbol{(n = 3)}$, a sample of which is listed in Table 3.1. Hermitian metric matrices $\vec{\boldsymbol{M}}$ and $\underset{\to}{\boldsymbol{M}}$ were constructed as in equations (3.1) and (3.2). We found that $\vec{\boldsymbol{M}}$ and $\underset{\to}{\boldsymbol{M}}$ always have the same eigenvalues but their eigenvectors are different. However, the eigenvectors of $\underset{\to}{\boldsymbol{M}}$ coincide with the canonically orthogonalized set $\boldsymbol{\Lambda = VUd^{-1/2}}$ obtained from $\vec{\boldsymbol{M}}$.

The eigenvectors of $\underset{\to}{\boldsymbol{M}}$ represent the so called 'principal components' [6], or the uncorrelated vectors which are ordered so that the first few retain most of the variation present in the original vectors. We have used the common criterion of sum of the projection-squares of the given vectors on the principal components [5, 11]. It is maximum along the first principal component. Thus, we found that, the principal component analysis (PCA) based on $\underset{\to}{\boldsymbol{M}}$ is the same as the canonical orthogonalization (CO) based on $\vec{\boldsymbol{M}}$ in the case of 3 vectors in 3-dimensions ($\boldsymbol{3\text{-}d}$). This should hold in general for $\boldsymbol{n}$ vectors in $\boldsymbol{n}$-dimensions.

Note that $\underset{\to}{\boldsymbol{M}}$ has information on overlap of pairs of features of the given vectors. Often one wants to look at the features in the background of the features possessed by all the vectors (e.g. feature no.1 in the background of this feature in all the vectors). Then one considers the correlation matrix $\underset{\to}{\boldsymbol{C}}$ obtained from $\underset{\to}{\boldsymbol{M}}$ by using normalized $\underset{\to}{\boldsymbol{v}}_{\boldsymbol{i}}$'s $(= \underset{\to}{\boldsymbol{v}}_{\boldsymbol{i}}/ \mid \underset{\to}{\boldsymbol{v}}_{\boldsymbol{i}} \mid)$ in place of $\underset{\to}{\boldsymbol{v}}_{\boldsymbol{i}}$'s. This amounts to starting with $\boldsymbol{V^T}$ and normalizing each of its columns. We call it as $\tilde{\boldsymbol{V}}$. We have applied the canonical orthogonalization to $\tilde{\boldsymbol{V}}^{\boldsymbol{T}}\tilde{\boldsymbol{V}}$ and the principal component analysis to $\tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^{\boldsymbol{T}}$. We found that the eigenvectors of $\tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^{\boldsymbol{T}}$ coincide with the vectors in the set $\tilde{\boldsymbol{V}}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{d}}^{-\mathbf{1/2}}$. However two of the three eigenvalues change a little bit as compared to the earlier values corresponding to $\boldsymbol{V^TV}$ and $\boldsymbol{VV^T}$.

Table 3.1: CO and PCA of 3 vectors in 3-dimension (each column is a vector)

$$\boldsymbol{V} = \begin{pmatrix} \mathbf{12} & \mathbf{5} & \mathbf{7} \\ \mathbf{19} & \mathbf{11} & \mathbf{9} \\ \mathbf{23} & \mathbf{37} & \mathbf{8} \end{pmatrix} \xrightarrow{Normalization} \begin{pmatrix} \mathbf{.3732} & \mathbf{.1285} & \mathbf{.5026} \\ \mathbf{.5909} & \mathbf{.2826} & \mathbf{.6462} \\ \mathbf{.7153} & \mathbf{.9506} & \mathbf{.5744} \end{pmatrix}$$

$\boldsymbol{V^TV}$:

$$\vec{M} = \begin{pmatrix} \mathbf{1} & \mathbf{.8949} & \mathbf{.9802} \\ \mathbf{.8949} & \mathbf{1} & \mathbf{.7932} \\ \mathbf{.89802} & \mathbf{.7932} & \mathbf{1} \end{pmatrix}$$

Eigenvalues; Eigenvectors

$$\boldsymbol{d} = \begin{pmatrix} \mathbf{2.7810} \\ \mathbf{.2178} \\ \mathbf{.0013} \end{pmatrix}; \boldsymbol{U} = \begin{pmatrix} \mathbf{.5974} & \mathbf{-.1760} & \mathbf{-.7824} \\ \mathbf{.5571} & \mathbf{.7929} & \mathbf{.2470} \\ \mathbf{.5769} & \mathbf{-.5834} & \mathbf{.5717} \end{pmatrix}$$

$\boldsymbol{VUd^{-1/2}}$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{.3505} & \mathbf{-.5508} & \mathbf{.7575} \\ \mathbf{.5296} & \mathbf{-.5505} & \mathbf{-.6453} \\ \mathbf{.7725} & \mathbf{.6273} & \mathbf{.0988} \end{pmatrix}$$

$$\boldsymbol{(V^T\Lambda)_{max} = 2.7810 = d_{max}}$$

$\boldsymbol{VV^T}$:

$$\underset{\rightarrow}{M} = \begin{pmatrix} \mathbf{.4083} & \mathbf{.5815} & \mathbf{.6777} \\ \mathbf{.5815} & \mathbf{.8465} & \mathbf{1.0642} \\ \mathbf{.6777} & \mathbf{1.0624} & \mathbf{1.7451} \end{pmatrix}$$

Eigenvalues; Eigenvectors

$$\boldsymbol{d} = \begin{pmatrix} \mathbf{2.7810} \\ \mathbf{.2178} \\ \mathbf{-.0013} \end{pmatrix}; \boldsymbol{U'} = \begin{pmatrix} \mathbf{.3505} & \mathbf{-.5508} & \mathbf{.7575} \\ \mathbf{.5296} & \mathbf{-.5505} & \mathbf{-.6453} \\ \mathbf{.7725} & \mathbf{.6273} & \mathbf{.0988} \end{pmatrix}$$

$$\boldsymbol{(V^TU')_{max} = 2.7810 = d_{max}}$$

We have considered the general case of $\boldsymbol{N} > \boldsymbol{n}$ to study how canonical orthogonalization can be applied in this case and to further analyse the equivalence of PCA and the CO. For this purpose, we have considered a set of 24 vectors (i.e. $\boldsymbol{N} = \mathbf{24}$) in 3-dimensions (i.e. $\boldsymbol{n} = \mathbf{3}$). An arbitrary set $\boldsymbol{W}$ of normalized vectors is listed in Table 3.2. The SSCP matrix $\boldsymbol{W}\boldsymbol{W}^{\boldsymbol{T}} \equiv \underset{\rightarrow}{\boldsymbol{M}}$ is constructed. The eigenvectors of $\underset{\rightarrow}{\boldsymbol{M}}$ represent the mutually orthogonal principal components. The sum of projection squares of $\vec{\boldsymbol{v}}^{\boldsymbol{i}}$'s maximises along one of the eigenvectors and is also equal to the largest eigenvalue of $\underset{\rightarrow}{\boldsymbol{M}}$. This eigenvector represents the first principal component.

It is clear that the canonical orthogonalization can not be applied to the given 24 vectors in 3-$\boldsymbol{d}$. Taking our cue from the example of 3 vectors in 3-$\boldsymbol{d}$ we will consider three vectors $\underset{\rightarrow}{\boldsymbol{v}_i}$'s in 24-dimensions, i.e. the $\mathbf{24} \times \mathbf{3}$ matrix $\tilde{\boldsymbol{W}}^{\boldsymbol{T}}$ as shown in Table 3.3. Note that the $\underset{\rightarrow}{\boldsymbol{v}_i}$'s have to be normalized as 24-$\boldsymbol{d}$ vectors while $\vec{\boldsymbol{v}}^{\boldsymbol{i}}$'s are normalized as 3-$\boldsymbol{d}$ vectors. Such normalized vectors are $\underset{\rightarrow}{\tilde{\boldsymbol{w}}_i}$'s and $\vec{\boldsymbol{w}}_{\boldsymbol{i}}$'s respectively. Note that the *tilde*, ($\tilde{}$) represents the difference in the mode of normalization. Thus, $\tilde{\boldsymbol{W}}^{\boldsymbol{T}}$ is not the simple transpose of $\boldsymbol{W}$. Since now the number of $\underset{\rightarrow}{\tilde{\boldsymbol{w}}_i}$'s is less than the dimensionality of the space, namely 24, we have computed $\vec{\boldsymbol{M}}$ as $(\tilde{\boldsymbol{W}}^{\boldsymbol{T}})^{\boldsymbol{T}}(\tilde{\boldsymbol{W}}^{\boldsymbol{T}})$, which turns out to be $\tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}}^{\boldsymbol{T}}$ and therefore represented as $\underset{\rightarrow}{\tilde{\boldsymbol{M}}}$ – this will be a $\mathbf{3} \times \mathbf{3}$ matrix. We have computed the eigenvalues and eigenvectors of $\underset{\rightarrow}{\tilde{\boldsymbol{M}}}$ and then the canonical orthonormal basis set $\boldsymbol{\Lambda} = \tilde{\boldsymbol{W}}^{\boldsymbol{T}}\boldsymbol{U}\boldsymbol{D}^{-\mathbf{1/2}}$. The projection-squares of $\underset{\rightarrow}{\tilde{\boldsymbol{w}}_i}$'s on $\underset{\rightarrow}{\boldsymbol{\lambda}_i}$'s add up to the eigenvalues of $\underset{\rightarrow}{\tilde{\boldsymbol{M}}}$, with the maximum being along $\underset{\rightarrow}{\boldsymbol{\lambda}_{\mathbf{1}}}$. The eigenvectors of $\underset{\rightarrow}{\boldsymbol{M}}$, which characterize the directions of the principal components, and the eigenvectors of $\underset{\rightarrow}{\tilde{\boldsymbol{M}}}$, which mark the directions of the canonically orthogonalized bases, are aligned parallel to each other within an accuracy of less than **4%**. Consequently, the PCA as applied to $\underset{\rightarrow}{M}$, is equivalent to the CO as applied to $\underset{\rightarrow}{\tilde{M}}$. These results are shown in Tables 3.4 and 3.5 respectively.

Table 3.2: 24 vectors in 3-dimensions (each column represents a vector)

$\boldsymbol{V}=$

17	**12**	**34**	**85**	**97**	**14**	**13**	**76**
21	**34**	**56**	**53**	**43**	**27**	**33**	**64**
93	**67**	**89**	**31**	**18**	**49**	**69**	**42**

19	**52**	**32**	**57**	**18**	**67**	**11**	**81**
28	**65**	**15**	**91**	**33**	**93**	**32**	**43**
37	**71**	**18**	**12**	**71**	**21**	**64**	**9**

23	**7**	**15**	**91**	**8**	**36**	**47**	**92**
8	**12**	**37**	**43**	**15**	**17**	**31**	**83**
15	**71**	**48**	**5**	**19**	**43**	**75**	**78**

$$\overset{Normalization(columnwise)}{\downarrow}$$

$\mathbf{W} =$

.1755	**.1577**	**.3077**	**.8106**	**.9013**	**.2428**	**.1676**	**.7045**
.2168	**.4469**	**.5067**	**.5055**	**.3996**	**.4682**	**.4254**	**.5933**
.9603	**.8806**	**.8053**	**.2956**	**.1673**	**.8496**	**.8894**	**.3894**

.3789	**.4753**	**.8068**	**.5276**	**.2241**	**.5750**	**.1519**	**.8970**
.5584	**.5941**	**.3782**	**.8422**	**.4108**	**.7981**	**.4420**	**.4666**
.7379	**.6490**	**.4538**	**.1111**	**.8838**	**.1802**	**.8840**	**.0977**

.8042	**.0968**	**.2403**	**.9030**	**.3138**	**.6143**	**.5012**	**.6284**
.2797	**.1659**	**.5926**	**.4267**	**.5883**	**.2901**	**.3306**	**.5669**
.5245	**.9814**	**.7688**	**.0496**	**.7452**	**.7338**	**.7997**	**.5327**

Table 3.3: 3 vectors in 24-dimensions (each column represents a vector)

$$\boldsymbol{V^T} = \begin{pmatrix} \mathbf{17} & \mathbf{21} & \mathbf{93} \\ \mathbf{12} & \mathbf{34} & \mathbf{67} \\ \mathbf{34} & \mathbf{56} & \mathbf{89} \\ \mathbf{85} & \mathbf{53} & \mathbf{31} \\ \mathbf{97} & \mathbf{43} & \mathbf{18} \\ \mathbf{14} & \mathbf{27} & \mathbf{49} \\ \mathbf{13} & \mathbf{33} & \mathbf{69} \\ \mathbf{76} & \mathbf{64} & \mathbf{42} \\ \mathbf{19} & \mathbf{28} & \mathbf{37} \\ \mathbf{52} & \mathbf{65} & \mathbf{71} \\ \mathbf{32} & \mathbf{15} & \mathbf{18} \\ \mathbf{57} & \mathbf{91} & \mathbf{12} \\ \mathbf{18} & \mathbf{33} & \mathbf{71} \\ \mathbf{67} & \mathbf{93} & \mathbf{21} \\ \mathbf{11} & \mathbf{32} & \mathbf{64} \\ \mathbf{81} & \mathbf{43} & \mathbf{9} \\ \mathbf{23} & \mathbf{8} & \mathbf{15} \\ \mathbf{7} & \mathbf{12} & \mathbf{71} \\ \mathbf{15} & \mathbf{37} & \mathbf{48} \\ \mathbf{91} & \mathbf{43} & \mathbf{5} \\ \mathbf{8} & \mathbf{15} & \mathbf{19} \\ \mathbf{36} & \mathbf{17} & \mathbf{43} \\ \mathbf{47} & \mathbf{31} & \mathbf{75} \\ \mathbf{92} & \mathbf{83} & \mathbf{78} \end{pmatrix} \overset{Normalization}{\longrightarrow} \boldsymbol{\tilde{W}^T} = \begin{pmatrix} \mathbf{.0670} & \mathbf{.0910} & \mathbf{.3532} \\ \mathbf{.0473} & \mathbf{.1474} & \mathbf{.2545} \\ \mathbf{.1341} & \mathbf{.2427} & \mathbf{.3380} \\ \mathbf{.3352} & \mathbf{.2297} & \mathbf{.1177} \\ \mathbf{.3825} & \mathbf{.1864} & \mathbf{.0684} \\ \mathbf{.0552} & \mathbf{.1170} & \mathbf{.1861} \\ \mathbf{.0513} & \mathbf{.1430} & \mathbf{.2621} \\ \mathbf{.2997} & \mathbf{.2774} & \mathbf{.1595} \\ \mathbf{.0749} & \mathbf{.1214} & \mathbf{.1405} \\ \mathbf{.2050} & \mathbf{.2818} & \mathbf{.2697} \\ \mathbf{.1262} & \mathbf{.0650} & \mathbf{.0684} \\ \mathbf{.2248} & \mathbf{.3945} & \mathbf{.0456} \\ \mathbf{.0710} & \mathbf{.1430} & \mathbf{.2697} \\ \mathbf{.2642} & \mathbf{.4031} & \mathbf{.0798} \\ \mathbf{.0434} & \mathbf{.1387} & \mathbf{.2431} \\ \mathbf{.3194} & \mathbf{.1864} & \mathbf{.0342} \\ \mathbf{.0907} & \mathbf{.0347} & \mathbf{.0570} \\ \mathbf{.0276} & \mathbf{.0520} & \mathbf{.2697} \\ \mathbf{.0591} & \mathbf{.1604} & \mathbf{.1823} \\ \mathbf{.3588} & \mathbf{.1864} & \mathbf{.0190} \\ \mathbf{.0315} & \mathbf{.0650} & \mathbf{.0722} \\ \mathbf{.1420} & \mathbf{.0737} & \mathbf{.1633} \\ \mathbf{.1853} & \mathbf{.1344} & \mathbf{.2849} \\ \mathbf{.3628} & \mathbf{.3598} & \mathbf{.2963} \end{pmatrix}$$

Table 3.4: Canonical Orthogonalization

$$(\tilde{\boldsymbol{W}}^{\boldsymbol{T}})^{\boldsymbol{T}}(\tilde{\boldsymbol{W}}^{\boldsymbol{T}}) = \tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}}^{\boldsymbol{T}} \equiv \underset{\to}{\tilde{\boldsymbol{M}}} = \begin{pmatrix} \mathbf{1} & \mathbf{.8857} & \mathbf{.5805} \\ \mathbf{.8857} & \mathbf{1} & \mathbf{.7317} \\ \mathbf{.5805} & \mathbf{.7317} & \mathbf{1} \end{pmatrix}$$
Eigenvalues $\boldsymbol{D} = \begin{pmatrix} \mathbf{2.4723} \\ \mathbf{.4398} \\ \mathbf{.0879} \end{pmatrix}$ & Eigenvectors $\boldsymbol{U} = \begin{pmatrix} \mathbf{.5804} & \mathbf{-.5565} & \mathbf{.5945} \\ \mathbf{.6146} & \mathbf{-.1796} & \mathbf{-.7681} \\ \mathbf{.5343} & \mathbf{.8122} & \mathbf{.2378} \end{pmatrix}$
$$(\tilde{\boldsymbol{W}}\boldsymbol{\Lambda})^{\mathbf{2}} = \left(\tilde{\boldsymbol{W}}\left[\tilde{\boldsymbol{W}}^{\boldsymbol{T}}\boldsymbol{U}\boldsymbol{D}^{-1/2}\right]\right)^{\mathbf{2}} = \begin{pmatrix} \mathbf{.8327} & \mathbf{.1362} & \mathbf{.0311} \\ \mathbf{.9339} & \mathbf{.0142} & \mathbf{.0519} \\ \mathbf{.7057} & \mathbf{.2894} & \mathbf{.0705} \end{pmatrix}$$
Sum of projection-squares of 3 vectors (in 24-$\boldsymbol{d}$) on $\boldsymbol{\Lambda}$ $= \text{sum of column elements in } (\tilde{\boldsymbol{W}}\boldsymbol{\Lambda})^{\mathbf{2}} = \begin{pmatrix} \mathbf{2.4723} \\ \mathbf{.4398} \\ \mathbf{.0879} \end{pmatrix}$

Table 3.5: Principal Component Analysis

$$\boldsymbol{WW^T} \equiv \underset{\rightarrow}{\boldsymbol{M}} = \begin{pmatrix} \mathbf{7.3297} & \mathbf{5.5624} & \mathbf{5.2763} \\ \mathbf{5.5624} & \mathbf{5.9067} & \mathbf{6.2432} \\ \mathbf{5.2763} & \mathbf{6.2432} & \mathbf{10.7636} \end{pmatrix}$$

Eigenvalues $\boldsymbol{d} = \begin{pmatrix} \mathbf{19.6485} \\ \mathbf{3.6672} \\ \mathbf{.6843} \end{pmatrix}$ & Eigenvectors $\boldsymbol{E} = \begin{pmatrix} \mathbf{.5237} & \mathbf{-.6959} & \mathbf{.4914} \\ \mathbf{.5189} & \mathbf{-.1969} & \mathbf{-.8318} \\ \mathbf{.6756} & \mathbf{.6906} & \mathbf{.2580} \end{pmatrix}$

Sum of projection-squares of the given 24-vectors (in 3-$\boldsymbol{d}$) on $\boldsymbol{E}$

$$= \text{sum of column elements in } (\boldsymbol{W^T E})^2 = \begin{pmatrix} \mathbf{19.6485} \\ \mathbf{3.6672} \\ \mathbf{.6843} \end{pmatrix}$$

Table 3.6: Hierarchical scheme of canonical orthogonalization of 24 vectors

$$\boldsymbol{W} \equiv \boldsymbol{W}^{(0)} = \left(\underbrace{\begin{matrix} .1755 & .1577 & .3077 \\ .2168 & .4469 & .5067 \\ .9603 & .8806 & .8053 \end{matrix}}_{1} \; \underbrace{\begin{matrix} .8106 & .9013 & .2428 \\ .5055 & .3996 & .4682 \\ .2956 & .1673 & .8496 \end{matrix}}_{2} \; \underbrace{\begin{matrix} .1676 & .7045 & .3789 \\ .4254 & .5933 & .5584 \\ .8894 & .3894 & .7379 \end{matrix}}_{3} \; \underbrace{\begin{matrix} .4753 & .8068 & .5276 \\ .5941 & .3782 & .8422 \\ .6490 & .4538 & .1111 \end{matrix}}_{4} \right.$$

group# 1 2 3 4

$$\left. \underbrace{\begin{matrix} .8042 & .0968 & .2403 \\ .2797 & .1659 & .5926 \\ .5245 & .9814 & .7688 \end{matrix}}_{5} \; \underbrace{\begin{matrix} .9030 & .3138 & .6143 \\ .4267 & .5883 & .2901 \\ .0496 & .7452 & .7338 \end{matrix}}_{6} \; \underbrace{\begin{matrix} .5012 & .6284 & .2241 \\ .3306 & .5669 & .4108 \\ .7997 & .5327 & .8838 \end{matrix}}_{7} \; \underbrace{\begin{matrix} .5750 & .1519 & .8970 \\ .7981 & .4420 & .4666 \\ .1802 & .8840 & .0977 \end{matrix}}_{8} \right)_{\mathbf{3\times 24}}$$

group# 5 6 7 8

$\boldsymbol{\lambda}_i^{(1)}$'s that maximize the sum of projection-squares in each group at stage (1)

↓

$$\boldsymbol{W}^{(1)} = \begin{pmatrix} \underbrace{\begin{matrix} .2161 & .5949 & .4345 \\ .3954 & .4678 & .5526 \\ .8927 & .6537 & .7112 \end{matrix}} & \underbrace{\begin{matrix} .6384 & .3278 & .7312 \\ .6369 & .5833 & .3585 \\ .4322 & .7431 & .5803 \end{matrix}} & \underbrace{\begin{matrix} .4991 & .5902 \\ .6038 & .4009 \\ .6215 & .7006 \end{matrix}} \end{pmatrix}_{\mathbf{3\times 8}}$$

$\boldsymbol{\lambda}_i^{(2)}$'s that maximize the sum of projection-squares in each group at stage (2)

↓

$$\boldsymbol{W}^{(2)} = \begin{pmatrix} .4243 & .5852 & .5485 \\ .4820 & .5428 & .5059 \\ .7665 & .6024 & .6657 \end{pmatrix}_{\mathbf{3\times 3}}$$

$\boldsymbol{\lambda}_i^{(3)}$ that maximizes the sum of projection-squares in this group at stage (3)

↓

$$\boldsymbol{\lambda}_1^{(3)} = \begin{pmatrix} .5521 \\ .5128 \\ .6815 \end{pmatrix}$$

Sum of projection-squares of 24 vectors in 3-dimension on

$$\boldsymbol{\lambda}_1^{(3)} = \left(\boldsymbol{W}^T \, \boldsymbol{\lambda}_1^{(3)}\right)^{\mathbf{2}} = \mathbf{19.6471.}$$

We have tried the hierarchical way of circumventing the problem of linear dependence of the same set of given vectors $\boldsymbol{W}$ in order to apply the canonical orthogonalization scheme. The 24 vectors in 3-$\boldsymbol{d}$ were divided into 8 groups of linearly-independent vectors and the canonical orthogonalization is applied to each one of them. From each group we have picked out that basis vector along which the projection-squares maximized in that group. We have again divided these 8 vectors into three sets comprising 3, 3 and 2 vectors and applied canonical orthogonalization to these three groups and picked out 3 main basis vectors along which the projection-squares maximized in each group. Finally we have applied the canonical orthogonalization to these three vectors and came up with the final canonically orthogonalized vector along which the sum of projection-squares of these three vectors is the maximum.

The projection-squares of the 24 given vectors on this final vector add up to 19.6471 which is very close to 19.6485, the largest eigenvalue of $\underset{\rightarrow}{M}$ constructed from the given 24 vectors for doing the PCA. The results of this scheme are summarized in Table 3.6. Thus we conclude that the hierarchical scheme of CO works successfully in determining the direction along which a given set of vectors projects maximally even though it may be a set of linearly-dependent vectors.

In sum we have found that it is possible to circumvent the restriction that a given set of vectors should be linearly-independent in order to apply Löwdin's canonical orthogonalization to find a direction in which the given vectors project maximally. We have made two proposals and have tested them numerically. It is also found that when the CO is used in these fashions, it is equivalent to the PCA, but the hierarchical scheme of CO proposed here is computationally more economical than the PCA. The same procedures can be very well applicable to the case of symmetric orthogonalization as well. We have extended this work to the *symmetric* orthogonalization for different data sets and tested its limitations from the computational point of view.

Chapter 4

New Classes of Orthogonal Polynomials

4.1 Introduction

The subject of orthogonal polynomials has extensive applications in physics, probability, statistics and mathematics. The computer revolution has increased the activities in approximation theory and numerical analysis and with that the orthogonal polynomials have come in focus again. The methods of generating orthogonal polynomials, their asymptotic limits, zeros and their applications to various branches have been developed regularly. There have been also new kind of orthogonal polynomials and their uses to numerical analysis, least-square fitting, computational techniques, etc. have been studied.

In this chapter, we have presented two new classes of orthogonal polynomials [13] which have novel properties. They are looking for problems where their novel properties would play a role.

A simple way of obtaining the known orthogonal polynomials is by applying Gram-Schmidt orthogonalization procedure [10, 2] to the monomials such that the orthonormality condition is satisfied with respect to different weight functions on different intervals – a combination of weight and interval yields a particular orthogonal polynomial sequence. If this procedure is repeated with Löwdin's democratic orthogonalizations [5], we find two new classes of orthogonal polynomial sets, a set for each orthogonality condition. These polynomials possess interesting new properties inherent in the two orthogonalization procedures due to Löwdin [1]. The new properties are however for a small price as we will describe them in the subsequent sections. They are new in that they possess novel properties in terms of their inner products with the monomials. Each class comprises sets of orthogonal polynomials which satisfy orthogonality conditions with respect to a weight function on a certain interval.

We have applied the Löwdin orthogonalization procedures to a set of monomials $\boldsymbol{\{x^N\}, N = 0, 1, 2, 3, \cdots, \infty}$, for the classical orthogonal polynomials such as the Legendre, Hermite, Laguerre, Chebyshev I, Chebyshev II, Gegenbauer, Jacobi and Bessel polynomials with their respective weight functions and limits of integration. We have named the newly obtained polynomials with the Löwdin name as prefix. The newly obtained polynomials satisfy the curious geometric property of Schweinler-Wigner matrix.

4.2 Classical Orthogonal Polynomials

The classical orthogonal polynomials such as the Legendre, Hermite, Laguerre etc. are obtained from a set of monomials/functions $\{x^N\}, N = 0, 1, \cdots, \infty$, which are orthogonal and orthonormal in a given interval $[a, b]$ with respect to a weight function $w(x)$. They can be generated one by one using the Gram-Schmidt orthogonalization procedure using a given set of monomials/functions. These set of monomials/functions can be taken one at a time in an increasing order. The polynomials obtained are a direct consequence of the choice of particular interval and weight function. The use of monomials with different choice of intervals and weight functions leads to different sets of orthogonal polynomials. In fact, these polynomials are solutions of particular differential equations. The choice of interval, weight function, orthogonality and orthonormalization conditions for different orthogonal polynomials are listed in Table 4.1.

4.3 An Overview of Löwdin Methods

We know that the Gram-Schmidt method orthogonalizes a given set of linearly independent vectors or monomials in a *sequence*, which can be arbitrary in the case of vectors but has to be in the increasing order in the case of monomials. The two methods due to Löwdin are however *democratic* in the sense that they handle all the given vectors/monomials simultaneously and treat them on equal footing. The derivation of the two procedures is described in detail in Chapter 2 and in references [1, 5]. However, we have given the gist of their derivation here for completeness and to highlight the interesting properties they possesses.

Table 4.1: List of the choice of interval, weight function, orthogonality and orthonormalization conditions for different classical orthogonal polynomials

Name of the Polynomials	Interval $[a, b]$	Weight Function $w(x)$	Orthogonality Condition	Orthognormalization Condition
Legendre	[-1, 1]	1	$\int_{-1}^{1} L_m(x)\ L_n(x)\ dx$	$\int_{-1}^{1} {L_n}^2(x)\ dx = 1$
Hermite	[-∞, ∞]	e^{-x^2}	$\int_{-\infty}^{\infty} H_m(x)\ H_n(x)\ dx$	$\int_{-\infty}^{\infty} {H_n}^2(x)\ dx = 1$
Laguerre	[0, ∞]	e^{-x}	$\int_{0}^{\infty} \mathcal{L}_m(x)\ \mathcal{L}_n(x)\ dx$	$\int_{0}^{\infty} {\mathcal{L}_n}^2(x)\ dx = 1$
Chebyshev	[-1, 1]	$(1-x^2)^{-1/2}$	$\int_{-1}^{1} C_m(x)\ C_n(x)\ dx$	$\int_{-1}^{1} {C_n}^2(x)\ dx = 1$
Chebyshev II	[0, -1]	$(1-x^2)^{1/2}$	$\int_{-1}^{1} C_m(x)\ C_n(x)\ dx$	$\int_{-1}^{1} {C_n}^2(x)\ dx = 1$
Gegenbauer	[-1, 1]	$(1-x^2)^{\alpha-1/2}$	$\int_{-1}^{1} G_m(x)\ G_n(x)\ dx$	$\int_{-1}^{1} {G_n}^2(x)\ dx = 1$
Jacobi	[-1, 1]	$(1-x)^{\alpha}\ (1+x)^{\beta}$	$\int_{-1}^{1} J_m(x)\ J_n(x)\ dx$	$\int_{-1}^{1} {J_n}^2(x)\ dx = 1$
Bessel	[0, ∞]	x	$\int_{0}^{\infty} B_m(x)\ B_n(x)\ dx$	$\int_{0}^{\infty} {B_n}^2(x)\ dx = 1$

Consider a set of linearly independent vectors $\boldsymbol{V} \equiv \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \cdots, \vec{v}_N\}$, in a $\boldsymbol{N}$-dimensional space which can in general be a complex vector space. We can define a general non-singular linear transformation $\boldsymbol{A}$ for the basis $\boldsymbol{V}$ to go to a new basis $\boldsymbol{Z}$:

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{A}\,. \tag{4.1}$$

The set $\boldsymbol{Z}(\equiv \{\vec{z}_\kappa\})$ will be orthonormal if

$$< \boldsymbol{Z}|\boldsymbol{Z} >=< \boldsymbol{V}\boldsymbol{A}|\boldsymbol{V}\boldsymbol{A} >= \boldsymbol{A}^\dagger < \boldsymbol{V}|\boldsymbol{V} > \boldsymbol{A} = \boldsymbol{A}^\dagger \boldsymbol{M}\boldsymbol{A} = \boldsymbol{I}\,, \tag{4.2}$$

where $\boldsymbol{M}$ is a Hermitian metric matrix of the given basis $\boldsymbol{V}$. The substitution

$$\boldsymbol{A} = \boldsymbol{M}^{-1/2}\boldsymbol{B}\ , \tag{4.3}$$

where $\boldsymbol{B}$ is an arbitrary unitary matrix, leads to the general solution of the orthogonalization problem.

The specific choice $\boldsymbol{B} = \boldsymbol{I}$ gives the *symmetric orthogonalization*,

$$\boldsymbol{Z} \equiv \boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{M}^{-1/2}. \tag{4.4}$$

The choice $\boldsymbol{B} = \boldsymbol{U}$, where $\boldsymbol{U}$ diagonalizes the metric matrix $\boldsymbol{M}$, gives the *canonical orthogonalization*,

$$\boldsymbol{Z} \equiv \boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{U}\boldsymbol{d}^{-1/2}. \tag{4.5}$$

4.4 Orthogonalization of Monomials

The curious geometric properties of the Löwdin orthogonalization methods [5] prompted us to try them on the monomials, which also form linearly-independent sets. Since these methods treat the given entities that need to be orthogonalized all in one go without any preferred order among them, one has to first pick a set of monomials, quite arbitrarily, and choose an orthonormality condition to be invoked with respect to a weight function on some interval – e.g. the weight could be $\mathbf{1}$ and interval $[\mathbf{-1, 1}]$ as in the case of Legendre polynomials, or the weight could be $\boldsymbol{e^{-x^2}}$ and interval $[\boldsymbol{-\infty, \infty}]$ as in Hermite polynomials, or the weight could be $\boldsymbol{e^{-x}}$ and interval $[\boldsymbol{0, \infty}]$ as in the case of Laguerre and so on. The orthonormality condition will be applied to the metric matrix $\boldsymbol{M}$ of the type in equations (4.2) and (4.3), but now it involves the monomials of the chosen set. A combination of weight and interval yields a particular orthogonal polynomial sequence. They are new in that they possess novel properties of the type described in section 2.8 of Chapter 2 in terms of their monomials.

4.5 Relationship between Orthogonal Polynomials and Monomials

Suppose the set of monomials is $\mathcal{M} = \{1, x^2, x^3, x^4, ..., x^N\}$ and a new set of orthogonal polynomials determined from them as above, is

$$\mathcal{P} = \{P_0(x), P_1(x), P_2(x), ..., P_N(x)\}.$$

If the set $\{P_i(x)\}$ is obtained through the *symmetric orthogonalization* and is called $\mathcal{S} = \{S_i(x)\}$, then its elements will satisfy,

$$\sum_i \left(x^k, S_i(x)\right)^2 = \sum_i \left(x^i, S_k(x)\right)^2, \text{for all } k\text{'s}. \tag{4.6}$$

The symmetric properties of the monomials are preserved in this set of orthogonal polynomials. This means that the elements of $\mathcal{S}$ will be closest to those of $\mathcal{M}$ in the least-squares sense that the i^{th} element S_i of $\mathcal{S}$ projects maximally on the i^{th} element M_i of $\mathcal{M}$. And, the sum of projection squares of all M_i's on a $S_k (k \neq i)$ equals the sum of projection squares of M_k on all S_i's.

If the set $\mathcal{P}$ is obtained through the *canonical orthogonalization* and called $\mathcal{C} = \{C_i\}$, then its elements will satisfy,

$$\sum_k \left(\sum_i \left(x^i, C_k(x)\right)^2\right)^2 = \textit{the maximum}, \tag{4.7}$$

for *any* set of orthogonal polynomials that can be generated from the given set of monomials.

Considering the set of monomials, we have applied the Löwdin's Orthogonalization methods. Then the monomials $\mathcal{M}$ are used with the orthogonality property of different orthogonal polynomials to construct the Hermitian metric matrix M.

The eigenvalues and eigenvectors of $\boldsymbol{M}$ are calculated to find the symmetric and canonical orthonormal set. Two new sets of finite and infinite orthogonal polynomials are obtained by applying the symmetric and canonical orthogonalizations. One higher monomial is added successively to apply these methods. These new orthogonal polynomials change every time the set of monomials changes because each time the entire set of monomials is used for generating the new set. These new sets of orthogonal polynomials exhibit some patterns of repetition within a set as well as between consecutive sets.

The most conspicuous property is that only those monomials that are mutually nonorthogonal combine to form Löwdin's symmetric and canonical orthogonal polynomials. The standard orthogonal polynomials such as Legendre, Chebyshev type I, Chebyshev type II, Hermite and Jacobi exhibit this property. The case of Löwdin orthogonal polynomials is different in the following sense. The set of monomials for the given $\boldsymbol{N}$ divides into two subsets of mutually nonorthogonal monomials. These polynomials use the monomials from these sets alternatively, and in each case all the monomials from a subset participate in forming orthogonal polynomials. For example, for $\boldsymbol{N = 4}$ the two subsets of mutually nonorthogonal monomials are $\boldsymbol{\{1, x^2, x^4\}}$ and $\boldsymbol{\{x, x^3\}}$ and the five orthogonal polynomials in the symmetric as well as canonical cases use all the elements from the two subsets alternatively. Since under the Bessel and Laguerre inner product condition all the monomials are nonorthogonal, all participate in the formation of all the Löwdin-Bessel and Löwdin-Laguerre orthogonal polynomials.

We have observed another interesting property which is common to Löwdin-Legendre, Löwdin-Chebyshev I and II, Löwdin-Jacobi and Löwdin-Hermite sets of orthogonal polynomials is that certain polynomials from one set, with a particular $\boldsymbol{N}$ value, repeat in the consecutive $(\boldsymbol{N}+\mathbf{1})$ set. In addition, there is a pattern to it; the $\mathbf{1}^{st}$ and $\mathbf{3}^{rd}$ in the $\boldsymbol{N}=\mathbf{2}$ set are the $\mathbf{1}^{st}$ and $\mathbf{3}^{rd}$ in the $\boldsymbol{N}=\mathbf{3}$ set, and the $\mathbf{2}^{nd}$ and$\mathbf{4}^{th}$ in $\boldsymbol{N}=\mathbf{3}$ are repeated as $\mathbf{2}^{nd}$ and $\mathbf{4}^{th}$ in the $\boldsymbol{N}=\mathbf{4}$ set. However, in the case of Löwdin-Bessel and Löwdin-Laguerre polynomials such repetitions are not there.

The most striking repetition property which is exhibited by the symmetric orthogonal polynomials is presented in Table 4.9 for all the new polynomials. It is inferred that the coefficients are repeated between polynomials consisting of either even-powered or odd-powered monomials, which respectively form nonorthogonal sets. Löwdin-Bessel and Löwdin-Laguerre are, of course, different because all monomials are nonorthogonal under their respective inner product condition and therefore showed this property more strongly. The symmetric orthogonal polynomials $\{\boldsymbol{S}_n(\boldsymbol{x})\}$ appear to have the form

$$\boldsymbol{S}_n(\boldsymbol{x}) = \boldsymbol{a}_n + \boldsymbol{a}_{n+1} + \cdots + \boldsymbol{a}_N \boldsymbol{x}^{N-n} + \boldsymbol{a}_0 \boldsymbol{x}^{N-n+1} + \cdots + \boldsymbol{a}_{n-1} \boldsymbol{x}^{N} \quad (4.8)$$

where, all the terms for $\boldsymbol{n}=\mathbf{0},\mathbf{1},\mathbf{2},\cdots,\boldsymbol{N}$ are present in all the Löwdin-Bessel and Löwdin-Laguerre polynomials. But in the case of Löwdin-Legendre, Löwdin-Chebyshev I, Löwdin-Chebyshev II, Löwdin-Jacobi and Löwdin-Hermite polynomials the terms only with even (odd) powers of $\boldsymbol{x}$ are present with even (odd) $\boldsymbol{n}$. However, the undesirable feature is that all the coefficients with the same suffix $\boldsymbol{n}$ appearing in different orthogonal polynomials do not necessarily have the same numerical value. But there is a pattern in the repetition of the values of the coefficient in the following matrices of coefficients for N=3:

$$\begin{array}{ccccc} & \boldsymbol{x^0} & \boldsymbol{x^1} & \boldsymbol{x^2} & \boldsymbol{x^3} \\ \boldsymbol{s_0}: & \boldsymbol{a_0'} & 0 & \boldsymbol{a_2} & 0 \\ \boldsymbol{s_1}: & 0 & \boldsymbol{a_2'} & 0 & \boldsymbol{a_0} \\ \boldsymbol{s_2}: & \boldsymbol{a_2} & 0 & \boldsymbol{a_0''} & 0 \\ \boldsymbol{s_3}: & 0 & \boldsymbol{a_0} & 0 & \boldsymbol{a_2''} \end{array} \qquad \begin{array}{ccccc} & \boldsymbol{x^0} & \boldsymbol{x^1} & \boldsymbol{x^2} & \boldsymbol{x^3} \\ \boldsymbol{s_0}: & \boldsymbol{a_0'} & \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} \\ \boldsymbol{s_1}: & \boldsymbol{a_1} & \boldsymbol{a_2'} & \boldsymbol{a_3'} & \boldsymbol{a_0} \\ \boldsymbol{s_2}: & \boldsymbol{a_2} & \boldsymbol{a_3'} & \boldsymbol{a_0''} & \boldsymbol{a_1} \\ \boldsymbol{s_3}: & \boldsymbol{a_3} & \boldsymbol{a_0} & \boldsymbol{a_1} & \boldsymbol{a_2''} \end{array}$$

The first matrix represents the arrangement of non-zero coefficient in the Löwdin-Legendre, Löwdin-Chebyshev I, Löwdin-Chebyshev II, Löwdin-Jacobi and Löwdin-Hermite polynomials and the second one represents those in the Löwdin-Bessel and Löwdin-Laguerre polynomials. The values of the coefficients are different in all the cases.

The symmetric polynomials are obtained by multiplying the $\boldsymbol{x^i}$'s at the top of the columns with the coefficients and adding them. The same symbol in two places within a matrix represents the same value, but those with and without primes and double primes represent different values. The new families of orthogonal polynomials are presented in Tables 4.2 to 4.8.

Table 4.2: Legendre $\{L_i\}$ and Löwdin-Legendre $\{L_i^{sy},\ L_i^{ca}\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Legendre $\{L_i\}$	$\{L_i^{sy}\}$	$\{L_i^{ca}\}$
$N = 2:$		
$\sqrt{\frac{1}{2}}$	$\rightarrow (0.8815 - 0.5899x^2)$	$\rightarrow (0.628 + 0.2274x^2)$
$\sqrt{\frac{3}{2}}\ x$	$(1.2247x)$	$(1.2247x)$
$\sqrt{\frac{5}{2}}\ \frac{1}{2}\ (3x^2 - 1)$	$\twoheadrightarrow (-0.5899 + 2.2973x^2)$	$\twoheadrightarrow (-0.8548 + 2.3609x^2)$
$N = 3:$		
$\sqrt{\frac{1}{2}}$	$\rightarrow (0.8815 - 0.5899x^2)$	$\rightarrow (0.628 + 0.2274x^2)$
$\sqrt{\frac{3}{2}}\ x$	$\hookrightarrow (2.311x - 2.0084x^3)$	$\hookrightarrow (0.8819x + 0.5568x^3)$
$\sqrt{\frac{5}{2}}\ \frac{1}{2}\ (3x^2 - 1)$	$\twoheadrightarrow (-0.5899 + 2.2973x^2)$	$\twoheadrightarrow (-0.8548 + 2.3609x^2)$
$\sqrt{\frac{7}{2}}\ \frac{1}{2}\ (5x^3 - 3x)$	$\mapsto (-2.0084x + 4.224x^3)$	$\mapsto (-2.932x + 4.6439x^3)$
$N = 4:$		
$\sqrt{\frac{1}{2}}$	$(0.9201 - 0.9036x^2$ $+0.3085x^4)$	$(0.5959 + 0.2249x^2$ $+0.1424x^4)$
$\sqrt{\frac{3}{2}}\ x$	$\hookrightarrow (2.311x - 2.0084x^3)$	$\hookrightarrow (0.8819x + 0.5568x^3)$
$\sqrt{\frac{5}{2}}\ \frac{1}{2}\ (3x^2 - 1)$	$(-0.9036 + 6.3692x^2$ $-5.2475x^4)$	$(-0.7748 + 1.2817x^2$ $+1.2176x^4)$
$\sqrt{\frac{7}{2}}\ \frac{1}{2}\ (5x^3 - 3x)$	$\mapsto (-2.0084x + 4.224x^3)$	$\mapsto (-2.932x + 4.6439x^3)$
$\sqrt{\frac{9}{2}}\ \frac{1}{8}\ (35x^4 - 30x^2 + 3)$	$(0.3085 - 5.2475x^2$ $+7.6499x^4)$	$(0.896 - 8.1992x^2$ $+9.2005x^4)$

Table 4.3: Hermite $\{H_i\}$ and Löwdin-Hermite $\{H_i^{sy},\ H_i^{ca}\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Hermite $\{H_i\}$	$\{H_i^{sy}\}$	$\{H_i^{ca}\}$
$N = 2:$		
$\sqrt{\frac{1}{\sqrt{\pi}}}$	$\rightarrow (0.8701 - 0.2986x^2)$	$\rightarrow (0.5021 + 0.392x^2)$
$\sqrt{\frac{2}{\sqrt{\pi}}}\ x$	$(1.0622x)$	$(1.0622x)$
$\sqrt{\frac{1}{2\sqrt{\pi}}}\ (2x^2 - 1)$	$\twoheadrightarrow (-0.2986 + 1.0194x^2)$	$\twoheadrightarrow (-0.7709 + 0.9873x^2)$
$N = 3:$		
$\sqrt{\frac{1}{\sqrt{\pi}}}$	$\rightarrow (0.8701 - 0.2986x^2)$	$\rightarrow (0.5021 + 0.392x^2)$
$\sqrt{\frac{2}{\sqrt{\pi}}}\ x$	$\hookrightarrow (1.608x - 0.4849x^3)$	$\hookrightarrow (0.2037x + 0.463x^3)$
$\sqrt{\frac{1}{2\sqrt{\pi}}}\ (2x^2 - 1)$	$\twoheadrightarrow (-0.2986 + 1.0194x^2)$	$\twoheadrightarrow (-0.7709 + 0.9873x^2)$
$\sqrt{\frac{1}{3\sqrt{\pi}}}\ (2x^3 - 3x)$	$\mapsto (-0.4849x + 0.7191x^3)$	$\mapsto (1.6671x - 0.7334x^3)$
$N = 4:$		
$\sqrt{\frac{1}{\sqrt{\pi}}}$	$(0.9118 - 0.4731x^2$ $+0.0507x^4)$	$(0.0383 + 0.0796x^2$ $+0.2649x^4)$
$\sqrt{\frac{2}{\sqrt{\pi}}}\ x$	$\hookrightarrow (1.608x - 0.4849x^3)$	$\hookrightarrow (0.2037x + 0.463x^3)$
$\sqrt{\frac{1}{2\sqrt{\pi}}}\ (2x^2 - 1)$	$(-0.4731 + 2.0205x^2$ $-0.4546x^4)$	$(0.7138 + 0.2085x^2$ $-0.1657x^4)$
$\sqrt{\frac{1}{3\sqrt{\pi}}}\ (2x^3 - 3x)$	$\mapsto (-0.4849x + 0.7191x^3)$	$\mapsto (1.6671x - 0.7334x^3)$
$\sqrt{\frac{1}{24\sqrt{\pi}}}\ (4x^4 - 12x^2 + 3)$	$(0.0507 - 0.4546x^2$ $+0.4085x^4)$	$(-0.7395 + 2.1127x^2$ $-0.5277x^4)$

Table 4.4: Laguerre $\{\mathcal{L}_i\}$ and Löwdin-Laguerre $\{\mathcal{L}_i^{sy},\ \mathcal{L}_i^{ca}\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Laguerre $\{\mathcal{L}_i\}$	$\{\mathcal{L}_i^{sy}\}$	$\{\mathcal{L}_i^{ca}\}$
$N = 2:$		
$\frac{1}{0!}$	$(1.515 - 0.8343x$ $+0.0939x^2)$	$(0.0174 + 0.0488x$ $+0.1902x^2)$
$\frac{1}{1!}\,(x-1)$	$(-0.8343 + 2.0364x$ $-0.3961x^2)$	$(-0.7621 - 0.5061x$ $+0.1995x^2)$
$\frac{1}{2!}\,(x^2 - 4x + 2)$	$(0.0939 - 0.3961x$ $+0.2903x^2)$	$(1.5553 - 2.1775x$ $+0.4171x^2)$
$N = 3:$		
$\frac{1}{0!}$	$(1.6191 - 1.1499x$ $+0.2364x^2 - 0.0143x^3)$	$(0.0003 + 0.0012x$ $+0.0061x^2 + 0.0362x^3)$
$\frac{1}{1!}\,(x-1)$	$(-1.1499 + 3.3506x$ $-1.2007x^2 + 0.099x^3)$	$(-0.1202 - 0.199x$ $-0.3623x^2 + 0.0685x^3)$
$\frac{1}{2!}\,(x^2 - 4x + 2)$	$(0.2364 - 1.2007x$ $+0.9939x^2 - 0.1219x^3)$	$(1.0515 + 0.2839x$ $-0.4926x^2 + 0.0639x^3)$
$\frac{1}{3!}\,(x^3 - 9x^2 + 18x - 6)$	$(-0.0143 + 0.099x$ $-0.1219x^2 + 0.0539x^3)$	$(-1.697 + 3.7256x$ $-1.4581x^2 + 0.133x^3)$

Table 4.5: Chebyshev I $\{\mathcal{C}_i - I\}$ and Löwdin-Chebyshev I $\{\mathcal{C}_i^{sy} - I,\ \mathcal{C}_i^{ca} - I\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Chebyshev I $\{\mathcal{C}_i - I\}$	$\{\mathcal{C}_i^{sy} - I\}$	$\{\mathcal{C}_i^{ca} - I\}$
$N = 2:$		
$\frac{1}{\sqrt{\pi}}$	$(0.8062 - 0.5535x^2)$	$(0.4367 + 0.2421x^2)$
$\sqrt{\frac{2}{\pi}}x$	$(0.7979x)$	$(-0.7979x)$
$\sqrt{\frac{8}{\pi}}(x^2 - \frac{1}{2})$	$(-0.5535 + 1.4976x^2)$	$(0.8750 - 1.5781x^2)$
$N = 3:$		
$\frac{1}{\sqrt{\pi}}$	$(0.8062 - 0.5535x^2)$	$(-0.4367 - 0.2421x^2)$
$\sqrt{\frac{2}{\pi}}x$	$(1.9161x - 1.6425x^3)$	$(0.4994x + 0.3899x^3)$
$\sqrt{\frac{8}{\pi}}(x^2 - \frac{1}{2})$	$(-0.5535 + 1.4976x^2)$	$(-0.8750 + 1.5781x^2)$
$\sqrt{\frac{32}{\pi}}(x^3 - \frac{3}{4}x)$	$(-1.6425x + 2.7374x^3)$	$(2.4738x - 3.1684x^2)$
$N = 4:$		
$\frac{1}{\sqrt{\pi}}$	$(0.8617 - 0.8829x^2$ $+0.2666x^4)$	$(0.3718 + 0.2194x^2$ $+0.1738x^4)$
$\sqrt{\frac{2}{\pi}}x$	$(1.9161x - 1.6425x^3)$	$(0.4994x + 0.3899x^3)$
$\sqrt{\frac{8}{\pi}}(x^2 - \frac{1}{2})$	$(-0.8829 + 5.1370x^2$ $-4.0078x^4)$	$(-0.8020 + 0.6973x^2)$ $+0.8352x^4)$
$\sqrt{\frac{32}{\pi}}(x^3 - \frac{3}{4}x)$	$(-1.6425x + 2.7374x^3)$	$(2.4738x - 3.1684x^2)$
$\sqrt{\frac{128}{15\pi}}(x^4 - \frac{3}{4}x^2)$	$(0.2666 - 4.0078x^2$ $+4.9535x^4)$	$(-0.9010 + 6.5342x^2$ $-6.3201x^4)$

Table 4.6: Chebyshev II $\{\mathcal{C}_i - II\}$ and Löwdin-Chebyshev II $\{\mathcal{C}_i^{sy} - II,\ \mathcal{C}_i^{ca} - II\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Chebyshev II $\{\mathcal{C}_i - II\}$	$\{\mathcal{C}_i^{sy} - II\}$	$\{\mathcal{C}_i^{ca} - II\}$
$N = 2:$		
$\sqrt{\frac{2}{\pi}}$	$(0.9383 - 0.6266x^2)$	$(0.7469 + 0.1984x^2)$
$\sqrt{\frac{8}{\pi}}x$	$(1.5960x)$	$(-1.5960x)$
$\sqrt{\frac{32}{\pi}}(x^2 - \frac{1}{4})$	$(-0.6266 + 3.1319x^2)$	$(0.8467 - 3.1878x^2)$
$N = 3:$		
$\sqrt{\frac{2}{\pi}}$	$(0.9383 - 0.6266x^2)$	$(0.7469 + 0.1984x^2)$
$\sqrt{\frac{8}{\pi}}x$	$(2.6659x - 2.3692x^3)$	$(1.2571x + 0.6610x^3)$
$\sqrt{\frac{32}{\pi}}(x^2 - \frac{1}{4})$	$(-0.6266 + 3.1319x^2)$	$(0.8467 - 3.1878x^2)$
$\sqrt{\frac{128}{\pi}}(x^3 - \frac{1}{2}x)$	$(-2.3692x + 5.9262x^3)$	$(3.3376x - 6.3479x^3)$
$N = 4:$		
$\sqrt{\frac{2}{\pi}}$	$(0.9382 - 0.9227x^2 + 0.3423x^4)$	$(0.7323 + 0.1990x^2 + 0.1038x^4)$
$\sqrt{\frac{8}{\pi}}x$	$(2.6659x - 2.3692x^3)$	$(1.2571x + 0.6610x^3)$
$\sqrt{\frac{32}{\pi}}(x^2 - \frac{1}{4})$	$(-0.9227 + 7.5676x^2 - 6.5306x^4)$	$(-0.7593 + 1.9640x^2 + 1.5897x^4)$
$\sqrt{\frac{128}{\pi}}(x^3 - \frac{1}{2}x)$	$(-2.3692x + 5.9262x^3)$	$(-3.3376x + 6.3479x^3)$
$\sqrt{\frac{128}{9\pi}}(x^4 - x^2 - \frac{1}{6})$	$(0.3423 - 6.5306x^2 + 10.8701x^4)$	$(-0.8907 + 9.8423x^2 - 12.5852x^4)$

Table 4.7: Bessel $\{\mathcal{B}_i\}$ and Löwdin-Bessel $\{\mathcal{B}_i^{sy},\ \mathcal{B}_i^{ca}\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Bessel $\{\mathcal{B}_i\}$	$\{\mathcal{B}_i^{sy}\}$	$\{\mathcal{B}_i^{ca}\}$
$N = 2:$		
$\sqrt{2}$	$(4.0379 - 6.3172x$ $+2.7409x^2)$	$(0.7946 + 0.5642x$ $+0.4398x^2)$
$\sqrt{4}(3x-2)$	$(-6.3172 + 22.4834x$ $-16.0595x^2)$	$(-3.1585 + 1.8791x$ $+3.2956x^2)$
$\sqrt{6}(10x^2$ $-12x+3)$	$(2.7409 - 16059x$ $+16.7201x^2)$	$(-7.2881 + 28.2749x$ $-23.1069x^2)$
$N = 3:$		
$\sqrt{2}$	$(4.7385 - 11.0582x$ $+11.5325x^2 - 4.8945x^3)$	$(0.7033 + 0.5081x$ $+0.4006x^2 + 0.3315x^3)$
$\sqrt{4}(3x-2)$	$(-11.0582 + 61.1441x$ $-94.7366x^2 + 45.7835x^3)$	$(-2.6590 + 0.7670x$ $1.8504x^2 + 2.2294x^3)$
$\sqrt{6}(10x^2$ $-12x+3)$	$(11.5325 - 94.7366x$ $+190.9662x^2 - 108.8258x^3)$	$(6.0269 - 16.5150x$ $-0.9015x^2 + 13.6180x^3)$
$\sqrt{8}(35x^3 - 60x^2$ $+30x-4)$	$(-4.8945 + 45.7835x$ $-108.8258x^2 + 72.7413x^3)$	$(16.0563 - 121.0720x$ $+239.6138x^2 - 138.0721x^3)$

Table 4.8: Jacobi $\{\mathcal{J}_i\}$ and Löwdin-Jacobi $\{\mathcal{J}_i^{sy},\ \mathcal{J}_i^{ca}\}$ Polynomials

Gram-Schmidt Orthogonalization	Löwdin's Symmetric Orthogonalization	Löwdin's Canonical Orthogonalization
Jacobi $\{\mathcal{J}_i\}$	$\{\mathcal{J}_i^{sy}\}$	$\{\mathcal{J}_i^{ca}\}$
$N = 2:$		
$\frac{\sqrt{3}}{2}$	$(0.9864 - 0.6590x^2)$	$(-0.8306 - 0.1737x^2)$
$\frac{\sqrt{15}}{2}x$	$(1.9367x)$	$(1.9367x)$
$\frac{1}{2}\sqrt{\frac{21}{8}}(5x^2 - 1)$	$(-0.6590 + 3.9989x^2)$	$(-0.8470 + 4.0491x^2)$
$N = 3:$		
$\frac{\sqrt{3}}{2}$	$(0.9864 - 0.6590x^2)$	$(-0.8306 - 0.1737x^2)$
$\frac{\sqrt{15}}{2}x$	$(2.9935x - 2.7170x^3)$	$(-1.6181x - 0.7264x^3)$
$\frac{1}{2}\sqrt{\frac{21}{8}}(5x^2 - 1)$	$(-0.6590 + 3.9989x^2)$	$(-0.8470 + 4.0491x^2)$
$\frac{1}{2}\sqrt{\frac{45}{8}}(7x^3 - 3x)$	$(-2.7170x + 7.8259x^3)$	$(-3.7047x + 8.2522x^3)$
$N = 4:$		
$\frac{\sqrt{3}}{2}$	$(1.0098 - 0.9479x^2 + 0.3855x^4)$	$(0.8233 + 0.1745x^2 + 0.0773x^4)$
$\frac{\sqrt{15}}{2}x$	$(2.9935x - 2.7170x^3)$	$(-1.6181x - 0.7264x^3)$
$\frac{1}{2}\sqrt{\frac{21}{8}}(5x^2 - 1)$	$(-0.9479 + 8.8860x^2 - 8.0560x^4)$	$(-0.7598 + 2.7271x^2 + 1.9364x^4)$
$\frac{1}{2}\sqrt{\frac{45}{8}}(7x^3 - 3x)$	$(-2.7170x + 7.8259x^3)$	$(3.7047x - 8.2522x^3)$
$\frac{1}{2}\sqrt{\frac{165}{64}}(21x^4 - 14x^2 - 1)$	$(0.3855 - 8.0560x^2 + 14.9266x^4)$	$(-0.9010 + 11.7171x^2 - 16.8551x^4)$

Table 4.9: Samples of *Symmetric* Orthogonal Polynomials

Löwdin-Legendre

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}\}$:
$0.8815 - 0.5899\boldsymbol{x^2}$
$1.2247\boldsymbol{x}$
$-0.5899 + 2.2973\boldsymbol{x^2}$

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}, \boldsymbol{x^3}\}$:
$0.8815 - 0.5899\boldsymbol{x^2}$
$2.3110\boldsymbol{x} - 2.0084\boldsymbol{x^3}$
$-0.5899 + 2.2973\boldsymbol{x^2}$
$-2.0084\boldsymbol{x} + 4.2240\boldsymbol{x^3}$

Löwdin-Hermite

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}\}$:
$0.8701 - 0.2986\boldsymbol{x^2}$
$1.0622\boldsymbol{x}$
$-0.2986 + 1.0194\boldsymbol{x^2}$

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}, \boldsymbol{x^3}\}$:
$0.8701 - 0.2986\boldsymbol{x^2}$
$1.6080\boldsymbol{x} - 0.4849\boldsymbol{x^3}$
$-0.2986 + 1.0194\boldsymbol{x^2}$
$-0.4849\boldsymbol{x} + 0.7191\boldsymbol{x^3}$

Löwdin-Laguerre

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}\}$:
$1.5150 - 0.8343\boldsymbol{x} + 0.0939\boldsymbol{x^2}$
$-0.8343 + 2.0364\boldsymbol{x} - 0.3961\boldsymbol{x^2}$
$0.0939 - 0.3961\boldsymbol{x} + 0.2903\boldsymbol{x^2}$

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}, \boldsymbol{x^3}\}$:
$1.6191 - 1.1499\boldsymbol{x} + 0.2364\boldsymbol{x^2} - 0.0143\boldsymbol{x^3}$
$-1.1499 + 3.3506\boldsymbol{x} - 1.2007\boldsymbol{x^2} + 0.0990\boldsymbol{x^3}$
$0.2364 - 1.2007\boldsymbol{x} + 0.9939\boldsymbol{x^2} - 0.1219\boldsymbol{x^3}$
$-0.0143 + 0.0990\boldsymbol{x} - 0.1219\boldsymbol{x^2} + 0.0539\boldsymbol{x^3}$

Löwdin-Chebyshev-I

$\{1, x, x^2\}$: $0.8062 - 0.5535x^2$

$0.7979x$

$-0.5535 + 1.4976x^2$

$\{1, x, x^2, x^3\}$: $0.8062 - 0.5535x^2$

$1.9161x - 1.6425x^3$

$-0.5535 + 1.4976x^2$

$-1.6425x + 2.7374x^3$

Löwdin-Chebyshev-II

$\{1, x, x^2\}$: $0.9383 - 0.6266x^2$

$1.5960x$

$-0.6266 + 3.1319x^2$

$\{1, x, x^2, x^3\}$: $0.9393 - 0.6266x^2$

$2.6659x - 2.3692x^3$

$-0.6266 + 3.1319x^2$

$-2.3692x + 5.9262x^3$

Löwdin-Bessel

$\{1, x, x^2\}$: $4.0379 - 6.3172x + 2.7409x^2$

$-6.3172 + 22.4834x - 16.0595x^2$

$2.7409 - 16.0595x + 16.7201x^2$

$\{1, x, x^2, x^3\}$: $4.7385 - 11.0582x + 11.5325x^2 - 4.8945x^3$

$-11.0582 + 61.1441x - 94.7366x^2 + 45.7835x^3$

$11.5325 - 94.7366x + 190.9662x^2 - 108.8258x^3$

$-4.8945 + 45.7835x - 108.8258x^2 + 72.7413x^3$

Löwdin-Jacobi

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}\}$: 0.9864−0.6590$\boldsymbol{x^2}$

1.9367$\boldsymbol{x}$

−0.6590+3.9989$\boldsymbol{x^2}$

$\{\mathbf{1}, \boldsymbol{x}, \boldsymbol{x^2}, \boldsymbol{x^3}\}$: 0.9864−0.6590$\boldsymbol{x^2}$

2.9935$\boldsymbol{x}$−2.7170$\boldsymbol{x^3}$

−0.6590+3.9989$\boldsymbol{x^2}$

−2.7170+7.8259$\boldsymbol{x^3}$

4.6 Asymptotic Limits and Zeros

We have thoroughly checked the asymptotic limits of all the newly obtained orthogonal polynomials. In the case of Löwdin-Legendre, Löwdin-Chebyshev I and II polynomials, $\boldsymbol{N}$ is $\geq \mathbf{8}$; for Löwdin-Hermite polynomials, it is $\geq \mathbf{13}$; for Löwdin-Jacobi polynomials, it is $\geq \mathbf{10}$ and for Löwdin Bessel polynomials, it is $\geq \mathbf{6}$. In the case of Löwdin-Bessel and Löwdin-Laguerre polynomials the orthogonality begins to show signs of not being so accurate as early as $\boldsymbol{N} = \mathbf{3}$.

In spite of the above apparent shortcomings in the new polynomials their main strengths lie in the special properties they possess in terms of their inner products with the monomials as illustrated in the equations (4.6) and (4.7). It therefore appears that they will be useful in the problems where these special properties play an important role. The application aspect is however open and we are unable at this stage to make any definite proposal. We have found the Löwdin's symmetric and canonical coefficients for the case of newly generated Legendre, Hermite and Laguerre orthogonal polynomials and are listed in Tables 4.10 to 4.12.

Table 4.10: Löwdin-Legendre - *Symmetric* and *Canonical* Coefficients

Löwdin-Legendre Symmetric Orthogonalization	Löwdin-Legendre Canonical Orthogonalization
N = 2:	N = 2:
$\boldsymbol{s_0 = 1.0782, s_1 = 0, s_2 = -0.3933}$	$\boldsymbol{c_0 = 0.7038, c_1 = 0, c_2 = 0.1516}$
$\boldsymbol{s_0 = 0, s_1 = 1.2247, s_2 = 0}$	$\boldsymbol{c_0 = 0, c_1 = 1.2247, c_2 = 0}$
$\boldsymbol{s_0 = 0.1759, s_1 = 0, s_2 = 1.5315}$	$\boldsymbol{c_0 = -0.0678, c_1 = 0, c_2 = 1.5739}$
N = 3:	N = 3:
$\boldsymbol{s_0 = 1.0782, s_1 = 0, s_2 = -0.3933, s_3 = 0}$	$\boldsymbol{c_0 = 0.7038, c_1 = 0, c_2 = 0.1516, c_3 = 0}$
$\boldsymbol{s_0 = 0, s_1 = 1.1059, s_2 = 0, s_3 = -0.8034}$	$\boldsymbol{c_0 = 0, c_1 = 1.2160, c_2 = 0, c_3 = 0.2227}$
$\boldsymbol{s_0 = 0.1759, s_1 = 0, s_2 = 1.5315, s_3 = 0}$	$\boldsymbol{c_0 = -0.0678, c_1 = 0, c_2 = 1.5739, c_3 = 0}$
$\boldsymbol{s_0 = 0, s_1 = 0.5260, s_2 = 0, s_3 = 1.6896}$	$\boldsymbol{c_0 = 0, c_1 = -0.1456, c_2 = 0, c_3 = 1.8576}$
N = 4:	N = 4:
$\boldsymbol{s_0 = 0.6806, s_1 = 0, s_2 = -0.4261,}$ $\boldsymbol{s_3 = 0, s_4 = 0.0705}$	$\boldsymbol{c_0 = 0.6993, c_1 = 0, c_2 = 0.2312,}$ $\boldsymbol{c_3 = 0, c_4 = 0.0325}$
$\boldsymbol{s_0 = 0, s_1 = 1.1059, s_2 = 0,}$ $\boldsymbol{s_3 = -0.8034, s_4 = 0}$	$\boldsymbol{c_0 = 0, c_1 = 1.2160, c_2 = 0,}$ $\boldsymbol{c_3 = 0.2227, c_4 = 0}$
$\boldsymbol{s_0 = 0.1700, s_1 = 0, s_2 = 1.2476,}$ $\boldsymbol{s_3 = 0, s_4 = -1.1994}$	$\boldsymbol{c_0 = -0.1041, c_1 = 0, c_2 = 1.5502,}$ $\boldsymbol{c_3 = 0, c_4 = 0.2783}$
$\boldsymbol{s_0 = 0, s_1 = 0.5260, s_2 = 0,}$ $\boldsymbol{s_3 = 1.6896, s_4 = 0}$	$\boldsymbol{c_0 = 0, c_1 = -0.1456, c_2 = 0,}$ $\boldsymbol{c_3 = 1.8576, c_4 = 0}$
$\boldsymbol{s_0 = 0.0893, s_1 = 0, s_2 = 0.8728,}$ $\boldsymbol{s_3 = 0, s_4 = 1.7485}$	$\boldsymbol{c_0 = 0.0031, c_1 = 0, c_2 = -0.2086,}$ $\boldsymbol{c_3 = 0, c_4 = 2.1030}$

Table 4.11: Löwdin-Hermite - *Symmetric* and *Canonical* Coefficients

Löwdin-Hermite Symmetric Orthogonalization	Löwdin-Hermite Canonical Orthogonalization
N = 2:	N = 2:
$s_0 = 0.7207, s_1 = 0, s_2 = -0.0747$	$c_0 = 0.6981, c_1 = 0, c_2 = 0.0980$
$s_0 = 0, s_1 = 0.5311, s_2 = 0$	$c_0 = 0, c_1 = 0.5311, c_2 = 0$
$s_0 = 0.2112, s_1 = 0, s_2 = 0.2549$	$c_0 = -0.2773, c_1 = 0, c_2 = 0.2468$
N = 3:	N = 3:
$s_0 = 0.7207, s_1 = 0, s_2 = -0.0747, s_3 = 0$	$c_0 = 0, c_1 = 0.4493, c_2 = 0, c_3 = 0.0579$
$s_0 = 0, s_1 = 0.4404, s_2 = 0, s_3 = -0.0606$	$c_0 = -0.6981, c_1 = 0, c_2 = -0.098, c_3 = 0$
$s_0 = 0.2112, s_1 = 0, s_2 = 0.2549, s_3 = 0$	$c_0 = 0.2773 c_1 = 0, c_2 = -0.2468, c_3 = 0$
$s_0 = 0, s_1 = 0.2970, s_2 = 0, s_3 = 0.0899$	$c_0 = 0, c_1 = 0.2834, c_2 = 0, c_3 = -0.0917$
N = 4:	N = 4:
$s_0 = 0.7136, s_1 = 0, s_2 = -0.0799,$ $s_3 = 0, s_4 = 0.0032$	$c_0 = 0.2773, c_1 = 0, c_2 = 0.2191,$ $c_3 = 0, c_4 = 0.0166$
$s_0 = 0, s_1 = 0.4404, s_2 = 0,$ $s_3 = -0.0606, s_4 = 0$	$c_0 = 0, c_1 = -0.8985, c_2 = 0,$ $c_3 = -0.0579, c_4 = 0$
$s_0 = 1.5595, s_1 = 0, s_2 = 0.8459,$ $s_3 = 0, s_4 = -0.0284$	$c_0 = -0.6932, c_1 = 0, c_2 = -0.0727,$ $c_3 = 0, c_4 = -0.0104$
$s_0 = 0, s_1 = 0.2970, s_2 = 0,$ $s_3 = 0.0899, s_4 = 0$	$c_0 = 0, c_1 = 0.2834, c_2 = 0,$ $c_3 = -0.0917, c_4 = 0$
$s_0 = 0.1295, s_1 = 0, s_2 = 0.1924,$ $s_3 = 0, s_4 = 0.0255$	$c_0 = -0.0791, c_1 = 0, c_2 = 0.1322,$ $c_3 = 0, c_4 = 2.1030$

Table 4.12: Löwdin-Laguerre - *Symmetric* and *Canonical* Coefficients

Löwdin-Laguerre Symmetric Orthogonalization	Löwdin-Laguree Canonical Orthogonalization
N = 2:	N = 2:
$\boldsymbol{s_0 = 0.8685, s_1 = 0.4587, s_2 = 0.0939}$	$\boldsymbol{c_0 = 0.1995, c_1 = -0.8096, c_2 = 0.1902}$
$\boldsymbol{s_0 = 0.4185, s_1 = -0.4700, s_2 = -0.3916}$	$\boldsymbol{c_0 = -0.8692, c_1 = -0.2919, c_2 = 0.1995}$
$\boldsymbol{s_0 = 0.2784, s_1 = -0.7651, s_2 = 0.2903}$	$\boldsymbol{c_0 = 0.2140, c_1 = 0.5071, c_2 = 0.4171}$
N = 3:	N = 3:
$\boldsymbol{s_0 = 0.8652, s_1 = 0.4617,}$ $\boldsymbol{s_2 = 0.1077, s_3 = 0.0143}$	$\boldsymbol{c_0 = 0.2309, c_1 = -0.6772,}$ $\boldsymbol{c_2 = 0.3319, c_3 = -0.0362}$
$\boldsymbol{s_0 = 0.3933, s_1 = -0.3298,}$ $\boldsymbol{s_2 = -0.3097, s_3 = 0.0990}$	$\boldsymbol{c_0 = -0.7961, c_1 = 0.5785,}$ $\boldsymbol{c_2 = 0.2542, c_3 = -0.0685}$
$\boldsymbol{s_0 = 0.2921, s_1 = -0.5807,}$ $\boldsymbol{s_2 = -0.1032, s_3 = 0.1219}$	$\boldsymbol{c_0 = 0.7336, c_1 = 0.5363,}$ $\boldsymbol{c_2 = 0.0825, c_3 = -0.0639}$
$\boldsymbol{s_0 = 0.1643, s_1 = -0.5816,}$ $\boldsymbol{s_2 = 0.3632, s_3 = -0.0539}$	$\boldsymbol{c_0 = -0.0896, c_1 = -0.2872,}$ $\boldsymbol{c_2 = -0.2611, c_3 = -0.1330}$

We have calculated the zeros of all the newly obtained polynomials and are listed in Tables from 4.13 to 4.19. The zeros possess the following properties.

- They are real and distinct on the interval $[\boldsymbol{a}, \boldsymbol{b}]$ and are located in the interior of the interval.
- Each interval contains exactly one root.
- There are cases where the roots are imaginary but is of exactly one root.
- The real or imaginary root lies either in the interval $[\boldsymbol{a}, \boldsymbol{b}]$ or in the exterior of $[\boldsymbol{a}, \boldsymbol{b}]$.

Table 4.13: Zeros of Löwdin-Legendre Orthogonal Polynomials

Löwdin-Legendre Symmetric Polynomials	Löwdin-Legendre Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.2224, -1.2224}$	$\mathbf{0 + 1.6618\ \mathit{i}, 0 - 0.6618\ \mathit{i}}$
$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0.5067, -0.5067}$	$\mathbf{0.6017, -0.6017}$
N = 3:	N = 3:
$\mathbf{1.2224, -1.2224}$	$\mathbf{0 + 1.6618\ \mathit{i}, 0 - 0.6618\ \mathit{i}}$
$\mathbf{0, 1.0727, -1.0727}$	$\mathbf{0, 0 + 1.2585\ \mathit{i}, 0 - 1.2585\ \mathit{i}}$
$\mathbf{0.5067, -0.5067}$	$\mathbf{0.6017, -0.6017}$
$\mathbf{0, 0.6895, -0.6895}$	$\mathbf{0, 0.7946, -0.7946}$
N = 4:	N = 4:
$\mathbf{-1.2632 + 0.3623\ \mathit{i}, -1.2632 - 0.3623\ \mathit{i},}$ $\mathbf{1.2632 + 0.3623\ \mathit{i}, 1.2632 - 0.3623\ \mathit{i}}$	$\mathbf{-0.7925 + 1.1907\ \mathit{i}, -0.7925 - 1.1907\ \mathit{i},}$ $\mathbf{0.7925 + 1.1907\ \mathit{i}, 0.7925 + 1.1907\ \mathit{i}}$
$\mathbf{0, 1.0727, -1.0727}$	$\mathbf{0, 0 + 1.2585\ \mathit{i}, 0 - 1.2585\ \mathit{i}}$
$\mathbf{1.0246, -1.0246, 0.4050, -0.4050}$	$\mathbf{0 + 1.2147\ \mathit{i}, 0 - 1.2174\ \mathit{i}, 0.6553, -0.6553}$
$\mathbf{0, 0.6895, 0.6895}$	$\mathbf{0.0.7946, -0.7946}$
$\mathbf{0.7880, -0.7880, 0.2548, -0.2548}$	$\mathbf{0.8739, -0.8739, 0.3571, -0.3571}$

Table 4.14: Zeros of Löwdin-Hermite Orthogonal Polynomials

Löwdin-Hermite Symmetric Polynomials	Löwdin-Hermite Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.7070, -1.7070}$	$\mathbf{0 + 1.1318\ \boldsymbol{i}, 0 - 0.1318\ \boldsymbol{i}}$
$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0.5412, -0.5412}$	$\mathbf{0.8836, -0.8836}$
N = 3:	N = 3:
$\mathbf{1.7070, -1.7070}$	$\mathbf{0 + 1.1318\ \boldsymbol{i}, 0 - 0.1318\ \boldsymbol{i}}$
$\mathbf{0, 1.8210, -1.8210}$	$\mathbf{0, 0 + 0.6633\ \boldsymbol{i}, 0 - 0.6633\ \boldsymbol{i}}$
$\mathbf{0.5412, -0.5412}$	$\mathbf{0.8836, -0.8836}$
$\mathbf{0, 0.8212, -0.8212}$	$\mathbf{0, 1.5077, -1.5077}$
N = 4:	N = 4:
$\mathbf{2.5712, -2.5712,}$ $\mathbf{1.6493, -1.6493}$	$\mathbf{-0.3391 + 0.5150\ \boldsymbol{i}, -0.3391 - 0.5150\ \boldsymbol{i},}$ $\mathbf{0.3391 + 0.5150\ \boldsymbol{i}, 0.3391 - 0.5150\ \boldsymbol{i}}$
$\mathbf{0, 1.8210, -1.8210}$	$\mathbf{0, 0 + 0.6633\ \boldsymbol{i}, 0 - 0.6633\ \boldsymbol{i}}$
$\mathbf{2.0486, -2.0486, 0.4980, -0.4980}$	$\mathbf{1.6727, -1.6727, 0 + 1.2408\ \boldsymbol{i}, 0 - 1.2408\ \boldsymbol{i}}$
$\mathbf{0, 0.8212, -0.8212}$	$\mathbf{0.1.5077, -1.5077}$
$\mathbf{0.9935, -0.9935, 0.3546, -0.3546}$	$\mathbf{1.9016, -1.9016, 0.6225, -0.6255}$

Table 4.15: Zeros of Löwdin-Laguerre Orthogonal Polynomials

Löwdin-Laguerre Symmetric Polynomials	Löwdin-Laguerre Canonical Polynomials
N = 2:	N = 2:
$\mathbf{6.3403, 2.5447}$	$\mathbf{-0.1283 + 0.2739\ \boldsymbol{i}, -0.1283 - 0.2739\ \boldsymbol{i}}$
$\mathbf{4.6922, 0.4489}$	$\mathbf{3.5984, -1.0616}$
$\mathbf{1.0590, 0.3054}$	$\mathbf{4.3666, 0.8539}$
N = 3:	N = 3:
$\mathbf{8.9828, 5.0553, 2.4933}$	$\mathbf{-0.2045, 0.0180 + 0.2005\ \boldsymbol{i}, 0.0180 - 0.2005\ \boldsymbol{i}}$
$\mathbf{8.1509, 3.5793, 0.3981}$	$\mathbf{4.7558, 0.2666 + 0.5458\ \boldsymbol{i}, 0.2666 - 0.5458\ \boldsymbol{i}}$
$\mathbf{6.7333, 1.1750, 0.2451}$	$\mathbf{6.6737, 2.1710, -1.1358}$
$\mathbf{1.0398 + 0.6141\ \boldsymbol{i}, 1.0398 - 0.6141\ \boldsymbol{i}, 0.1819}$	$\mathbf{7.4195, 2.9634, 0.5803}$

Table 4.16: Zeros of Löwdin-Chebyshev I Orthogonal Polynomials

Löwdin-Chebyshev I Symmetric Polynomials	Löwdin-Chebyshev I Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.2069, -1.2069}$	$\mathbf{0 + 1.3431\ \mathit{i}, 0 - 1.3431\ \mathit{i}}$
$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0.6079, -0.6079}$	$\mathbf{0.7446, -0.7446}$
N = 3:	N = 3:
$\mathbf{1.2069, -1.2069}$	$\mathbf{0 + 1.3431\ \mathit{i}, 0 - 1.3431\ \mathit{i}}$
$\mathbf{0, 1.0801, -1.0801}$	$\mathbf{0, 0 + 1.1317\ \mathit{i}, 0 - 1.1317\ \mathit{i}}$
$\mathbf{0.6079, -0.6079}$	$\mathbf{0.7446, -0.7446}$
$\mathbf{0, 0.7746, -0.7746}$	$\mathbf{0, 0.8836, -0.8836}$
N = 4:	N = 4:
$\mathbf{1.3141 + 0.2664\ \mathit{i}, 1.3141 - 0.2664\ \mathit{i},}$ $\mathbf{-1.3141 + 0.2664\ \mathit{i}, -1.3141 - 0.2664\ \mathit{i}}$	$\mathbf{0.6448 + 1.0232\ \mathit{i}, 0.6448 - 1.0232\ \mathit{i},}$ $\mathbf{-0.6448 + 1.0232\ \mathit{i}, -0.6448 - 1.0232\ \mathit{i}}$
$\mathbf{0, 1.0801, -1.0801}$	$\mathbf{0, 0 + 1.1317\ \mathit{i}, 0 - 1.1317\ \mathit{i}}$
$\mathbf{1.0379, -1.0379, 0.4522, -0.4522}$	$\mathbf{0 + 1.2176\ \mathit{i}, 0 - 1.2176\ \mathit{i}, 0.8048, -0.8048}$
$\mathbf{0, 0.7746, -0.7746}$	$\mathbf{0, 0.8836, -0.8836}$
$\mathbf{0.8579, -0.8579, 0.2704, -0.2704}$	$\mathbf{0.9327, -0.9327, 0.4048, -0.4048}$

Table 4.17: Zeros of Löwdin-Chebyshev II Orthogonal Polynomials

Löwdin-Chebyshev II Symmetric Polynomials	Löwdin-Chebyshev II Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.2237, -1.2237}$	$\mathbf{0 + 1.9403\ i, 0 - 1.9403\ i}$
$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0.4473, -0.4473}$	$\mathbf{0.5154, -0.5154}$
N = 3:	N = 3:
$\mathbf{1.2237, -1.2237}$	$\mathbf{0 + 1.9403\ i, 0 - 1.9403\ i}$
$\mathbf{0, 1.0622, -1.0622}$	$\mathbf{0, 0 + 1.3791\ i, 0 - 1.3791\ i}$
$\mathbf{0.4473, -0.4473}$	$\mathbf{0.5154, -0.5154}$
$\mathbf{0, 0.6341, -0.6341}$	$\mathbf{0, 0.7251, -0.7251}$
N = 4:	N = 4:
$\mathbf{1.2254 + 0.3923\ i, 1.2254 - 0.3923\ i,}$ $\mathbf{-1.2254 + 0.3923\ i, -1.2254 - 0.3923\ i}$	$\mathbf{0.9213 + 1.3444\ i, 0.9213 - 1.3444\ i,}$ $\mathbf{-0.9213 + 1.3444\ i, -0.9213 - 1.3444\ i}$
$\mathbf{0, 1.0622, -1.0622}$	$\mathbf{0, 0 + 1.3791\ i, 0 - 1.3791\ i}$
$\mathbf{1.0101, -1.0101, 0.3721, -0.3721}$	$\mathbf{0 + 1.2428\ i, 0 - 1.2428\ i, 0.5561, -0.5561}$
$\mathbf{0, 0.6314, -0.6314}$	$\mathbf{0.0.7251, -0.7251}$
$\mathbf{0.7367, -0.7367, 0.2409, -0.2409}$	$\mathbf{0.8232, -0.8232, 0.3232, -0.3232}$

Table 4.18: Zeros of Löwdin-Bessel Orthogonal Polynomials

Löwdin-Besesl Symmetric Polynomials	Löwdin-Bessel Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.1524 + 0.3810\ i, 1.1524 - 0.3810\ i}$	$\mathbf{-0.6414 + 1.1812\ i, -0.6414 - 1.1812\ i}$
$\mathbf{1.0109, 0.3891}$	$\mathbf{-1.3047, 0.7346}$
$\mathbf{0.7385, 0.2220}$	$\mathbf{0.8546, 0.3691}$
N = 3:	N = 3:
$\mathbf{1.1041, 0.6261 + 0.6963\ i, 0.6261 - 0.6963\ i}$	$\mathbf{-1.2925, 0.0420 + 1.2805\ i, 0.0420 - 1.2805\ i}$
$\mathbf{0.8845 + 0.1484\ i, 0.8845 - 0.1484\ i, 0.3003}$	$\mathbf{-0.7969 + 0.9625\ i, -0.7969 - 0.9625\ i, 0.7638}$
$\mathbf{0.9699, 0.6040, 0.1809}$	$\mathbf{-1.2223, 0.8743, 04141}$
$\mathbf{0.8463, 0.4863, 0.1635}$	$\mathbf{0.9188, 0.6087, 0.2079}$

Table 4.19: Zeros of Löwdin-Jacobi Orthogonal Polynomials

Löwdin-Jacobi Symmetric Polynomials	Löwdin-Jacobi Canonical Polynomials
N = 2:	N = 2:
$\mathbf{1.2234, -1.2234}$	$\mathbf{0 + 2.1867\ i, 0 - 2.1867\ i}$
$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0.4059, -0.4059}$	$\mathbf{0.4574, -0.4574}$
N = 3:	N = 3:
$\mathbf{1.2234, -1.2234}$	$\mathbf{0 + 2.1867\ i, 0 - 2.1867\ i}$
$\mathbf{0, 1.0497, -1.0497}$	$\mathbf{0.4574, -0.4574}$
$\mathbf{0, 0.5892, -0.5892}$	$\mathbf{0, 0.6700, -0.6700}$
N = 4:	N = 4:
$\mathbf{1.1933 + 0.4410\ i, 1.1933 - 0.4410\ i,}$ $\mathbf{-1.1933 + 0.4410\ i, -1.1933 - 0.4410\ i}$	$\mathbf{-1.0332 + 1.4819\ i, -1.0332 - 1.4819\ i,}$ $\mathbf{1.0332 + 1.4819\ i, 1.0332 - 1.4819\ i}$
$\mathbf{0, 1.0497, -1.0497}$	$\mathbf{0, 0 + 1.4925\ i, 0 - 1.4925\ i}$
$\mathbf{0.9917, -0.9917, 0.3459, -0.3459}$	$\mathbf{-0 + 1.2832\ i, -0 - 1.2832\ i, 0.4882, -0.4882}$
$\mathbf{0, 0.5892, -0.5892}$	$\mathbf{0.0.6700, -0.6700}$
$\mathbf{0.6796, -0.6796, 0.2304, -0.2304}$	$\mathbf{0.7792, -0.7792, 0.2967, -0.2967}$

4.7 A Unified View of All the Polynomials

The new classes of orthogonal polynomials are plotted with the case of orthogonal polynomials obtained through Gram-Schmidt method for the case of $\boldsymbol{N = 3}$. Different colors are given to the different orthogonal polynomials obtained through Gram-Schmidt, symmetric and canonical orthogonalization methods. The orthogonal polynomials are presented in different graphs with uniform colouring pattern. In these three methods, a particular colour indicates a particular kind of orthogonal polynomial. Plots of the Löwdin-Legendre, Löwdin-Hermite, Löwdin-Laguerre, Lowdin-Chebyshev I, Löwdin-Chebyshev II, Löwdin-Bessel and Löwdin-Jacobi plots together with the standard Legendre, Hermite, Laguerre, Chebyshev I, Chebyshev II, Bessel and Jacobi polynomials are shown in Figures 4.1 to 4.7. The different colours; *magenta, green, red* and *blue* represent the $\boldsymbol{0^{th}, 1^{st}, 2^{nd}}$ and $\boldsymbol{3^{rd}}$ polynomials respectively.

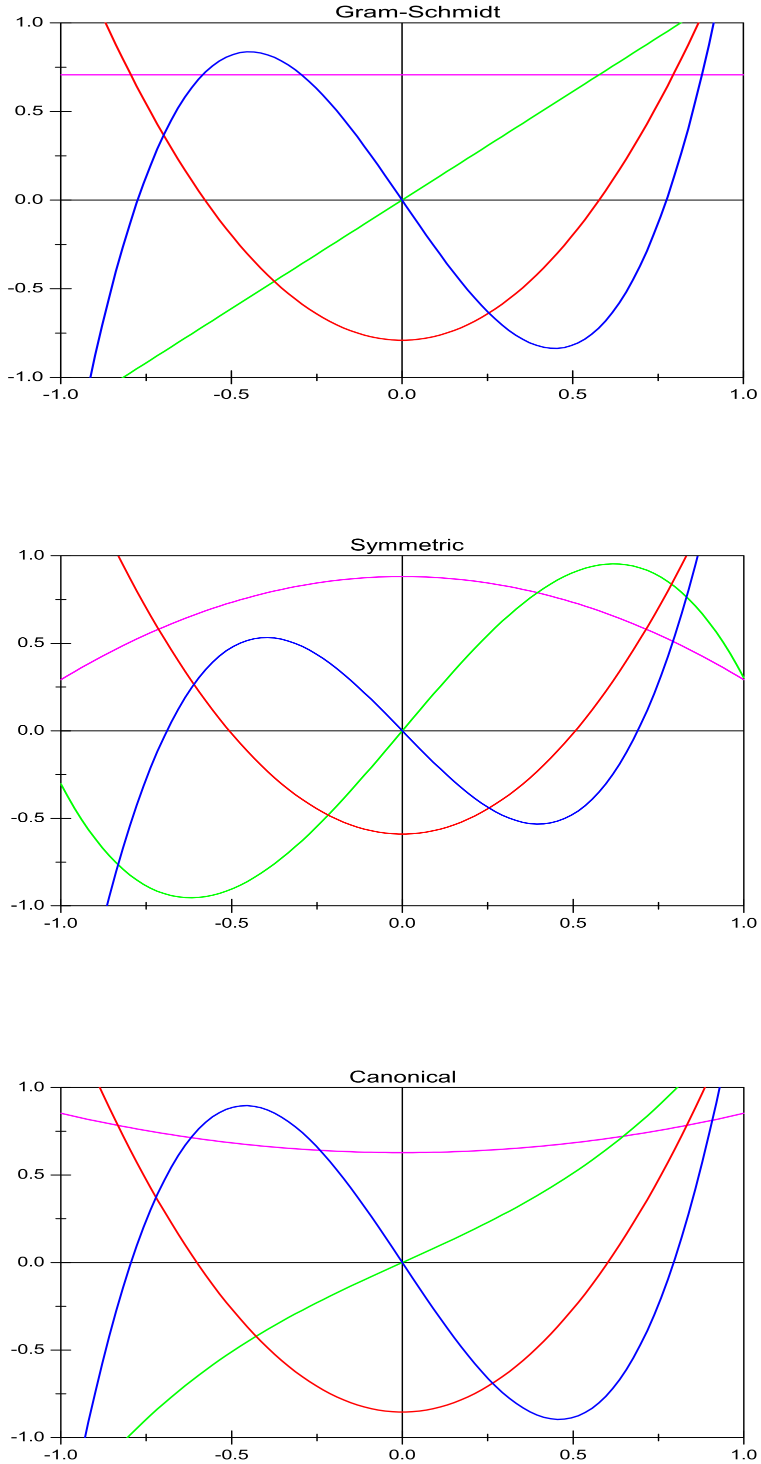


Figure 4.1: Legendre and Löwdin-Legendre Orthogonal Polynomials

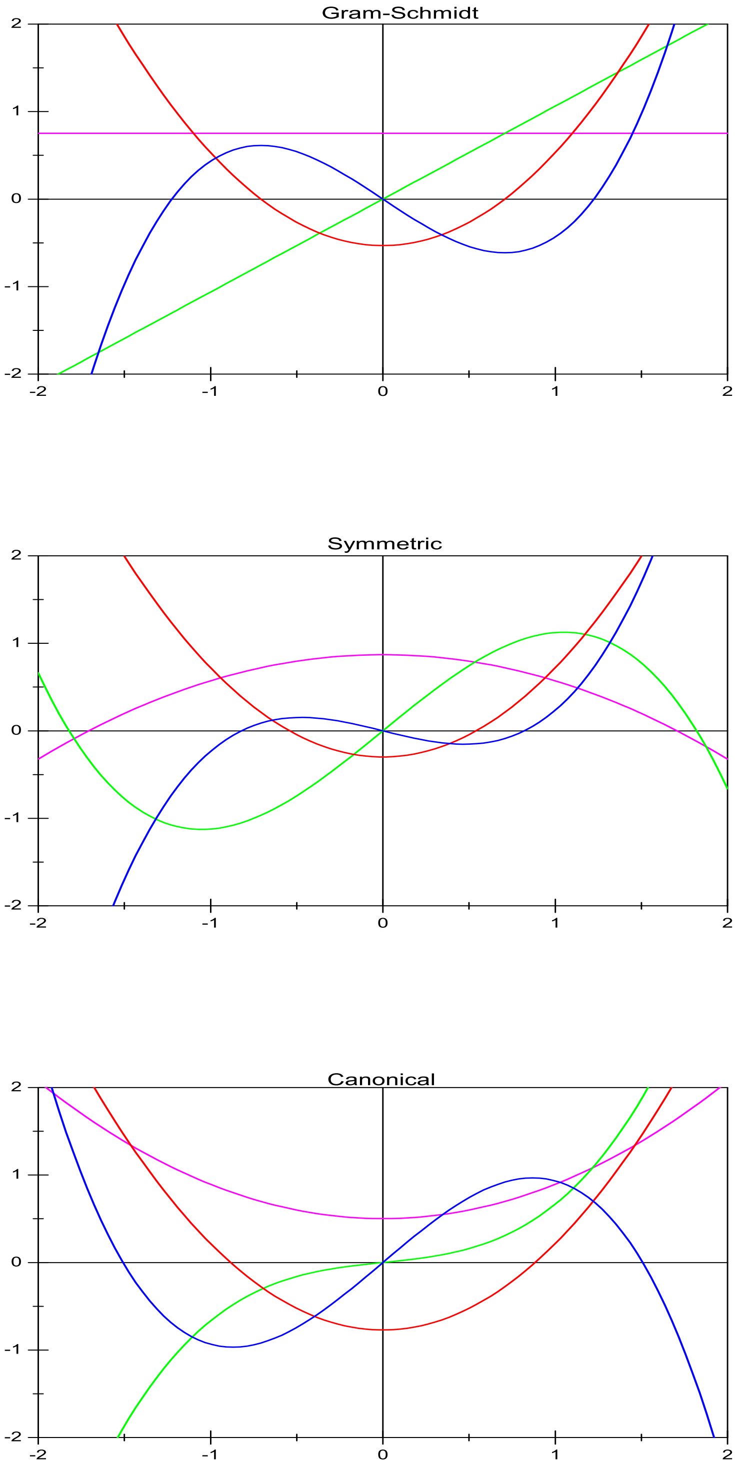


Figure 4.2: Hermite and Löwdin-Hermite Orthogonal Polynomials

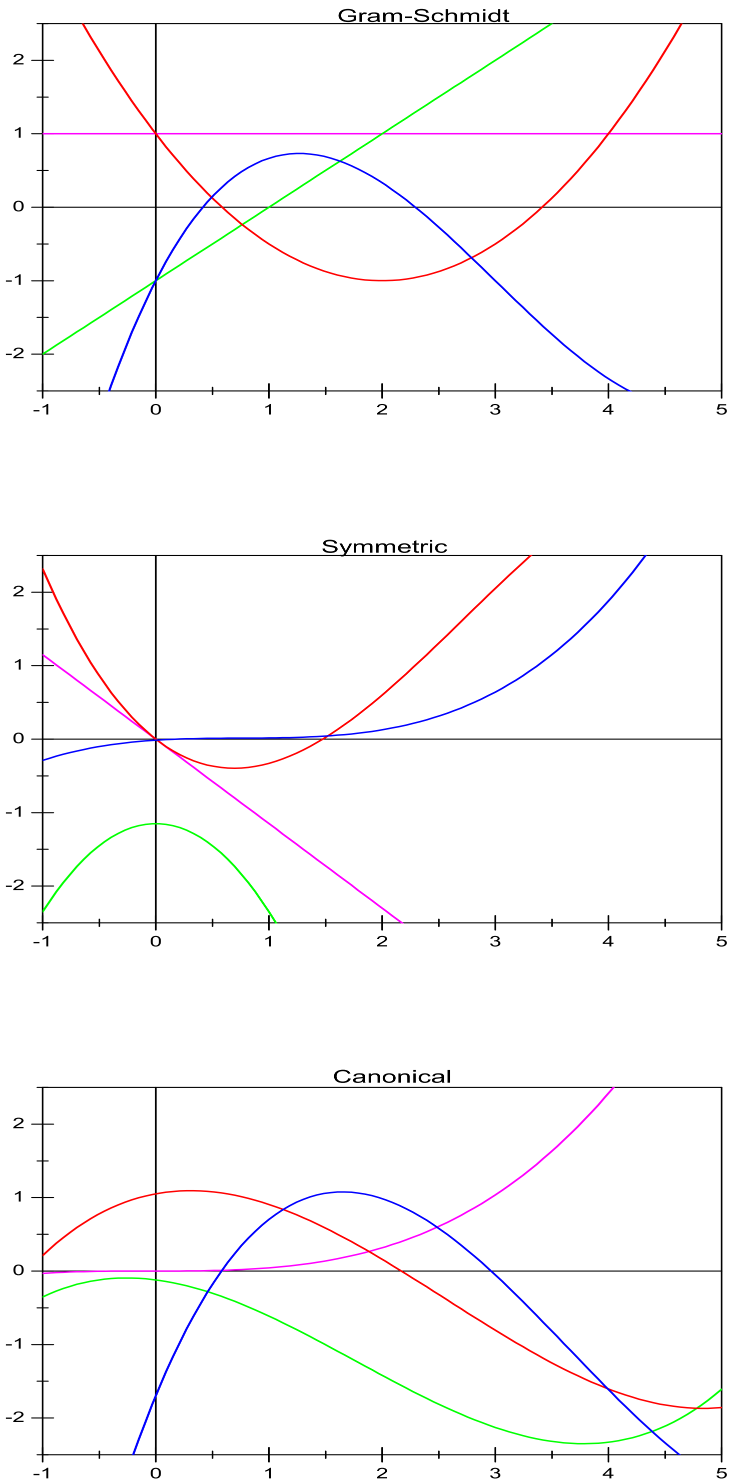


Figure 4.3: Lagurre and Löwdin-Laguerre Orthogonal Polynomials

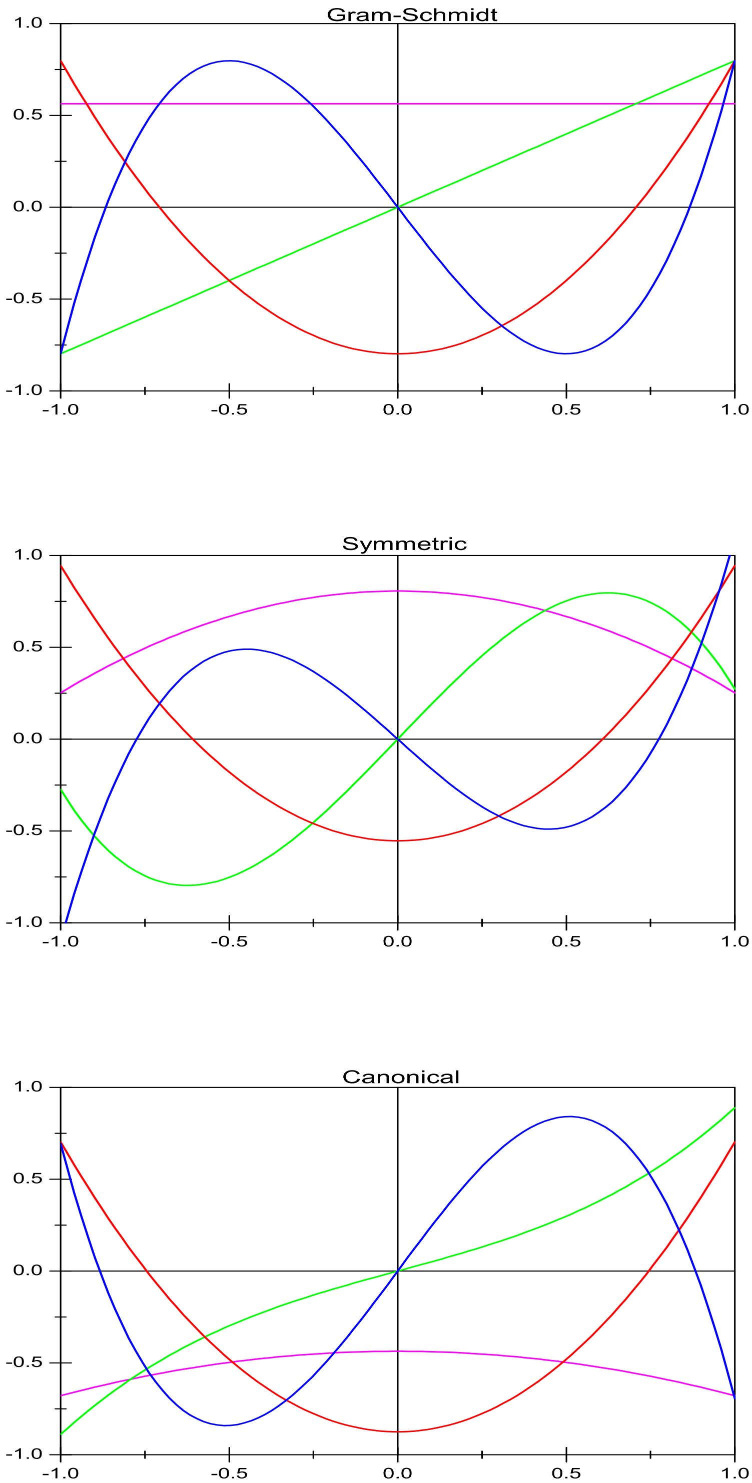


Figure 4.4: Chebyshev I and Löwdin-Chebyshev I Orthogonal Polynomials

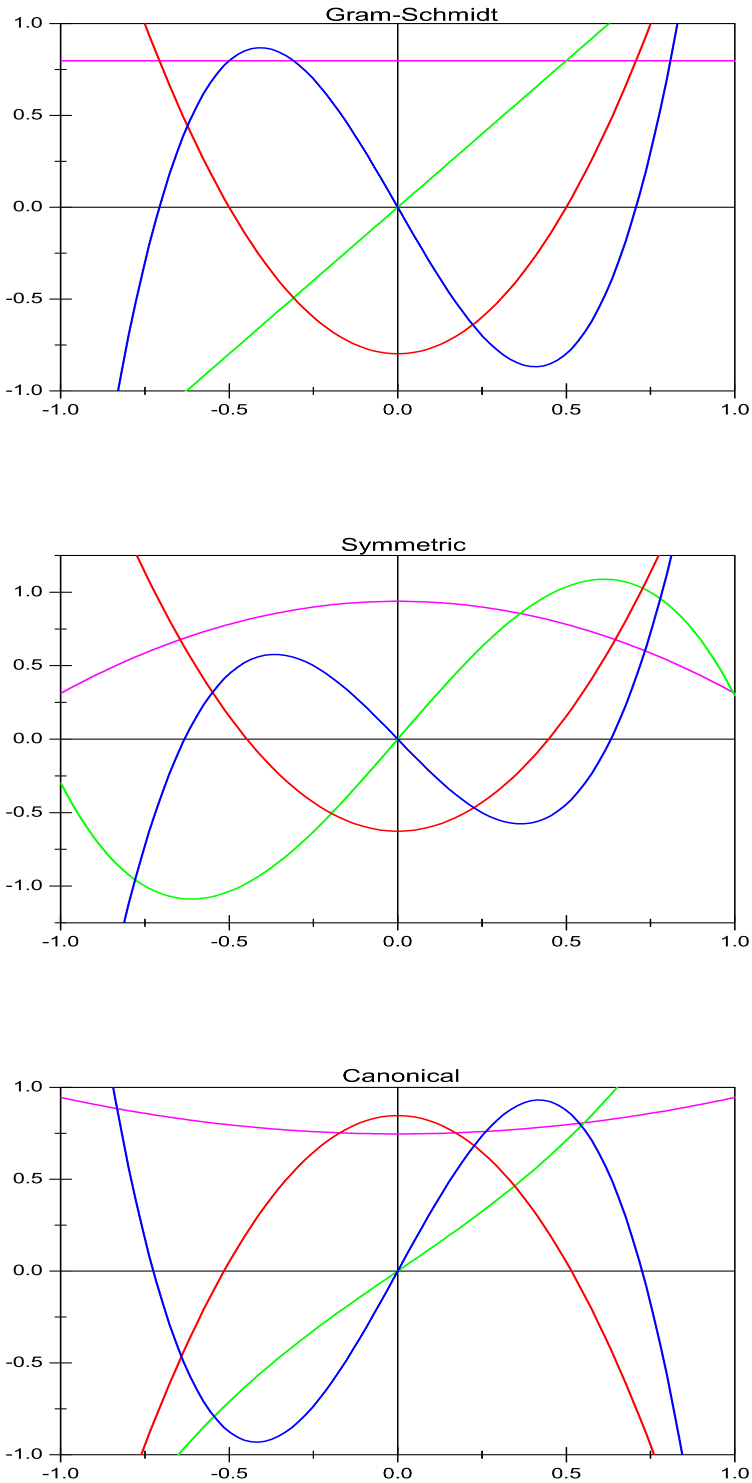


Figure 4.5: Chebyshev II and Löwdin-Chebyshev II Orthogonal Polynomials

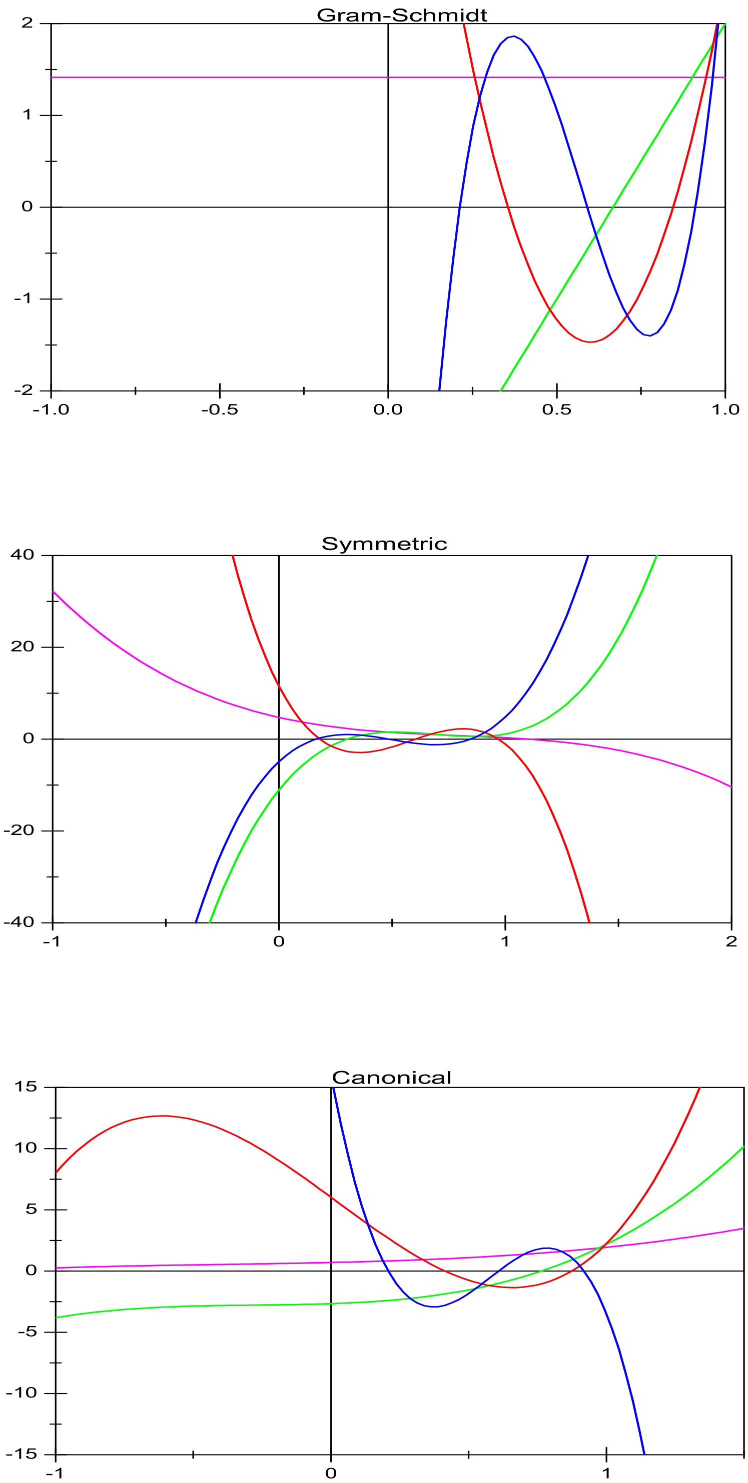


Figure 4.6: Bessel and Löwdin-Bessel Orthogonal Polynomials

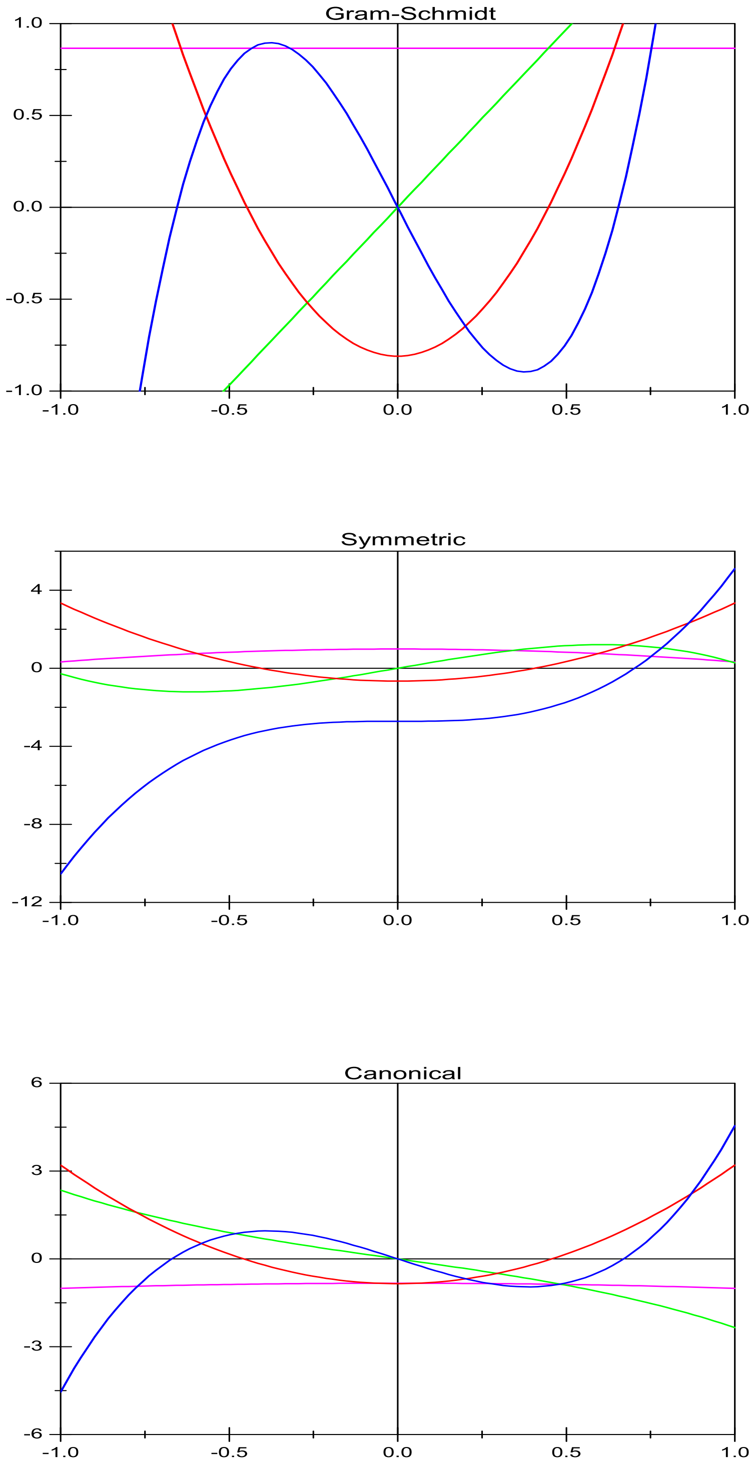


Figure 4.7: Jacobi and Löwdin-Jacobi Orthogonal Polynomials

A unified view of all the polynomials obtained using the three methods viz., the Gram-Schmidt, symmetric and canonical are presented in a single graph as shown in Figures 4.8 to 4.14. This unification helps to visualize the three kinds of orthogonal polynomials. We have presented the plot of polynomials of Löwdin-Legendre, Löwdin-Chebyshev I, Löwdin-Chebyshev II, Löwdin-Jacobi, Löwdin-Bessel, Löwdin-Hermite and Löwdin-Laguerre in a single graph. We have used different symbols for different kind of polynomials. In this unified graphical representation, except Bessel and Laguerre orthogonal polynomials big differences are there in other polynomials. The standard polynomials obtained through Gram-Schmidt method are represented by *solid lines* (—), the Löwdin symmetric orthogonal polynomials are represented by *dashed lines* (— — —) and the Löwdin canonical orthogonal polynomials are represented by *dotted lines* (...). A particular colour represent one kind of polynomial in each of the three polynomials.

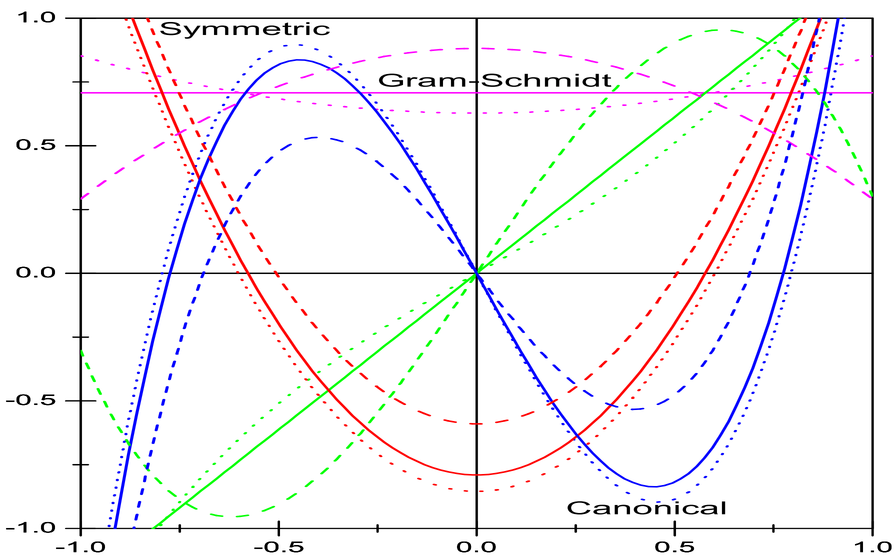


Figure 4.8: Unified view of Legendre and Löwdin-Legendre Polynomials

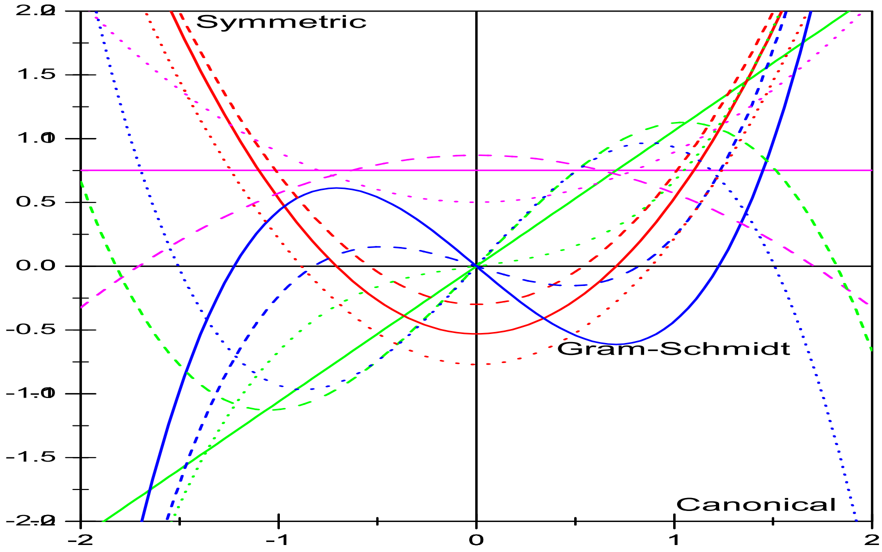


Figure 4.9: Unified view of Hermite and Löwdin-Hermite Polynomials

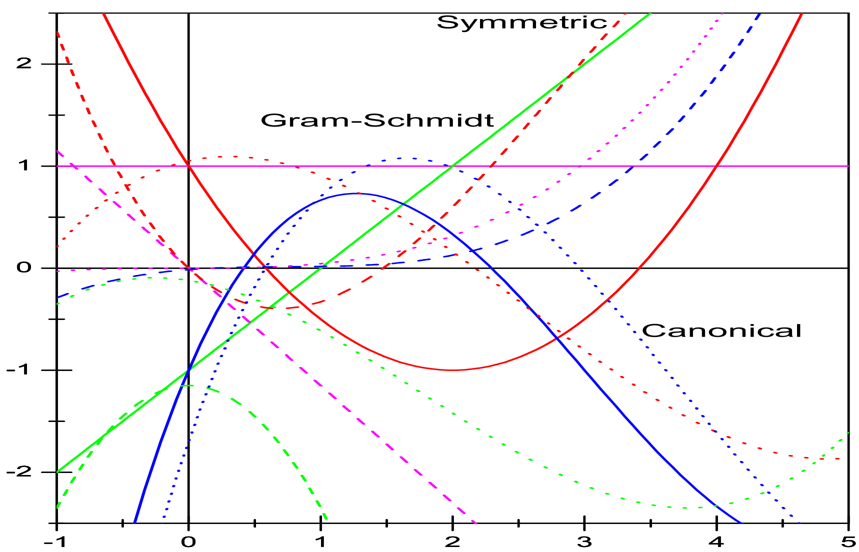


Figure 4.10: Unified view of Lagurre and Löwdin-Laguerre Polynomials

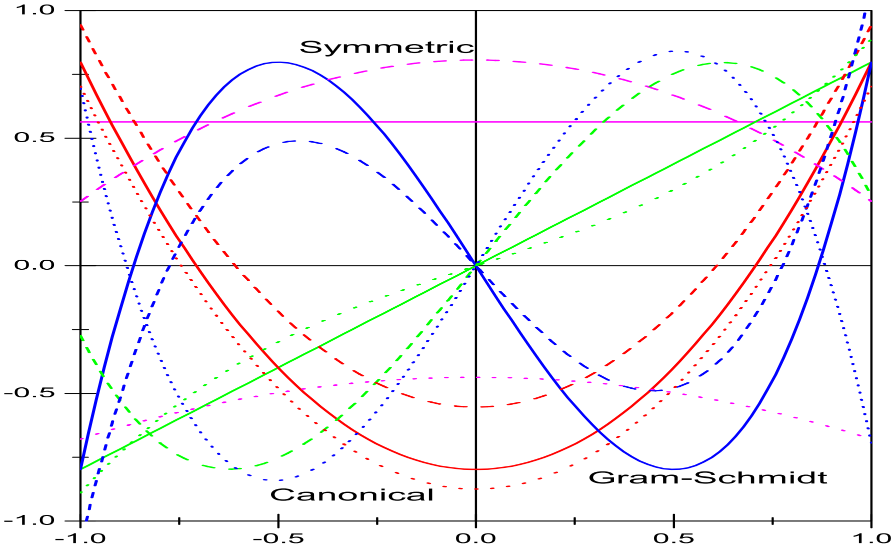


Figure 4.11: Unified view of Chebyshev I and Löwdin-Chebyshev I Polynomials

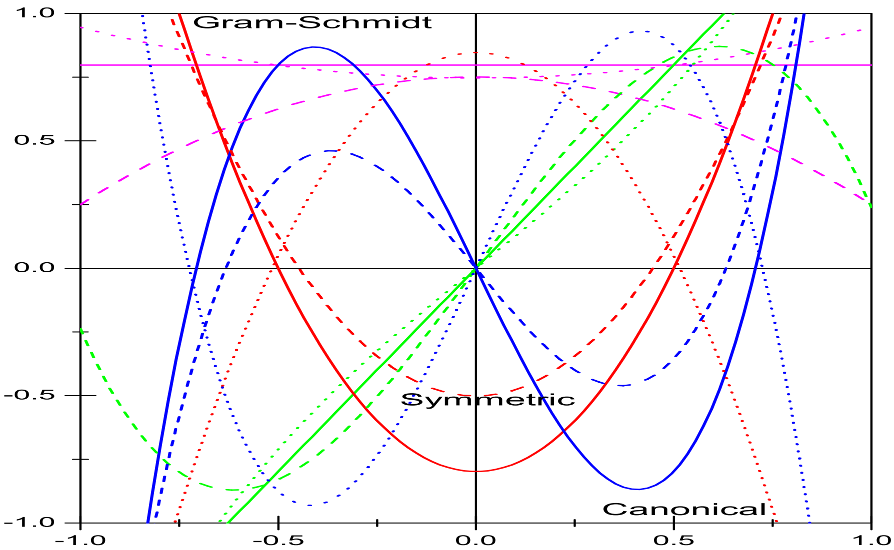


Figure 4.12: Unified view of Chebyshev II and Löwdin-Chebyshev II Polynomials

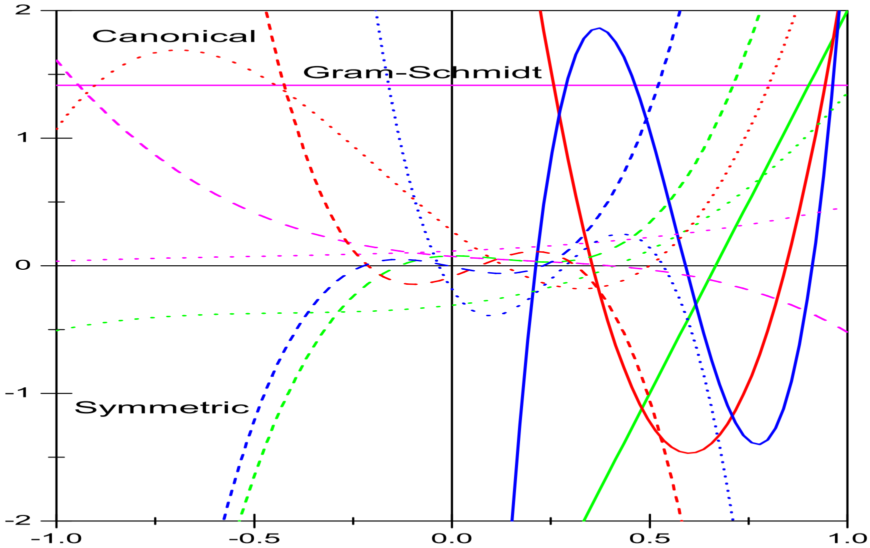


Figure 4.13: Unified view of Bessel and Löwdin-Bessel Polynomials

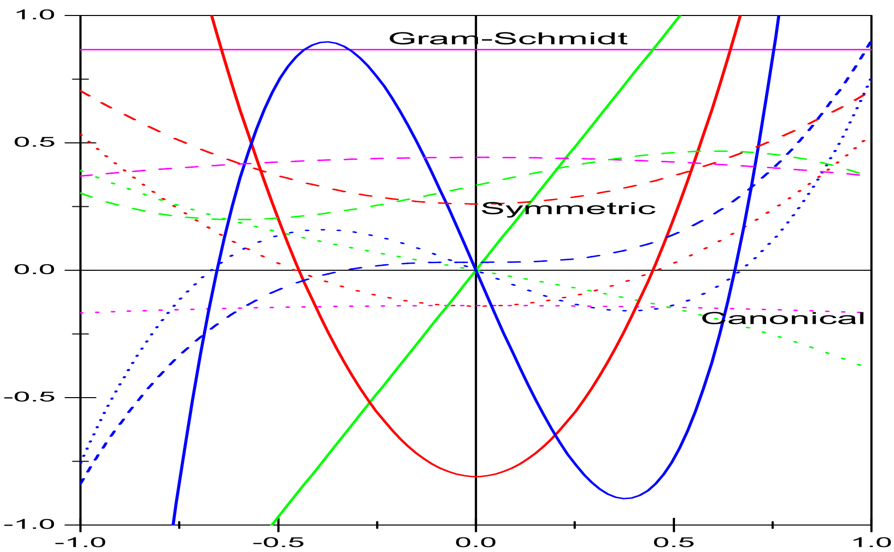


Figure 4.14: Unified view of Jacobi and Löwdin-Jacobi Polynomials

4.8 Computations, Results and Discussion

We have obtained new sets of orthogonal polynomials as the set of monomials changes, one in the *symmetric* class and another in the *canonical* class. We have only enlarged them by adding one higher monomial successively. They change every time the set of monomials changes because each time the entire set of monomials is used in generating the orthogonal polynomials. The new sets of orthogonal polynomials exhibit some pattern within a set as well as between consecutive sets. The pattern is somewhat stronger in the *symmetric* class. However the patterns do not lead to a clue that might help in inventing Rodrigues like formulae or some other generating functions. We have written the coefficients in decimal form, but are trying to see if there are any patterns hidden in them by writing them as fractions.

The newly generated polynomials using the monomials have shown some repetitive patterns for the case of symmetric orthogonalization. The observed patterns in the repetition of the coefficients are quite apparent, but they are not sufficiently abundant to hint at possible generating functions and recurrence relationships. Löwdin's orthogonalization methods are applied to discover the new and perhaps richer patterns. All the newly obtained polynomials are plotted and are compared with the standard orthogonal polynomials obtained through the Gram-Schmidt orthogonalization procedure.

Quite understandably, the biggest difference arises in the Löwdin-Bessel and Löwdin-Laguerre polynomials when compared with their standard orthogonal polynomials. There are some shortcomings in finding the eigenvalues and eigenvectors of the Hermitian metric matrix $\boldsymbol{M}$ for higher values of $\boldsymbol{N}$. An unfavorable feature of the Gram matrix $\boldsymbol{M}$ is that it begins to develop negative eigenvalues beyond a certain order $\boldsymbol{N}$ of the monomials.

Löwdin-Legendre polynomials obtained through *symmetric* and *canonical* orthogonalization procedures are compared with the standard Legendre polynomials obtained through Gram-Schmidt orthogonalization procedure for $\boldsymbol{N = 2, 3}$ and $\boldsymbol{4}$ (i.e. $\boldsymbol{3, 4}$ and $\boldsymbol{5}$ monomials). All the polynomials are normalized to unity. Note that alternate polynomials in a set repeat in the subsequent set. This is depicted by different types of arrows. This is shown in Table 4.2 for Legendre polynomials, Table 4.3 for Hermite polynomials and Table 4.4 for Laguerre polynomials.

In sum, the general concept of orthogonal polynomials has been introduced describing many new classes of orthogonal polynomials. Two new sets of polynomials can be easily obtained by applying the concept of Löwdin orthogonalization methods. This theoretical framework offers the possibility to search for new iterative approaches where desired properties can be immediately obtained from the theoretical investigations. Moreover, a classification of the qualitative behaviour of these polynomials with a different view point has been introduced. The Löwdin orthogonalization concepts and these new classes of orthogonal polynomials could be the basis to develop new methods and new polynomials. From these results, research for efficient and robust techniques could be stimulated. With the new classes of orthogonal polynomials, the subject of orthogonal polynomials will witness the possible openings of several new avenues of research. They may be open problems in numerical analysis, solutions of non linear differential equations, least-square curve fitting etc.

Chapter 5

The Centrality of Löwdin Orthogonalization Methods

5.1 Introduction

Löwdin orthogonalization methods [1] were discovered for the purpose of orthogonalizing hybrid electron orbits in quantum chemistry. A few other orthogonalization methods have been developed independently by mathematicians and computer scientists to deal with specific problems.

In this chapter, we show different relationships between the Löwdin orthogonalizations and other orthogonalization techniques. We have found that other orthogonalization methods such as the polar decomposition, introduced by Autonne in 1902 [15], the principal component analysis (PCA) [6] and the singular value decomposition (SVD) [14] can be derived from Löwdin methods. We have analytically found that the polar decomposition is contained in the symmetric orthogonalization. We have also found that the *canonical orthogonalization* can be reduced to the form of *reduced SVD* or vice-versa. The analytic relations between the symmetric and canonical orthogonalization methods are also established.

5.2 Polar Decomposition

The polar decomposition of a matrix $\mathbf{V} \in \mathbb{R}^{n \times m}$, where $\mathbb{R}$ is a set of real numbers, can be obtained from the Löwdin's symmetric orthogonalization. The symmetric orthonormal basis of a matrix $\boldsymbol{V}$ is expressed as (recall eqn. (2.12))

$$\boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{M}^{-1/2}, \tag{5.1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{M} \in \mathbb{R}^{m \times m}$. Now multiplying both sides of eqn. (5.1) from right by $\boldsymbol{M}^{1/2}$, we get

$$\begin{aligned} \boldsymbol{\Phi}\boldsymbol{M}^{1/2} &= \boldsymbol{V}\boldsymbol{M}^{-1/2}\boldsymbol{M}^{1/2}, \\ \text{or}, \boldsymbol{V} &= \boldsymbol{\Phi}\boldsymbol{M}^{1/2}. \end{aligned} \tag{5.2}$$

This is called as the polar decomposition [15] of the matrix $\boldsymbol{V}$.

5.3 Principal Component Analysis

For $\boldsymbol{N}$ vectors in $\boldsymbol{n}$ dimensions, we have found that the principal components obtained through the principal component analysis of a square matrix $\boldsymbol{V}$ are equal to the orthonormal vectors obtained through the Löwdin's canonical orthogonalization.

Let $\boldsymbol{V} \in \mathbb{R}^{N \times n}$, then the sum of squares and cross-prodcuts (SSCP) mattrix $\boldsymbol{S}$ can be constructed using $\boldsymbol{V}\boldsymbol{V}^{\boldsymbol{T}}$. The SSCP matrix is the covariance matrix without subtracting the mean. The diagonal values are sums of squares and the off-diagonal values are sums of cross products. The eigenvalues and eigenvectors of the SSCP matrix $\boldsymbol{S}$ are constructed. We have found that the eigenvalues are the same as the eigenvalues of the Gram matrix constructed using $\boldsymbol{V^T V}$ in the case of Löwdin orthogonalizations. The eigenvectors in the two cases are different. However, the eigenvectors of the SSCP matrix $\boldsymbol{S}$, called as the principal components of $\boldsymbol{V}$, are the same as those obtained using the canonical orthogonalized set $\boldsymbol{\Lambda} = \boldsymbol{VUd^{-1/2}}$. The eigenvectors of $\boldsymbol{S}$ are ordered so that the first two principal components retain most of the variation present in the original set of vectors. We have computed the sum of the projection-squares of the given vectors onto principal components. This gives the eigenvalues of $\boldsymbol{S}$ and is the same as the sum of projection-squares of the original vectors on the canonical orthonormal vectors, i.e., the eigenvalues of $\boldsymbol{M}$. Hence in the case of square matrices, the principal component analysis of the pure SSCP matrix is equivalent to the canonical orthogonalization.

5.4 Singular Value Decomposition

The singular value decomposition is the most commonly applicable orthogonalization technique for the orthogonal-diagonal-orthogonal type matrix decompositions. The importance of SVD is widely known in various application fields such as in numerical linear algebra, least-squares problem and data compression. The SVD contains a great deal of information and is very useful as a theoretical and practical tool for data visualization. The singular value decomposition of a non-singular matrix $\boldsymbol{V}$ can be obtained from the Löwdin's canonical orthogonalization.

The canonical orthogonalization of a matrix $\boldsymbol{V}$ **is** written as (recall eqn. (2.17))

$$\boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{U}\boldsymbol{d}^{-1/2}. \tag{5.3}$$

Multiplying both sides of equation (5.3) from its right with $\boldsymbol{d}^{1/2}$, we have

$$\begin{aligned} \boldsymbol{\Lambda}\boldsymbol{d}^{1/2} &= \boldsymbol{V}\boldsymbol{U}\underbrace{\boldsymbol{d}^{-1/2}\boldsymbol{d}^{1/2}} \\ \boldsymbol{\Lambda}\boldsymbol{d}^{1/2} &= \boldsymbol{V}\boldsymbol{U}\boldsymbol{I} = \boldsymbol{V}\boldsymbol{U}. \end{aligned} \tag{5.4}$$

Now multiplying equation (5.4) on both sides from right with $\boldsymbol{U}^{\boldsymbol{T}}$, we get

$$\begin{aligned} \boldsymbol{\Lambda}\boldsymbol{d}^{1/2}\boldsymbol{U}^{\boldsymbol{T}} &= \boldsymbol{V}\underbrace{\boldsymbol{U}\boldsymbol{U}^{\boldsymbol{T}}} \\ \text{or}, \ \boldsymbol{V} &= \boldsymbol{\Lambda}\boldsymbol{d}^{1/2}\boldsymbol{U}^{\boldsymbol{T}} \end{aligned} \tag{5.5}$$

This is known as singular value decomposition and is called as the reduced SVD form of the canonical orthogonalization.

5.5 Analytic Relations between Symmetric and Canonical Orthogonalizations

We can analytically obtain the relationships between symmetric and canonical orthogonalizations. If we know one of them from the Gram matrix, say canonical or symmetric, then the other can be obtained from the following fundamental relations. With the help of these relations, we have shown that the symmetric orthogonalization is contained in the reduced form of the singular value decomposition.

The symmetric orthonormal basis is given by

$$\boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{M}^{-1/2}. \tag{5.6}$$

From the equation

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{A}, \tag{5.7}$$

We have

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{M}^{-1/2}\boldsymbol{U} = \boldsymbol{\Lambda} \tag{5.8}$$

Or

$$\boldsymbol{\Lambda} = \underbrace{\boldsymbol{V}\boldsymbol{M}^{-1/2}}\boldsymbol{U} = \boldsymbol{\Phi}\boldsymbol{U}. \tag{5.9}$$

This relation is useful to construct the canonical orthonormal basis directly using symmetric orthonormal basis and the eigenvectors of the Hermitian metric matrix.

We can also directly obtain the symmetric orthonormal basis from the canonical orthonormal basis by multiplying both sides of equation (5.9) from its right with $\boldsymbol{U}^{\dagger}$, as shown below

$$\begin{aligned} \boldsymbol{\Lambda}\boldsymbol{U}^{\dagger} &= \boldsymbol{\Phi}\underbrace{\boldsymbol{U}\boldsymbol{U}^{\dagger}} && (5.10)\\ \boldsymbol{\Lambda}\boldsymbol{U}^{\dagger} &= \boldsymbol{\Phi}\boldsymbol{I} = \boldsymbol{\Phi}. && (5.11)\\ \text{Or, } \boldsymbol{\Phi} &= \boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}. && (5.12) \end{aligned}$$

5.6 Symmetric Orthogonalization in SVD

We can analytically obtain the symmetric orthogonalization from the singular value decomposition.

Let $\boldsymbol{V} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{m}}$ with $\boldsymbol{n} \geq \boldsymbol{m}$ be a non-singular matrix. Then the singular value decomposition of $\boldsymbol{V}$ can be written from (5.5) as

$$\boldsymbol{V} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & w_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mm} \end{pmatrix}$$

$$\equiv \boldsymbol{W}\,\boldsymbol{\Sigma}\,\boldsymbol{U}^{\boldsymbol{T}}\ (\boldsymbol{\Sigma} = \boldsymbol{d}^{1/2}),$$

where $\boldsymbol{W} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ are matrices with their columns as orthonormal vectors. The columns $\boldsymbol{W_j},\ j = 1, 2, \cdots, m$ of $\boldsymbol{W}$ are called *left singular vectors* of $\boldsymbol{V}$ and the columns of $\boldsymbol{U_j},\ j = 1, 2, \cdots, m$ of $\boldsymbol{U}$ (or rows of $\boldsymbol{U^T}$)are called *right singular vectors* of $\boldsymbol{V}$. And $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ is square and diagonal matrix with $\boldsymbol{\sigma_i}$'s as the *singular values* of $\boldsymbol{V}$. By convention, the singular values are arranged in a descending order as $\boldsymbol{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0}$. This form of singular value decomposition is known as the *reduced singular value decomposition.* This reduced SVD can give the symmetric orthogonalization. Using the analytic relationship between the symmetric and canonical orthogonalizations, the symmetric orthogonalization of the matrix $\boldsymbol{V}$ can be written from the reduced SVD as follows,

$$\boldsymbol{\Phi} = \boldsymbol{W}\boldsymbol{U}^{\boldsymbol{T}}, \tag{5.13}$$

$$\boldsymbol{\Phi}_{ij} = \boldsymbol{W}_{i1}\left(\boldsymbol{U}^{\boldsymbol{T}}\right)_{1j} + \boldsymbol{W}_{i2}\left(\boldsymbol{U}^{\boldsymbol{T}}\right)_{2j} + \cdots + \boldsymbol{W}_{im}\left(\boldsymbol{U}^{\boldsymbol{T}}\right)_{mj} \tag{5.14}$$

$$= \boldsymbol{W}_{i1}\left(\boldsymbol{U}\right)_{j1} + \boldsymbol{W}_{i2}\left(\boldsymbol{U}\right)_{j2} + \cdots + \boldsymbol{W}_{im}\left(\boldsymbol{U}\right)_{jm}, \tag{5.15}$$

where $\boldsymbol{W}$ is the same as the canonical orthonormal basis $\boldsymbol{\Lambda}$. The symmetric orthonormal basis $\boldsymbol{\Phi}$ is *unique* since it takes the linearly independent set of columns of $\boldsymbol{V}$ as input and gives the orthonormal columns of $\boldsymbol{WU^T}$ as output.

Chapter 6

Löwdin Orthogonalization Methods to Analyze Microarray Gene Expression Data

6.1 Introduction

Microarray experiments provide large quantities of genome-wide data on the gene expression patterns. The microarray data analysis gives a better understanding of gene expression patterns and their regulatory mechanisms. The knowledge of various processes at a molecular level can advance research on diagnosis and drug design of diseases such as cancer, alzheimer's and parkinson's etc. It is therefore important to develop computational methods to analyze the cluster microarray data and extract meaningful biological information. During the machine process of *c*DNA microarray chips, some spots on the microarray may be missed due to several factors. Since the repetition of experiment is very costly and also time consuming, molecular biologists, mathematicians, statisticians, computer scientists and physicists make attempts to recover the missing gene expressions by systematic analysis.

Various computational tools are available such as the Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) etc. But the methods that are used to analyze the data can have profound influence on the interpretation of the results. A basic understanding of these computational tools is therefore necessary for the optimal experimental design and meaningful data analysis. We have used the two orthogonalization methods developed by P. O. Löwdin [1], namely the 'symmetric' and 'canonical' orthogonalization methods, to analyze the large volumes of available microarray gene expression data and to interpret the results.

Microarray gene expression data are organized as a gene expression matrix $\boldsymbol{V}$ with $\boldsymbol{n}$ rows, which correspond to genes, and $\boldsymbol{m}$ columns, which correspond to different experiments. Typically, the number of genes, $\boldsymbol{n}$, is much larger than the number of experiments, $\boldsymbol{m}$. From the gene expression data matrix, the analysis of missing gene expressions on the array would translate to recover the missed entries in the gene expression matrix values.

In this chapter, we have applied two orthogonalization methods in which the estimation of missing entries is done simultaneously, i.e. the estimation of one missing entry influences the estimation of the other missing entries. If the gene expression data matrix $\boldsymbol{V}_{\boldsymbol{n}\times\boldsymbol{m}}$ has missing data, we want to complete its entries to obtain a matrix $\hat{\boldsymbol{V}}$, such that the rank of $\hat{\boldsymbol{V}}$ is equal to $\boldsymbol{d}^{1/2}$, where $\boldsymbol{d}^{1/2}$ is taken to be the number of significant singular values of $\boldsymbol{V}$.

6.2 Biological Background

Cells and organisms are basically divided into two classes; procaryotic (such as bacteria), which do not have nucleus and eucaryotic [16], which have a nucleus. The cell is enclosed by a membrane; embedded in the cell's cytoplasm in its nucleus, surrounded and protected by its own membrane. The nucleus contains DNA (Deoxyrybo Nuclei Acid), which has four kinds of bases, denoted by $\boldsymbol{A}$ (Adenine), $\boldsymbol{G}$ (Guanine), $\boldsymbol{C}$ (Cyanine) and $\boldsymbol{T}$ (Thymine). The two strands contain complimentary base sequences and are held together by hydrogen bonds that connect matching pairs of bases; $\boldsymbol{A}-\boldsymbol{T}$ (two hydrogen bonds) and $\boldsymbol{G}-\boldsymbol{C}$ (three hydrogen bonds).

6.2.1 Genes and Gene Expression

A *gene* is a segment of DNA, which contains the formula for the chemical composition of one particular protein. Proteins are the fundamental working molecules of life. Most biological processes that take place in a cell are carried out by proteins. Topologically, a protein is a chain; each link is an amino acid, with neighbours along the chain connected by covalent bonds. All the proteins are made up of 20 different amino acids. So the chemical formula of a protein of length N is an N-letter word, whose letters are taken from a 20 letter alphabet. A *gene* lists the order in which the amino acids are to be strung when the corresponding protein is synthesized. Genetic information is encoded in the linear sequence in which the base pairs on the two strands of DNA are ordered along the molecule. The genetic code is a universal translation table with specific triplets of consecutive bases coding for every amino acid.

The *genome* contains the collection of all the genes that code the chemical formulae of all the proteins (and RNA) that an organism needs and produces. The genome of a sample organism such as yeast contains about 6000 genes; the human genome has between 30,000 to 40,000. It was established that an overwhelming majority (98%) of human DNA contains non-coding regions, i.e., strands that do not code for any particular protein. But they play a role in the regulation level of synthesis of the different proteins. Every cell of a multicellular organism contains its entire genome. Even though each cell contains the same set of genes, there is a *differentiation*. Cells of complex organism taken from different organs, have entirely different functions and proteins that perform these functions are entirely different. For example the cells in our retina need photosensitive molecules whereas the colour of our hair does not contain these molecules.

A gene is expressed in a cell when the protein it encodes for is synthesized. In an average human cell, about 10,000 genes are expressed. The set of genes that indicate the expression level of each of these levels is called the *expression profile* of the cell.

The large majority of abundantly expressed genes are associated with common functions, such as metabolism, and hence co-expressed in all cells. However, there will be differences between the expression profiles of different cells, and even in single cell, expression of genes will vary with time, in a manner dictated by external and internal probes that reflects the state of the organism and the cell itself.

6.2.2 Transcription and Translation

The obvious solution of information transfer would be to rip out the piece of DNA that contains the gene that is to be expressed, and transport it to the cytoplasm. When a gene receives a function to be expressed, corresponding segment of the double helix of DNA opens and a precise copy of the information, as written on one of the strands is prepared which is called the *messenger* RNA. The process of its production is called *transcription*. The subsequent reading of $\boldsymbol{m}$RNA, deciphering the message, written using the base triplets, into amino acids and synthesis of the corresponding proteins at the ribosomes is called *translation* [17]. The ribosomes, where the synthesization of proteins takes place, are the mechanics that read the information written on the DNA and synthesize the proteins according to the instructions.

When a large number of molecules of a certain protein are needed, the cell produces many corresponding $\boldsymbol{m}$RNAs. These $\boldsymbol{m}$RNAs are transformed to the cytoplasm through the nuclear membrane and are *read* by several ribosomes. Thus the single master copy of the instructions contained in the DNA, generates many copies of the proteins. This transcription strategy is prudent and safe, preserving the previous master copy. At the same time it also serves as a remarkable amplifier of genetic information.

6.3 DNA Microarrays

A DNA chip is the instrument that measures simultaneously the concentration of thousands of $\boldsymbol{m}$RNA molecules. It is also referred to as DNA microarray [16].

DNA microarrays can measure simultaneously the expression levels of up to 30,000 genes. Schematically, the microarrays are produced as follows. A chip (a glass plate of about 10 mm across) is divided into *pixels* – each pixel is assigned to one gene. Several millions of 25 base-pair long pieces, called oligonucleotides, of single strand DNA are, copied from a particular segment of gene. These are photolitigraphically synthesized on the assigned pixel (referred to as *probes*). The $\boldsymbol{m}$RNA molecules are extracted from the cells taken from the tissue of interest such as a cancer tissue or brain tumor tissue etc. Next, the resulting DNA is transcribed back into fluorescently marked single strand RNA diffuse over the dense forest of single strand DNA probes. When such an $\boldsymbol{m}$RNA encounters a bit of the probe,of which the RNA is a perfect copy, it hybridizes to this strand - i.e., attaches to it with a high affinity. When the $\boldsymbol{m}$RNA solution is washed off, only those molecules that find their perfect match remain stuck to the chip. Now the chip is illuminated by a laser, and these stuck *targets* fluoresce. Now by measuring the light intensity emanating from each pixel, one can obtain a measure of the number of tangents that stuck, which form a chip on which $\boldsymbol{n}$ genes are placed. These number $\boldsymbol{n}$ represents the expression level of these genes in that tissue. A typical experiment provides the expression profiles of several tens of samples ($\boldsymbol{m \approx 100}$), over several thousand ($\boldsymbol{n}$) genes. These numbers are summarized in an $\boldsymbol{n \times m}$ expression table. Each entry of such an expression table stands for the expression level of a particular gene in a particular experiment.

6.4 Gene Expression Data Matrix

The gene expression data matrix consists of the expression levels of a large number of genes at a smaller number of time points. The gene expression data matrix $\boldsymbol{V}$ consists of $\boldsymbol{n}$ rows and $\boldsymbol{m}$ columns. Each row represents a gene and each column a time point. Hence the data can be viewed as: each row as a gene vector or each column as a time vector. Generally the data is taken on large number of genes and less number of time points. We have pre-processed the data by replacing each measurement with its logarithm, and normalizing each genes transcriptional response to have zero mean and unit standard deviation. The matrix $\boldsymbol{V}$, an $\boldsymbol{n} \times \boldsymbol{m}$ matrix $\in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{m}}$ is a gene expression data matrix with $\boldsymbol{n} \geq \boldsymbol{m}$ and can be expressed as follows,

$$\boldsymbol{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{pmatrix} = \begin{pmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_n^T \end{pmatrix} = (\vec{t}_1 \ \vec{t}_2 \ \cdots \ \vec{t}_m) \tag{6.1}$$

where,

$$\vec{v}_j^T = (v_{j1} \ v_{j2} \ \cdots \ v_{jm}), \ j = 1, 2, \cdots, n \tag{6.2}$$

and

$$\vec{t}_i = \begin{pmatrix} v_{1i} \\ v_{2i} \\ \vdots \\ v_{ni} \end{pmatrix}, \ i = 1, 2, \cdots, m. \tag{6.3}$$

The row vector $\vec{v}_j^T$ corresponds to the relative expression levels of the j^{th} gene in m experiments. The column vector $\vec{t}_i$ corresponds to the relative expression levels of the n genes in the i^{th} experiment.

This matrix is very difficult to analyze and one cannot get any useful information from this raw data regarding the functioning of a group of genes. The raw data has to be polished by doing the normalization procedure for the rows and columns, ending with row normalization. Now the aim is to analyze this normalized data to get some biological functioning of a large set of genes. Our aim is to bring out the clusters in the gene expression patterns using the Löwdin orthogonalization techniques. We have applied these democratic schemes to the gene expression data matrix and found some genes which are co-expressed. We have found that the clusters can be obtained by plotting the first versus the second symmetric and canonical orthonormal bases. We also found that the genes do exhibit the simple harmonic motion in their behaviour when we project the time vectors onto the orthonormal bases. Hence the Löwdin methods give more insight into the oscillatory behaviour of the genes when they are co-expressing.

6.5 Analysis of Gene Expression Data

The orthogonalization methods due to Löwdin [1, 5] are applied to three data sets of microarray gene expression experiments in order to understand the possible clusters in the data sets. We have considered data sets of yeast cell-cycle (cdc15-15 points and cdc15-12 points), sporulation and fibroblast from the literature [18].

The Hermitian metric matrix **M** is constructed using the eqn. (2.5) as follows.

$$\boldsymbol{M} = \boldsymbol{V^T V} = \begin{pmatrix} \vec{v}_1.\vec{v}_1 & \vec{v}_1.\vec{v}_2 & \cdots & \vec{v}_1.\vec{v}_m \\ \vec{v}^2.\vec{v}^1 & \vec{v}^2.\vec{v}^2 & \cdots & \vec{v}^2.\vec{v}_m \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_m.\vec{v}_1 & \vec{v}_m.\vec{v}_2 & \cdots & \vec{v}_m.\vec{v}_m \end{pmatrix}. \quad (6.4)$$

Each component of the Hermitian metric matrix contains mixture of the expression levels of all the genes of two time points and the diagonal elements contain the expression levels of all the genes at the same time point, i.e., the expression levels of all the genes at a particular time point are overlapped along the diagonal. Its eigenvalues and eigenvectors are calculated. Let $\boldsymbol{d}$ be an $\boldsymbol{m \times m}$ diagonal matrix which contains eigenvalues of the Gram matrix in a descending order and $\boldsymbol{U}$ be an $\boldsymbol{m \times m}$ matrix whose columns represent the normalized eigenvectors. The square roots of these matrices and their respective inverse matrices are constructed by using the transformations discussed in Chapter 2. They are given by $\boldsymbol{M^{-1/2}}$ and $\boldsymbol{d^{-1/2}}$ respectively as follows.

$$\boldsymbol{M^{-1/2}} = \begin{pmatrix} M_{11}^{-1/2} & M_{12}^{-1/2} & \cdots & M_{1m}^{-1/2} \\ M_{21}^{-1/2} & M_{21}^{-1/2} & \cdots & M_{2m}^{-1/2} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1}^{-1/2} & M_{m1}^{-1/2} & \cdots & M_{mm}^{-1/2} \end{pmatrix}, \quad (6.5)$$

$$\boldsymbol{d^{-1/2}} = \begin{pmatrix} d_{11}^{-1/2} & 0 & \cdots & 0 \\ 0 & d_{22}^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{mm}^{-1/2} \end{pmatrix}. \quad (6.6)$$

These two matrices are used to construct the canonical and symmetric orthonormal basis for the gene expression data matrix $\boldsymbol{V}$. They are given by

$$\boldsymbol{\Lambda} = \boldsymbol{V U d}^{-\frac{1}{2}} = \left[\vec{\boldsymbol{\lambda}}_1, \vec{\boldsymbol{\lambda}}_2, \cdots, \vec{\boldsymbol{\lambda}}_m\right], \tag{6.7}$$

$$\boldsymbol{\Phi} = \boldsymbol{V M}^{-\frac{1}{2}} = \left[\vec{\boldsymbol{\phi}}_1, \vec{\boldsymbol{\phi}}_2, \cdots, \vec{\boldsymbol{\phi}}_m\right]. \tag{6.8}$$

The canonical and symmetric orthonormal bases consist of $\boldsymbol{m}$, $\boldsymbol{n}$-dimensional vectors. Each $\vec{\boldsymbol{\lambda}}$ or $\vec{\boldsymbol{\phi}}$ consists of information of all the genes at all the times. The symmetric orthonormal basis vectors are completely orthogonal and normal to each other if the given $\vec{\boldsymbol{v}}$'s are completely linearly-independent. If there is a dependency between any of the time points, it leads to near orthogonality of the orthonormal basis.

To analyze the data, recall,

$$\boldsymbol{V} = \boldsymbol{\Lambda}\ \boldsymbol{\Lambda}^{\boldsymbol{T}}\ \boldsymbol{V} = \boldsymbol{\Lambda}\ \boldsymbol{X}, \tag{6.9}$$

$$\text{where, } \boldsymbol{X} = \begin{pmatrix} \vec{\boldsymbol{X}}_1(\boldsymbol{t}) \\ \vec{\boldsymbol{X}}_2(\boldsymbol{t}) \\ \vdots \\ \vec{\boldsymbol{X}}_m(\boldsymbol{t}) \end{pmatrix} = \begin{pmatrix} \vec{\boldsymbol{v}}_1.\vec{\boldsymbol{\lambda}}_1 & \vec{\boldsymbol{v}}_2.\vec{\boldsymbol{\lambda}}_1 & \cdots & \vec{\boldsymbol{v}}_m.\vec{\boldsymbol{\lambda}}_1 \\ \vec{\boldsymbol{v}}_1.\vec{\boldsymbol{\lambda}}_2 & \vec{\boldsymbol{v}}_2.\vec{\boldsymbol{\lambda}}_2 & \cdots & \vec{\boldsymbol{v}}_m.\vec{\boldsymbol{\lambda}}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \vec{\boldsymbol{v}}_1.\vec{\boldsymbol{\lambda}}_m & \vec{\boldsymbol{v}}_2.\vec{\boldsymbol{\lambda}}_m & \cdots & \vec{\boldsymbol{v}}_m.\vec{\boldsymbol{\lambda}}_m \end{pmatrix} \tag{6.10}$$

represents the set of characteristic modes, $\{\boldsymbol{X}_i(\boldsymbol{t})\}$ associated with $\boldsymbol{V}$ as obtained from canonical orthogonalization. At the same time from the symmetric orthogonalization we have,

$$\boldsymbol{V} = \boldsymbol{\Phi}\ \boldsymbol{M}^{-1/2} = \boldsymbol{\Phi}\ \boldsymbol{V}^{\boldsymbol{T}}\ \boldsymbol{\Phi} = \boldsymbol{\Phi}\ \boldsymbol{\Phi}^{\boldsymbol{T}}\ \boldsymbol{V} \tag{6.11}$$

and in analogy with (6.9) we can represent characteristic modes for symmetric orthogonalization as,

$$\boldsymbol{\Phi^T\ V} = \begin{pmatrix} \vec{v}_1.\vec{\phi}_1 & \vec{v}_2.\vec{\phi}_1 & \cdots & \vec{v}_m.\vec{\phi}_1 \\ \vec{v}_1.\vec{\phi}_2 & \vec{v}_2.\vec{\phi}_2 & \cdots & \vec{v}_m.\vec{\phi}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_1.\vec{\phi}_m & \vec{v}_2.\vec{\phi}_m & \cdots & \vec{v}_m.\vec{\phi}_m \end{pmatrix}. \tag{6.12}$$

The temporal variation of the genes in $\boldsymbol{V}$ can thus be represented as linear combination of the characteristic modes – either (6.10) or (6.12) – with $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ as coefficients.

Note that the characteristic modes reflect the genome-wide expression pattern which are not gene-specific. This is because the $\{\vec{\Lambda}_i\}$ and $\{\vec{\phi}_i\}$ are constructed using all $\{\vec{v}_i\}$. So, they contain complete temporal information of expression levels of *all* the given genes under consideration in a convoluted manner. The coefficients that multiply the elements of the matrices in (6.10) and (6.12) are the matrix elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Phi}$ which are obtained typically by taking the projections of $\vec{\Lambda}_i$'s on $\vec{z}_j$'s, where $\vec{z}_j$ represents the temporal variation of the expression level of the j^{th} gene. Thus the coefficients are gene-specific.

As compared to the analyses [18, 19] that attempted to classify the genes based on their expression profiles the present analyses based on Löwdin's orthogonalization techniques give us physical insight into what entails the characteristic modes.

In the canonical orthogonalization approach we know that the basis set comprising $\vec{\Lambda}_i$'s is constructed such that one of them, say $\vec{\Lambda}_1$, samples the largest set of $\vec{v}_i$'s, another of them, say $\vec{\Lambda}_2$, samples the next largest set of $\vec{v}_i$'s, and so on, such that the projection squares of $\vec{v}_i$'s on $\vec{\Lambda}_1$ add up to a maximum, those on $\vec{\Lambda}_2$ to a smaller number, and so on. Note that the projections of $\vec{v}_i$'s on $\vec{\Lambda}_1, \vec{\Lambda}_2,$ etc.

will be large as well as small, and positive as well as negative, seemingly in a random manner. However, in the present context the Figures 6.1 (a-c) show that all these large and small, positive and negative projections (as in eqn. (6.10)) follow a periodic behaviour as a function of time.

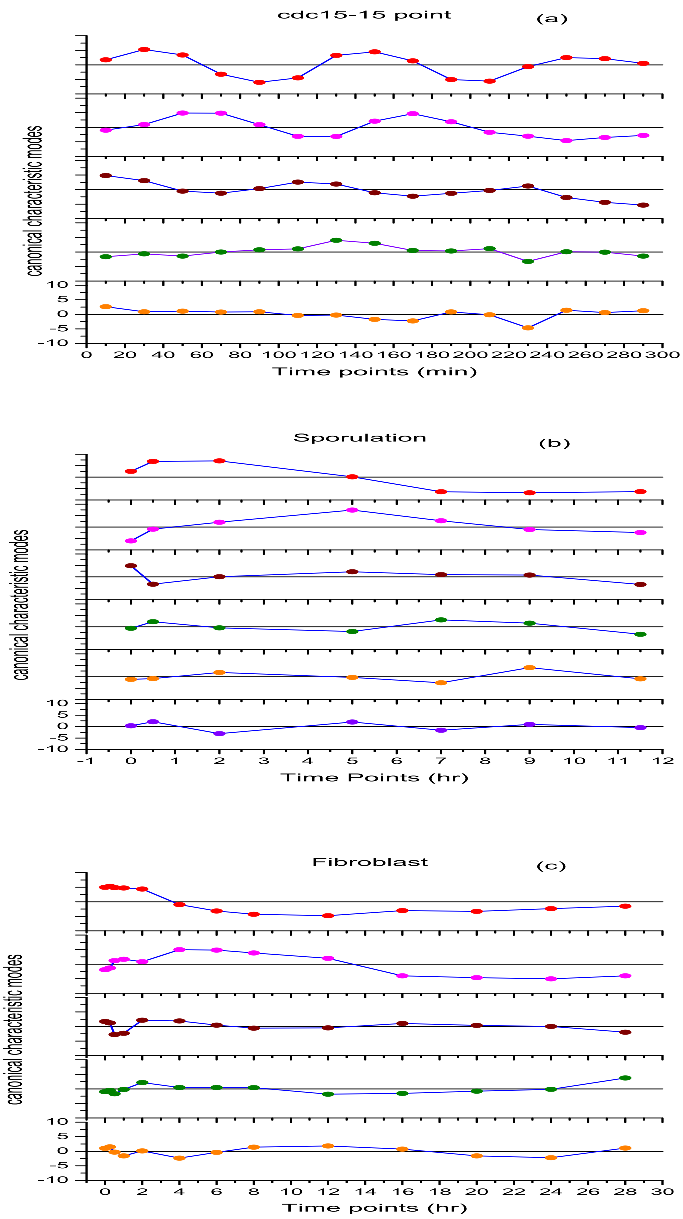


Figure 6.1: Canonical characteristic modes of (a) yeast cell-cycle cdc15-15 point, (b) Sporulation and (c) Fibroblast data.

Since $\vec{v}_i$'s represent *all* the genes at different time points the observed periodic behaviour points to a fact that the genes under study express as a function of time in a highly correlated manner.

Moreover, it is found that the entire given lot of genes can be divided up into primarily two classes. Their expression profiles exhibit similar $\boldsymbol{t}$-behaviour, namely oscillatory, but they are shifted by almost $\boldsymbol{\pi/2}$, i.e. at the time when the first lot shows zero or nearly zero expression (i.e. maximum repression) the other lot shows maximum expression. Also, between the two lots the second lot whose $\boldsymbol{t}$-vectors cluster around $\vec{\boldsymbol{\lambda}}_{\mathbf{2}}$ shows a weaker correlation as compared to the lot that clusters around $\vec{\boldsymbol{\lambda}}_{\mathbf{1}}$ apparently because the number of vectors diminishes as we go from $\vec{\boldsymbol{\lambda}}_{\mathbf{1}}$ to $\vec{\boldsymbol{\lambda}}_{\mathbf{2}}$. In fact, any periodic pattern is hardly seen beyond $\vec{\boldsymbol{\lambda}}_{\mathbf{2}}$ (see Fig. 6.1 (a)). Further grouping (or clustering) of genes becomes more apparent when we plot the coefficients $\vec{\boldsymbol{\lambda}}_{\mathbf{1}}$ and $\vec{\boldsymbol{\lambda}}_{\mathbf{2}}$ of the first two characteristic modes against each other as in Figures 6.2 (a), (b), (c) and (d). Clusters are seen to be formed along the periphery of the ellipses.

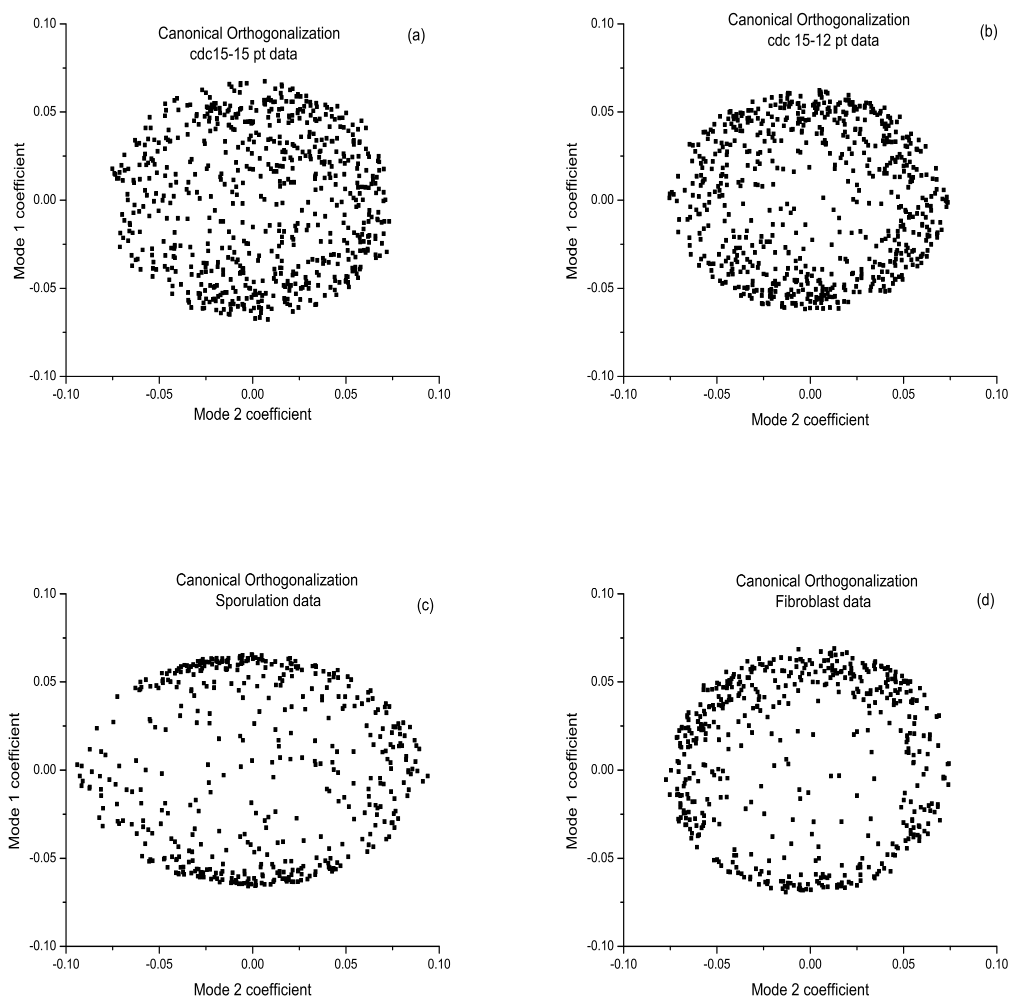


Figure 6.2: Plot of the coefficients for characteristic mode 1 against the coefficients for characteristic mode 2 for (a) cdc15-15 point, (b) cdc15-12 point, (c) Sporulation and (d) Fibroblast data using canonical orthogonalization.

The scenario with the symmetric orthogonalization applied to the given gene-expression data is entirely different from the above. While the canonical orthogonalization tries to classify or categorise (or make groups of) co-expressing genes in temporal terms, the symmetric orthogonalization does the opposite – it tends to scatter the data though in a very special manner. Most importantly all the characteristic modes become equivalent in the sense that the matrix (6.12) is symmetric about the diagonal; so all of them look almost alike except that they show a peak at the diagonal element, which naturally shifts one step at a time as we switch from one mode to the next. The modes show a gradual but systematic decay on either side of the peak at the diagonal element (see Figures 6.3 (a-c)). Thus the correlated behaviour in the collection of genes is again witnessed as a function of time but it is monotonic and almost the same in all characteristic modes.

The symmetric orthogonalization organises the basis set $\{\vec{\phi}_i\}$ such that $\vec{v}_i$ becomes in a sense equivalent to $\vec{\phi}_i$ presents us with a special scenario in which projections of all $\vec{v}_i$'s on a particular $\vec{\phi}_j$ are element by element equal to the projections of all $\vec{\phi}_i$'s on $\vec{v}_j$. The large magnitudes of $\vec{v}_j.\vec{\phi}_j$ as compared to rapidly diminishing values of all the rest of the elements $\vec{v}_i.\vec{\phi}_j$ as well as $\vec{v}_j.\vec{\phi}_i$ (as shown in Figures 6.3 (a-c)) indicates that $\vec{\phi}_j$ is maximally aligned along $\vec{v}_j$, i.e. as close to it as possible, such that the orthonormality of $\vec{\phi}_i$'s is also achieved.

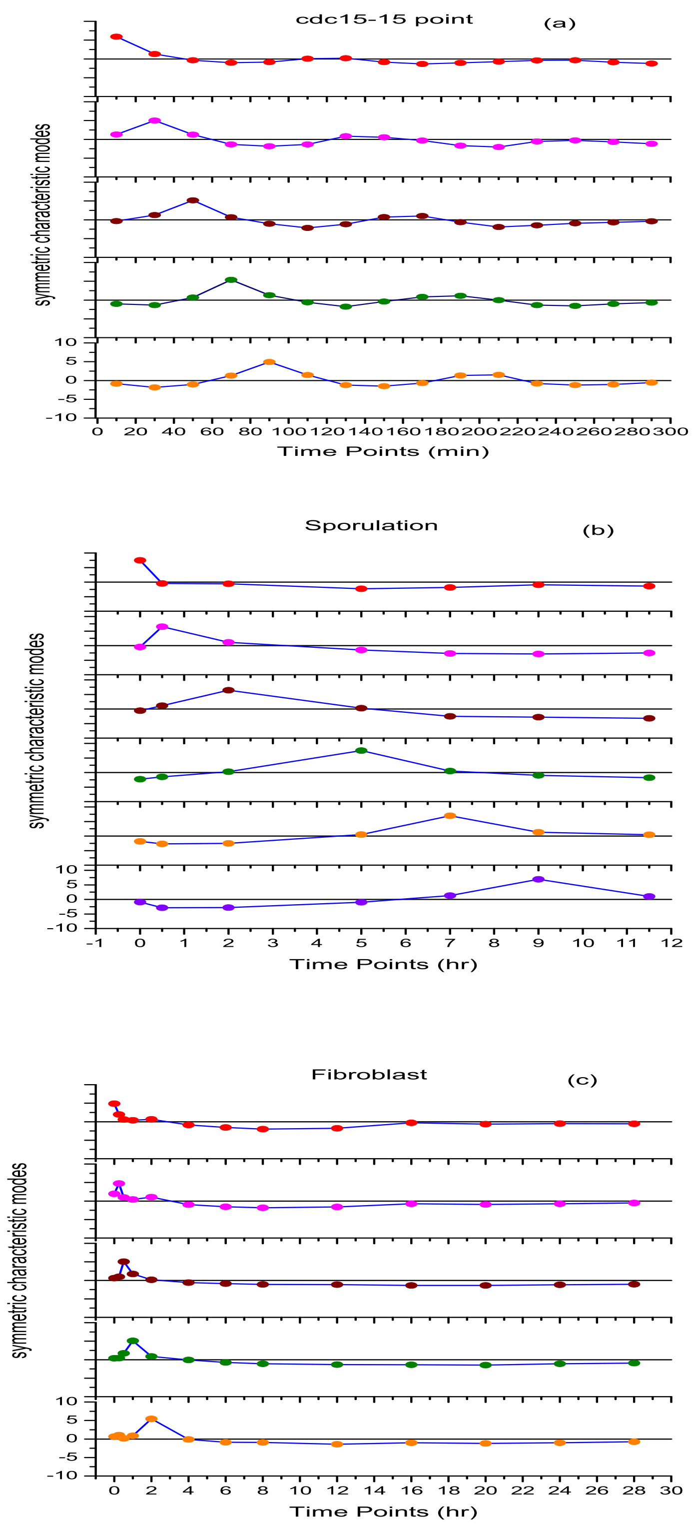


Figure 6.3: Symmetric characteristic modes of (a) yeast cell-cycle cdc15-15 point, (b) Sporulation and (c) Fibroblast data.

In Figures 6.4 (a, b, c; i, ii, iii) we have made **2**-d plots of coefficients $\boldsymbol{\phi_1} - \boldsymbol{\phi_2}$, $\boldsymbol{\phi_1} - \boldsymbol{\phi_3}$ and $\boldsymbol{\phi_1} - \boldsymbol{\phi_4}$ for cdc15-15 point, sporulation and fibroblast data sets. While the $\boldsymbol{\phi_1}$ *vs.* $\boldsymbol{\phi_2}$ plots show some clustering of genes, the data only gets increasingly smeared as we combine $\boldsymbol{\phi_1}$ with $\boldsymbol{\phi_3}$ and $\boldsymbol{\phi_4}$. No definite information is obtained but we are investing the data further.

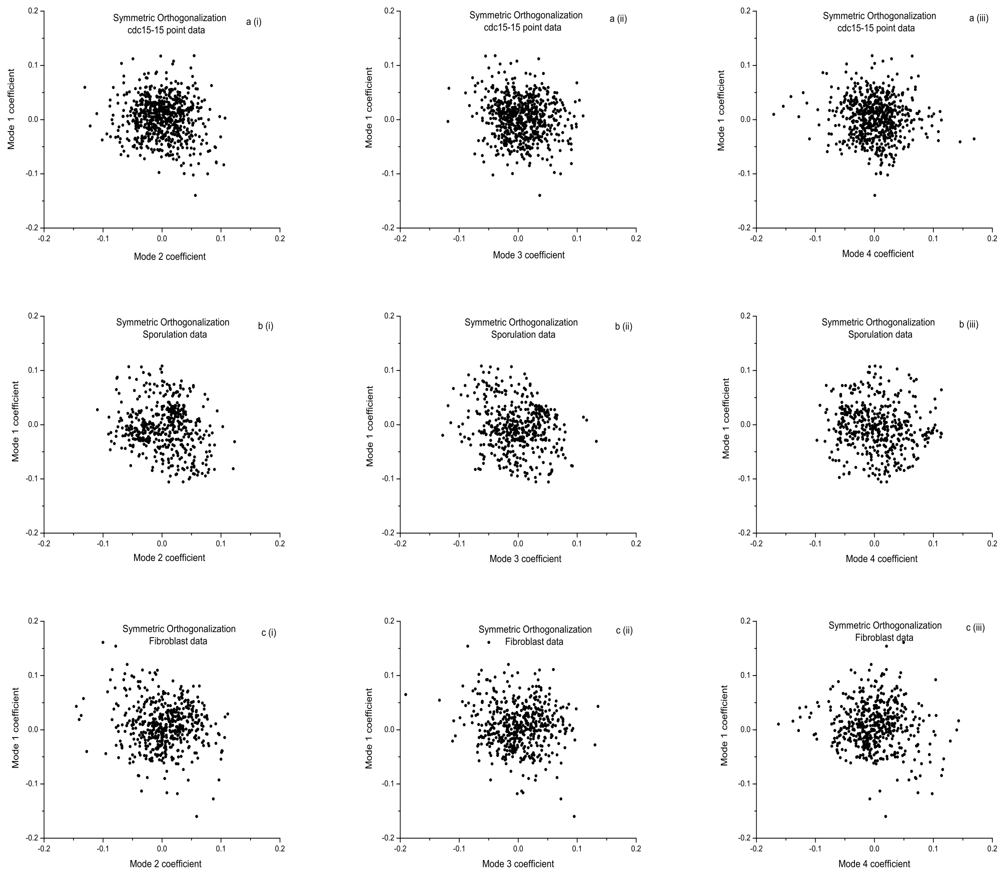


Figure 6.4: Plot of the coefficients for characteristic mode 1 against the coefficients for (i) characteristic mode 2, (ii) characteristic mode 3 and (iii) characteristic mode 4 respectively for (a) cdc15-15 point, (b) Sporulation and (c) Fibroblast data using symmetric orthogonalization.

6.6 Some Analytical Results

We have checked the correctness of the data generated by the implementation of canonical and symmetric orthogonalization schemes by testing the following mathematical identities.

For any gene $\boldsymbol{j},\ \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{n}$,

$$\sum_{i=1}^{m} \left(\boldsymbol{\Lambda}_{ji}, {d_i}^{1/2}\right)^2 = 1. \tag{6.13}$$

For any time $\boldsymbol{t},\ \boldsymbol{t} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{m}$,

$$\sum_{j=1}^{n} \left(\boldsymbol{\Lambda}_{ji}, {d_i}^{1/2}\right)^2 = d_{ii}. \tag{6.14}$$

For any gene $\boldsymbol{j},\ \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{n}$,

$$\sum_{i} \Big[\sum_{k} \left(\boldsymbol{\Phi}_{jk}, {M_{ki}}^{1/2}\right) \Big]^2 = 1. \tag{6.15}$$

For any time $\boldsymbol{t},\ \boldsymbol{t} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{m}$,

$$\sum_{j} \Big[\sum_{k} \left(\boldsymbol{\Phi}_{jk}, {M_{ki}}^{1/2}\right) \Big]^2 = M_{ii}. \tag{6.16}$$

Few other results which we have obtained are as follows.

$$\sum_{i} \Big[\sum_{k} \left(\boldsymbol{\Lambda}_{jk}, X_{ki}\right) \Big]^2 = 1. \tag{6.17}$$

$$\sum_{j} \Big[\sum_{k} \left(\boldsymbol{\Lambda}_{jk}, X_{ki}\right) \Big]^2 = M_{ii}. \tag{6.18}$$

$$\sum_{i} \Big[\sum_{k} \left(\boldsymbol{\Phi}_{jk}, X_{ik}\right) \Big]^2 = 1. \tag{6.19}$$

$$\sum_{j} \Big[\sum_{k} \left(\boldsymbol{\Phi}_{jk}, X_{ik}\right) \Big]^2 = d_{ii}. \tag{6.20}$$

Chapter 7

Neural Networks Using Löwdin Orthogonalizations

7.1 Introduction

The activity in neural networks gained considerable momentum after the associative memory model proposed by J. J. Hopfield [20, 21]. But this model has a serious constraint that there is an interference between the stored input patterns which turns *catastrophic* when the number of stored patterns becomes quite large. In the Hopfield model, in a system of $\boldsymbol{N}$ firing and quiescent neurons and $\boldsymbol{p}$ input patterns, the condition of associative memory fails when $\boldsymbol{p/N > 0.14}$, [22, 23, 24] i.e., this model gives a memory capacity of $\mathbf{0.14}\boldsymbol{N}$ [25]. The retrieval process of stored patterns breaks down and a memory blackout occurs when the number of patterns that come to be recorded exceeds $\mathbf{0.14}\boldsymbol{N}$. This is the *catastrophic interference* or *catastrophic forgetting*. It is also known as the *stability-plasticity* problem [26]. The cause for this *catastrophe* is the correlation among the patterns (or memories), which makes the system noisy as the number of stored memories increases until retrieval becomes minimal [22, 23, 24]. This catastrophic interference has been a serious constraint also in other connectionist models [27] of neural networks in which the learning of new information beyond a certain limit causes sudden and complete disappearance of previously stored information.

It was proposed by V. Srivastava and S. F. Edwards [2] that the process of orthogonalization could help to overcome the limitation of catastrophic forgetting on the memory capacity of the stored patterns. V. Srivastava et. al., [2, 33] have proposed models to understand how the brain discriminates and categorises different tasks. The mathematical computation of orthogonalization overcomes the noise among the learnt patterns and it circumvents the problem of catastrophic interference among the memories. The process of orthogonalization essentially compares a new object with those already in memory and it identifies their *similarities* and *differences* so that the new object can be placed in the right category. It is our understanding that the process of orthogonalization can overcome catastrophic interference but at the same time it can perform discrimination and categorization of different objects and tasks.

The Gram-Schmidt orthogonalization does this sequentially by taking patterns one by one and comparing them with the existing patterns in the memory. The sequential Gram-Schmidt orthogonalization process acts like decision making process wherein one object is compared with the stored ones and the nature of correlation with them decides, in an efficient and economical manner, in what form it should be stored. The noise is also eliminated. It is still unclear how the information in the brain is actually stored, processed and retrieved in biological neural networks.

In the process of understanding how the brain processes *similar* and *different* tasks simultaneously, a few researches have explored the finer aspects of distinction between the *same* and *different* kind of stimuli [28], how the brain isolates objects from their backgrounds [29], how rats compare scents to do a task efficiently [30, 31] and how the previous awareness affects the sensory learning [32]. But the problem in all these and many more cognitive functions is the mind's ability to *discriminate*. How it is carried out in the brain physiologically is not known. It is possible that the brain could perhaps do orthogonalization, or some thing similar to it, to discriminate and categorise the information.

There are several other processes of memory where one does not always handle the patterns sequentially. For such situations, Löwdin's non-sequential (or democratic) orthogonalization methods might play a role where one can handle all the input patterns at the same time. These *democratic* orthogonalization methods are incorporated in the Hopfield model to circumvent the problem of catastrophic interference in the retrieval of, say, episoidal memories.

In this chapter, we have numerically tested the condition of associative memory of Hopfield model for large number of synaptic patterns $\boldsymbol{p}$ with the number of firing and not-firing neurons $\boldsymbol{N}$. We have simulated the memory using Hopfield model by incorporating the sequential Gram-Schmidt and non-sequential Löwdin orthogonalization methods as the storage strategies for large data sets. We have numerically tested how the stored synaptic patterns can be retrieved.

7.2 The Hopfield Model

We first outline the Hopfield model briefly to elicit the mathematical representation of the brain functions involved in the process of learning, memory and recognition, etc. [22, 23, 24]. We have studied the Hopfield model in neural networks to understand how the brain stores, retrieves and processes the information and to understand the catastrophic interference. The Hopfield model is numerically tested for large data sets to understand the condition of memory catastrophe in the case of large data sets. In a system of N number of firing and not-firing neurons, a pattern is represented by an N-dimensional vector $\vec{\xi}$ whose N components ξ_i's are either +1 or -1, respectively. A pattern of +1 or -1 represents a stable state of the neural system, and hence a memory, if it minimises the Hamiltonian,

$$H = -\frac{1}{2}\,\Sigma_{i,j=1}^{N}\,W_{ij}\xi_i\xi_j, \tag{7.1}$$

where W_{ij} is the synaptic connection between the two neurons ξ_i and ξ_j.

The procedure involved in Hopfield model is outlined in the following algorithm.

Let p be the number of patterns to be memorized and let the memorized patterns be represented by $\vec{\xi}^{\mu}$, with $\mu = 1, 2, \cdots, p$. These patterns are randomly generated and each pattern is normalized. The weight matrix is formulated by using the Hebb's auto-association outer product rule,

$$W_{ij} = \frac{1}{N}\Sigma_{\mu=1}^{p}\,\xi_i^{\mu}\,\xi_j^{\mu},\; i, j = 1, 2, \cdots, N, \tag{7.2}$$

where the index μ represents the patterns. It gives the symmetric matrix $W_{ij} = W_{ji}$. Each memory is treated as independent, i.e. a new pattern (or memory) is added without reference to the previous ones.

As there is no self-feedback in a neuron, we set

$$\boldsymbol{W_{ii}} = \boldsymbol{0} \tag{7.3}$$

$$\text{So, that, } \boldsymbol{W_{ij}} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\boldsymbol{\Sigma}_{\mu=1}^{p}\, \boldsymbol{\xi_i^{\mu}\xi_j^{\mu}} - \frac{\boldsymbol{p}}{\boldsymbol{N}}\boldsymbol{I}, \tag{7.4}$$

where $\boldsymbol{I}$ is identity matrix.

Starting with the first pattern, the patterns are picked up one by one sequentially and weights are calculated according to (7.2) in a cumulative manner. The cumulative changes in the weights store the patterns or the memories. In order to test if the patterns are there in the memory, we present one of the $\boldsymbol{p}$ memories of learnt patterns, say $\boldsymbol{\mu} = \boldsymbol{\nu}$, to the brain for association. It produces local fields $\boldsymbol{h_i^{\nu}}$ on all the neurons. This can be expressed as

$$\boldsymbol{h_i^{\nu}} = \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{N}}}\boldsymbol{\Sigma}_{j=1}^{N}\left(\frac{\boldsymbol{1}}{\boldsymbol{N}}\boldsymbol{\Sigma}_{\mu=1}^{p}\, \boldsymbol{\xi_i^{\mu}}\ \boldsymbol{\xi_j^{\mu}}\right)\ \boldsymbol{\xi_j^{\nu}} \tag{7.5}$$

$$\text{or, } \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{N}}}\ \boldsymbol{h_i^{\nu}}\ \boldsymbol{\xi_i^{\nu}} = \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{N}}}\boldsymbol{\Sigma}_{j=1}^{N}\left(\frac{\boldsymbol{1}}{\boldsymbol{N}}\boldsymbol{\Sigma}_{\mu=1}^{p}\, \boldsymbol{\xi_i^{\mu}}\ \boldsymbol{\xi_j^{\mu}}\right)\ \boldsymbol{\xi_j^{\nu}}\ \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{N}}}\boldsymbol{\xi_i^{\nu}} \tag{7.6}$$

$$= \frac{\boldsymbol{1}}{\boldsymbol{N^2}}\boldsymbol{\Sigma}_{j=1}^{N}\left[\boldsymbol{\xi_j^1}\ \boldsymbol{\xi_i^1} + \boldsymbol{\xi_j^2}\ \boldsymbol{\xi_i^2} + \cdots + \boldsymbol{\xi_j^p}\ \boldsymbol{\xi_i^p}\right]\ \boldsymbol{\xi_j^{\nu}}\ \boldsymbol{\xi_i^{\nu}} \tag{7.7}$$

The success of retrieval is measured by the sign of $\boldsymbol{h_i^{\nu}}$ on all the sites $\boldsymbol{i}$ through the requirement $\frac{\boldsymbol{1}}{\sqrt{\boldsymbol{N}}}\ \boldsymbol{h_i^{\nu}}\ \boldsymbol{\xi_i^{\nu}} > \boldsymbol{0}$. If the sign of $\boldsymbol{h_i^{\nu}}$ turns out to be the same as that of $\boldsymbol{\xi_i^{\nu}}$ for almost all $\boldsymbol{i}$'s, then the presented $\boldsymbol{\xi^{\nu}}$ is supposed to associate well with the $\boldsymbol{\nu^{th}}$ pattern. Hence the condition for good retrieval of the memorised patterns can be written as follows.

$$\boldsymbol{sgn(h_i)} = \boldsymbol{sgn(\xi_i)}. \tag{7.8}$$

If this condition is satisfied on 97% or more of neurons, then it is considered to be a good retrieval.

We have numerically tested a network consisting of 100 to 1000 neurons storing $\boldsymbol{p}$ patterns for an overlap $(\vec{\boldsymbol{h}}^{\boldsymbol{\nu}}.\ \vec{\boldsymbol{\xi}}^{\boldsymbol{\nu}})$ of 97%, 98% and 99% or more. We have drawn the graphs for these three cases between the fraction of the presented pattern that are retrieved and $\boldsymbol{p/N}$, the number of presented patterns normalized by the number of neurons, $\boldsymbol{N}$, for 500 to 1000 neurons considering the overlap between the presented and retrieved patterns to be 97%, 98% and 99% and is shown in Figure 7.1. The results show that the memory capacity of the Hopfield model is limited to $\boldsymbol{p/N \sim 0.14}$. The fraction of retrieved patterns begins to drop as $\boldsymbol{p/N}$approaches $\mathbf{0.12}$ and then drops rapidly to *zero* around $\boldsymbol{p/N \simeq 0.18}$. The rate of drop does not appear to depend on the 'overlap' parameter although one would expect it to be steeper as overlap increases from 97% to 99%.

The above constraints in the model is known in the literature as *memory catastrophe* which is quite serious if we wish to have a large number of stored patterns. The reason for this is that there is an interference between the stored input patterns because they are correlated, i.e., their vectors have non-zero dot products. The correlation among the memories makes the system noisy as the number of stored patterns increases. As the stored patterns increase in number, at some point the storage capacity will be exhausted, and new patterns will interfere with old patterns and they will prevent the retrieval of stored patterns.

The problem of catastrophic interference among the memories could be overcome with the help of orthogonalization without having to resort to sparse coding as is sometimes suggested in the literature. The process of orthogonalization enables the network to compare a new object with those already in memory, and to identify their similarities and differences so that the new object can be placed in the right memory.

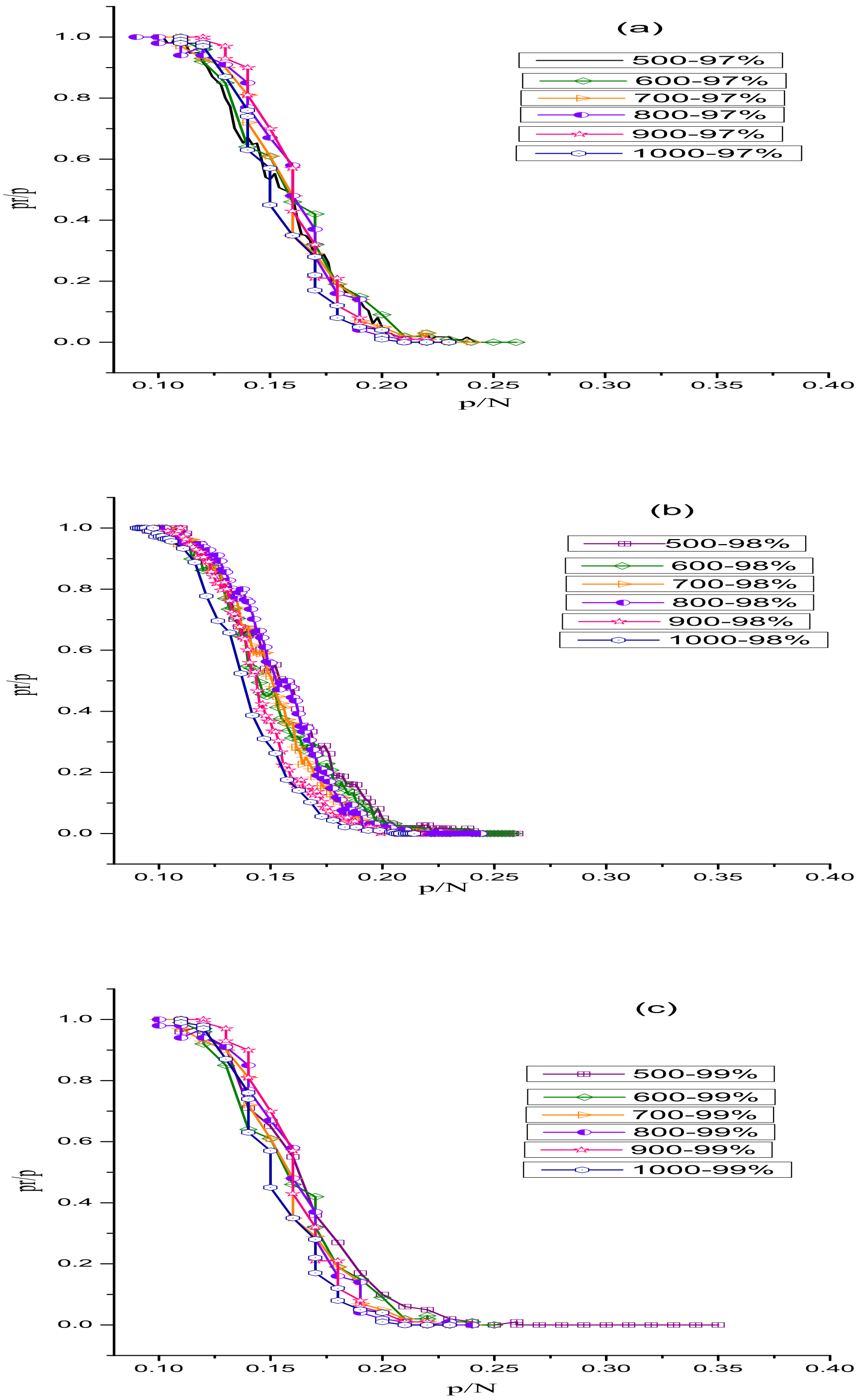


Figure 7.1: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no.of. neurons) for retrieval percentages (a) 97%, (b) 98% and (c) 99%.

7.3 Hopfield Model with Gram-Schmidt Orthogonalization

The Hopfield model is tested by incorporating the Gram-Schmidt orthogonalization process before the calculation of weight matrix $\boldsymbol{W_{ij}}$. The use of Gram-Schmidt orthogonalization was proposed by V. Srivastava and S. F. Edwards to eliminate the noise and to achieve high memory capacity but at the same time to handle the correlated memories. The orthonormalized patterns obtained through the Gram-Schmidt method are used to calculate the weight matrix $\boldsymbol{W_{ij}}$. The local field $\boldsymbol{h_i}$ is calculated by presenting the normalized raw patterns and the orthonormalized ones. Then the normalized raw patterns and orthonormalized patterns are presented to the weight matrix constructed by using the orthonormalized patterns. It is observed that the number of retrieved patterns is increased when the normalized raw patterns are presented to the weight matrix calculated using the orthonormalized patterns than in the usual Hopfield model. When the orthonormalized patterns are presented to the same weight matrix, all the patterns are completely retrieved, i.e., there is 100% retrieval of the presented patterns.

The Gram-Schmidt orthogonalization is incorporated in the Hopfield model as follows.

Let there be $\boldsymbol{p}$ patterns to be memorized in a system of $\boldsymbol{N}$ neurons. Let the memorized patterns be represented by $\vec{\boldsymbol{\xi}}^{\mu}, \boldsymbol{\mu} = \mathbf{1}, \mathbf{2}, \cdots, \boldsymbol{p}$. These patterns are randomly generated and each pattern is normalized.

The Gram-Schmidt orthogonalization process is used to calculate the orthonormal patterns. That is the process of exploring the store on the arrival of a new information, say $\vec{\xi}^n$, is represented mathematically as modifying the raw $\vec{\xi}^n$ so that it becomes orthogonal to the existing patterns. This process is defined through

$$\eta_k^n = \xi_k^n - \Sigma_{q=1}^{n-1} \eta_k^q \frac{\Sigma_{i=1}^N \eta_i^q \xi_i^n}{\Sigma_{i=1}^N \eta_i^q \eta_i^q}. \tag{7.9}$$

This amounts to exacting details from the n^{th} as well as the earlier patterns. The brain stores these $\vec{\eta}^n$'s rather than the raw $\vec{\xi}^n$'s as in the case of Hopfield model. The second term on the right-hand side represents the sum of projections of $\vec{\xi}^n$'s on all $\vec{\eta}^q$'s with $q < n$. Thus $\vec{\eta}^n$ is obtained after subtracting out from $\vec{\xi}^n$ its commonalities with all the earlier $\vec{\eta}^q$'s. The normalization $\vec{\eta}^\mu . \vec{\eta}^\mu = N$ remains for all μ. Note that

$$\Sigma_i \eta_i^n \eta_i^q = 0, \text{for all } q < n. \tag{7.10}$$

The weight matrix is obtained as before using $\vec{\eta}^n$'s.

$$W_{ij} = \frac{1}{N} \Sigma_{\mu=1}^p \ \eta_i^\mu \ \eta_j^\mu, i, j = 1, 2, \cdots, N, \tag{7.11}$$

where the index μ represents the patterns. It gives the symmetric matrix $W_{ij} = W_{ji}$. Each memory is independent, i.e., each time a new pattern (or memory) is added without reference to the previous ones.

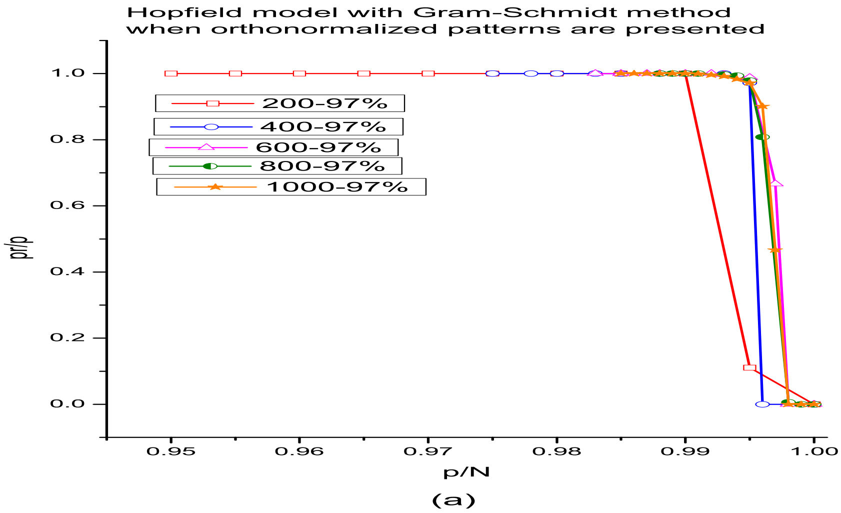
Again for $W_{ii} = 0$, we have the memory matrix

$$W_{ij} \quad = \quad \frac{1}{N} \Sigma_{\mu=1}^p \ \eta_i^\mu \eta_j^\mu - \frac{p}{N} I, \tag{7.12}$$

where I is the identity matrix.

If we now present $\vec{\eta}^\mu$'s for retrieval all learnt $\vec{\eta}^\mu$'s are retrieved perfectly as long as $p \leq N$. Thus the memory capacity is increased from $p/N = 0.14$ to $p/N = 1$. Interestingly the raw patterns $\vec{\xi}^\mu$'s are also retrieved for $p \leq N$. Figure 7.2 shows perfect retrieval of orthogonalized as well as raw patterns after

Gram-Schmidt scheme is invoked. In an $\boldsymbol{N}$-dimensional space there can be $\boldsymbol{N}$ orthogonal vectors. This restricts the capacity to $\boldsymbol{p/N = 1}$.



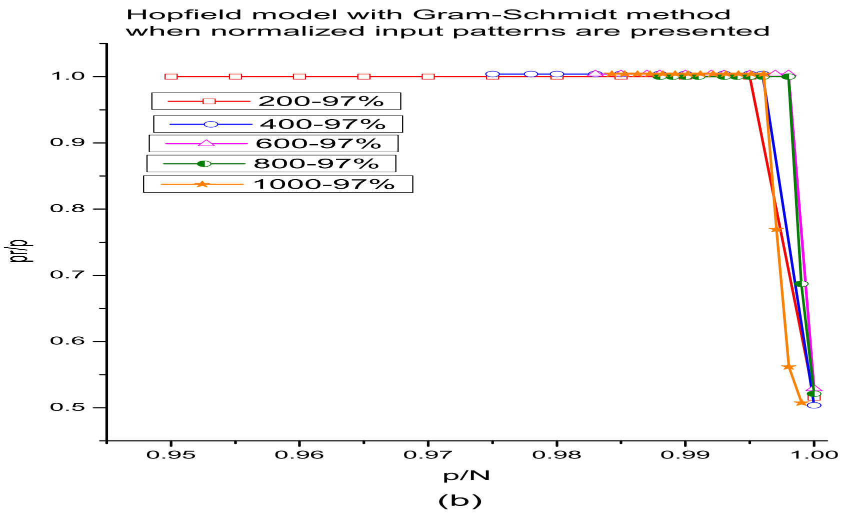


Figure 7.2: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no.of. neurons) for Gram-Schmidt method.

7.4 Hopfield Model using Löwdin's Symmetric and Canonical Orthogonalizations

We have tested the Hopfield model by incorporating the Löwdin's symmetric and canonical orthogonalization methods as the storage strategies. As with Gram-Schmidt we expect the patterns orthogonalized by these two schemes to be retrieved perfectly since they will also eradicate the noise in the same manner as Gram-Schmidt does. However we expect the orthogonalized patterns to possess different characters than those orthogonalized with Gram-Schmidt scheme. since the underlying schemes are democratic in nature and handle the given set of random patterns altogether yet quite differently, we expect that the fine differences between canonical and symmetrical methods and their differences with Gram-Schmidt will be useful in their applications to cognitive phenomena. We randomly generate the raw patterns of firing/not-firing neurons and normalize them. We then calculate the symmetric and canonical orthonormal bases and form the weight matrices using them. Then, we presented the normalized raw patterns to the symmetric and canonical weight matrices and calculated their local field to study retrieval.

7.4.1 Hopfield model with Symmetric Orthogonalization

The Löwdin's symmetric orthogonalization method can be incorporated in the Hopfield model as follows.

Let $\boldsymbol{p}$ patterns $\{\boldsymbol{\xi^\mu}\}, \boldsymbol{\mu = 1, 2, \cdots, p}$ in $\boldsymbol{N}$ dimensions be written as matrix Ξ, whose columns represent the input patterns. The Gram matrix $\boldsymbol{M}$ is constructed as $\boldsymbol{M} = \Xi\ \Xi^{\boldsymbol{T}}$ and its eigenvalues $\boldsymbol{d}$ and normalized eigenvectors $\boldsymbol{U}$ are calculated.

The symmetric orthonormal bases are obtained using

$$\boldsymbol{\Phi} = \boldsymbol{\Xi}\ \boldsymbol{M}^{-1/2}. \tag{7.13}$$

The brain stores these $\vec{\boldsymbol{\phi}}^{\boldsymbol{n}}$'s rather than the raw $\vec{\boldsymbol{\xi}}_{\boldsymbol{i}}$'s. The weight matrix is now

$$\boldsymbol{W_{ij}} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\boldsymbol{\Sigma}^{\boldsymbol{p}}_{\boldsymbol{\mu=1}}\ \boldsymbol{\phi^{\mu}_{i}\phi^{\mu}_{j}} - \frac{\boldsymbol{p}}{\boldsymbol{N}}\boldsymbol{I}, \tag{7.14}$$

where $\boldsymbol{I}$ is the identity matrix.

7.4.2 Hopfield model with Canonical Orthogonalization

The Löwdin's canonical orthogonalization method is incorporated in the Hopfield model in the same fashion as above. The canonical orthonormal bases are obtained using

$$\boldsymbol{\Lambda} = \boldsymbol{\Xi}\ \boldsymbol{U}\ \boldsymbol{d}^{-1/2}. \tag{7.15}$$

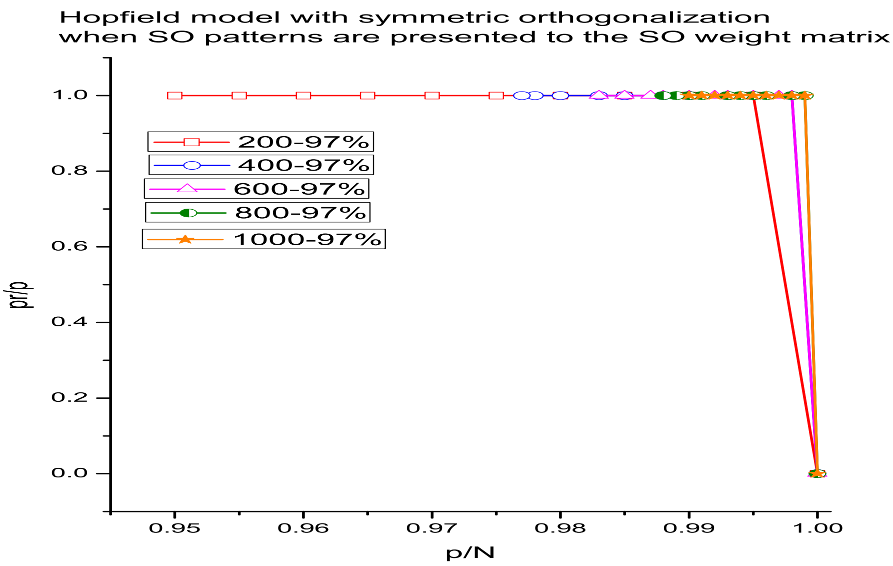
and the weight matrix is written as

$$\boldsymbol{W_{ij}} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\boldsymbol{\Sigma}^{\boldsymbol{p}}_{\boldsymbol{\mu=1}}\ \boldsymbol{\lambda^{\mu}_{i}\lambda^{\mu}_{j}} - \frac{\boldsymbol{p}}{\boldsymbol{N}}\boldsymbol{I}, \tag{7.16}$$

where $\boldsymbol{I}$ is the identity matrix.

Quite expectedly, the memory capacity in both 'symmetric' and 'canonical' orthogonalizations is $\boldsymbol{p/N = 1}$ as shown in Figures 7.3 and 7.4.

The major difference between the Gram-Schmidt and the Löwdin methods is that in the latter as one new pattern is added after a set of $\boldsymbol{q}$ patterns is orthogonalized, the entire lot of $\boldsymbol{q+1}$ patterns is orthogonalized all over again. Thus the sets of $\boldsymbol{q}$ and $\boldsymbol{q+1}$ orthogonalized patterns differ from each other. The differences between the sets also depend on whether we employ canonical or symmetric orthogonalization. We have discussed these changes in the following section.



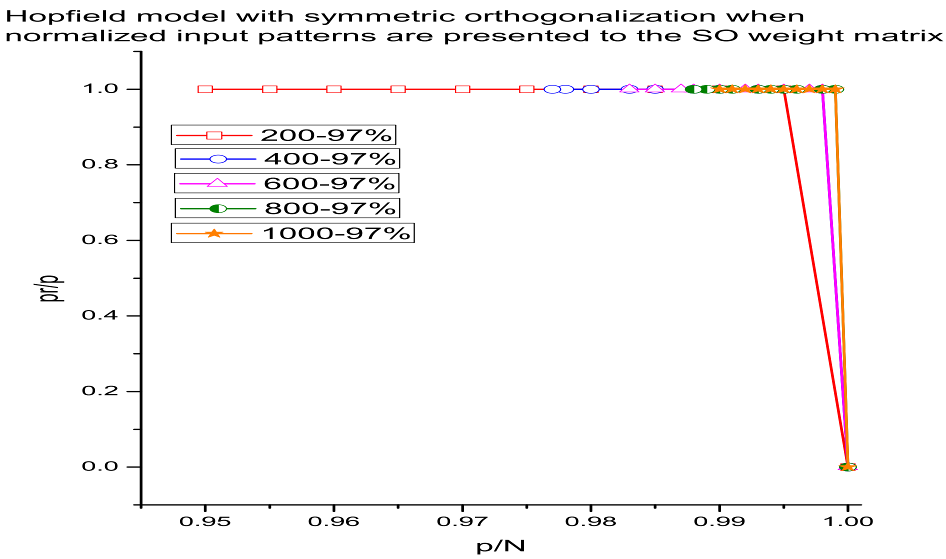
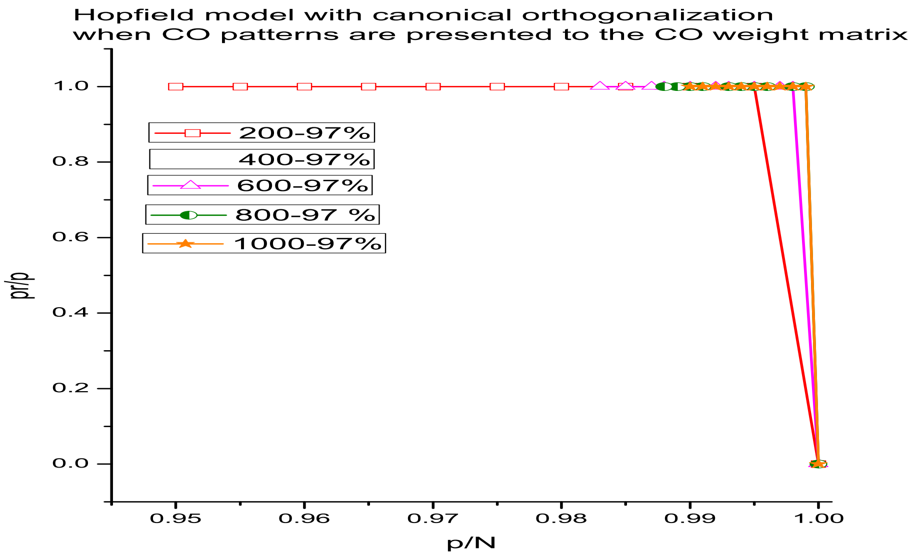


Figure 7.3: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no.of. neurons) for symmetric method.



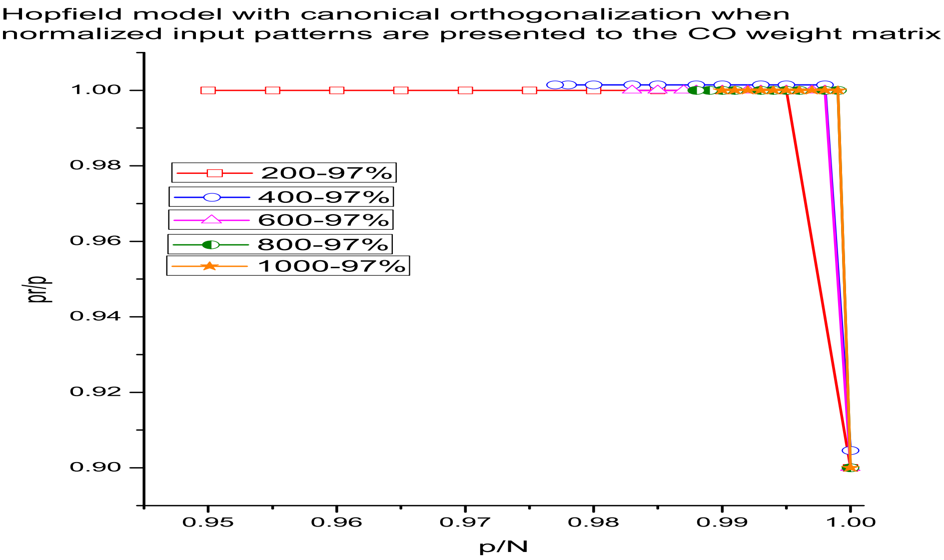


Figure 7.4: Plots of (no. of retrieved patterns)/(no. of presented patterns) versus (no. of presented patterns)/(no.of. neurons) for canonical method.

7.5 Computations, Results and Discussion

We have numerically tested the Hopfield model by presenting patterns of firing/not-firing neurons starting with the smaller data sets. We have applied the Löwdin's symmetric and canonical orthogonalizations for the case of **5** randomly generated normalized input patterns of **10** firing or not-firing neurons. When a new pattern is added each time without disturbing the previously stored patterns, it changes all the weights and triggers electric activity among neurons and generates local field on each of them for symmetric and canonical orthogonalizations. The addition of new pattern will change all the previously stored **5** patterns in both symmetric and canonical schemes, and generate new ones. The randomly generated **5** patterns and their corresponding symmetric and canonical patterns are shown in Figure 7.5. Figure 7.6 shows the results of symmetric and canonical orthogonalizations when the new pattern is added.

We have added the $\mathbf{6^{th}}$ pattern, with 70% dissimilarity to the $\mathbf{5^{th}}$ pattern and all the six patterns are orthogonalized using the symmetric and canonical orthogonalizations. We have projected the normalized input patterns onto the symmetric and canonical orthonormal patterns to check how they have changed in a gross sense. In the case of symmetric orthogonalization, the sum of squared projections of any of the normalized input patterns onto any of the symmetric orthonormal patterns is equal to *unity*. When all the normalized input patterns are projected individually onto the canonical patterns picked up one by one, then the sum of squared projections is found to be the largest on the canonical pattern which corresponds to the largest eigenvalue.

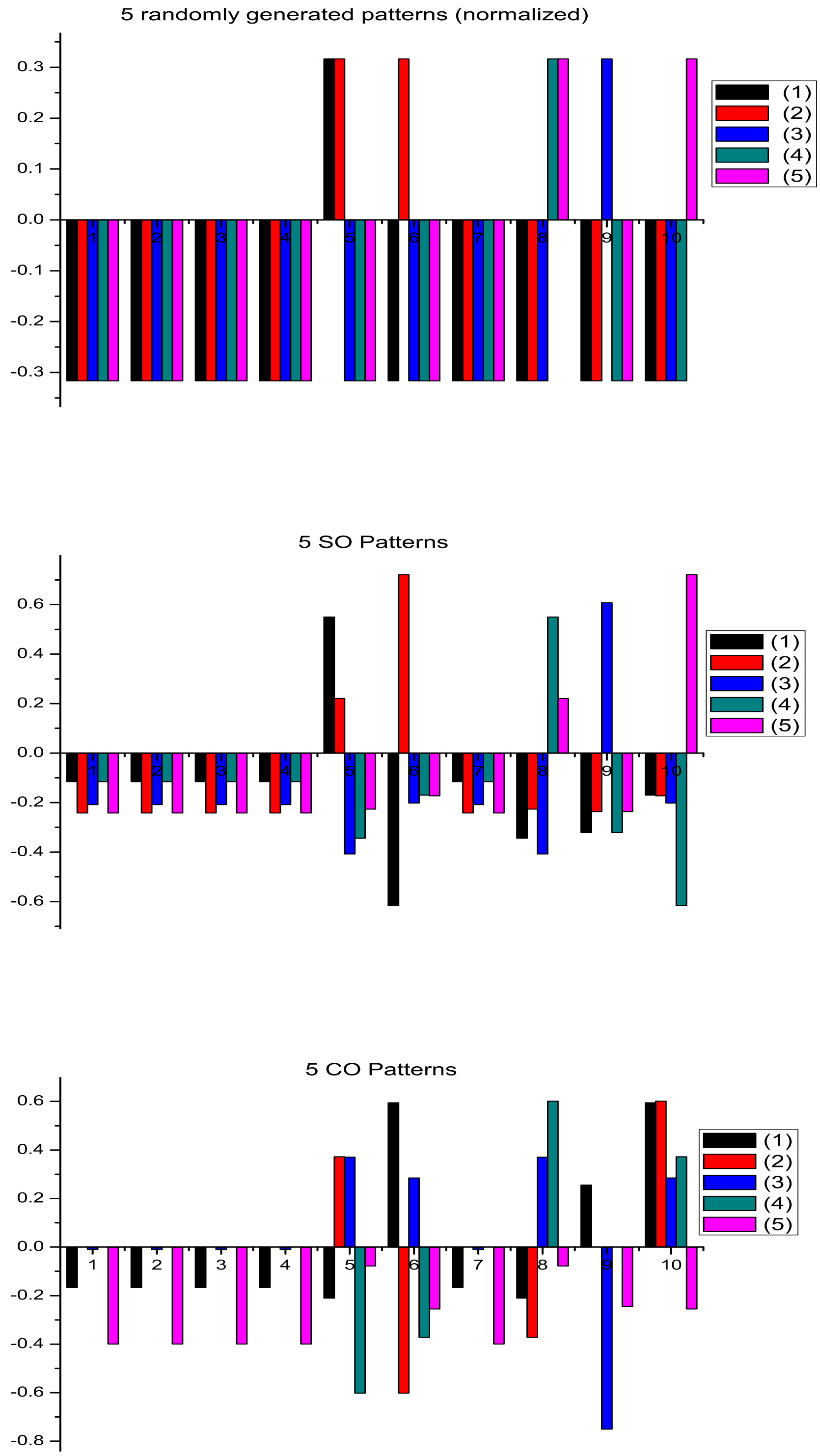


Figure 7.5: **5** randomly generated patterns and their corresponding symmetrically and canonically orthogonalized patterns.

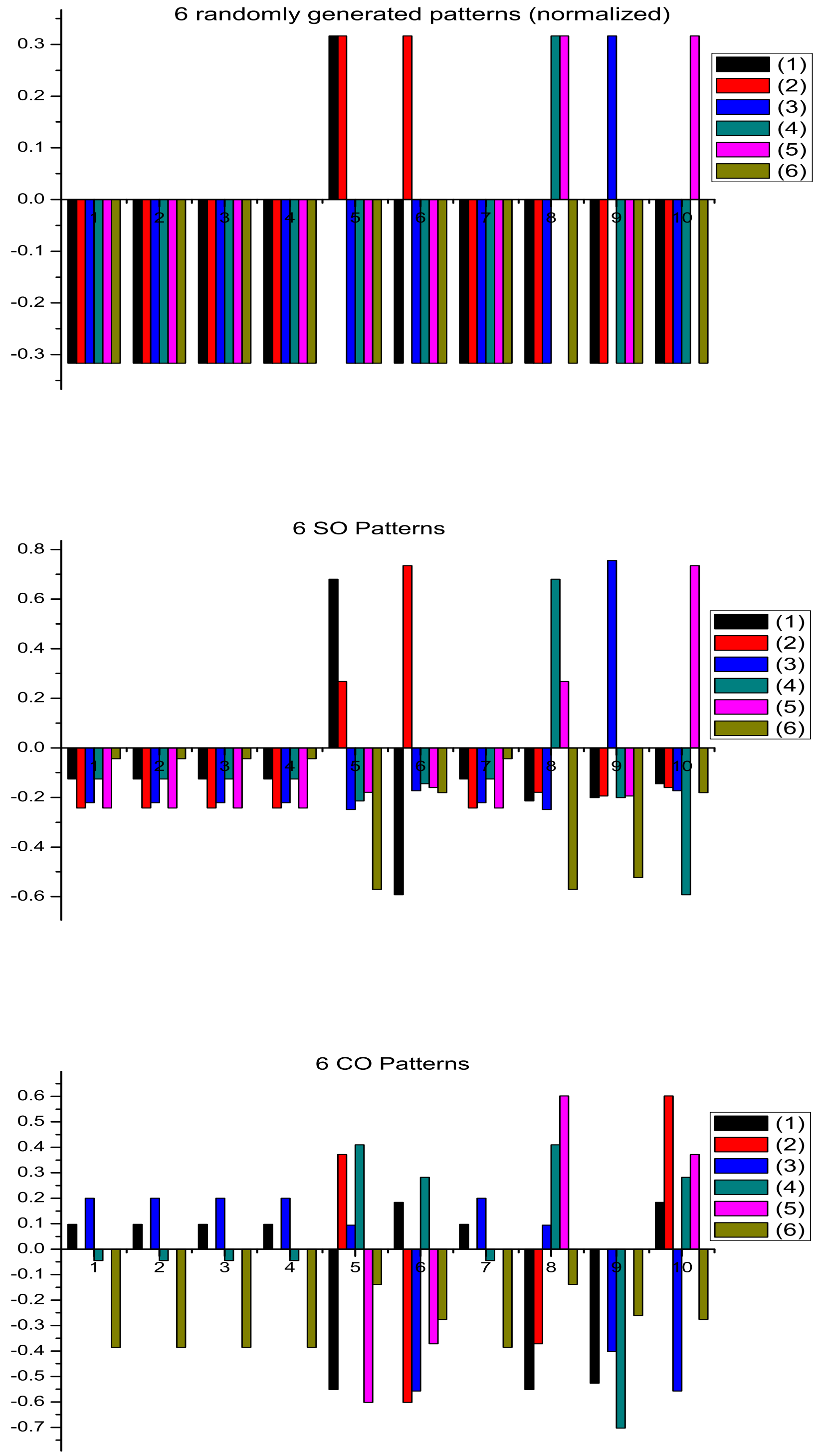


Figure 7.6: **6** randomly generated patterns and their corresponding symmetrically and canonically orthogonalized patterns.

Starting from **5** patterns we have added one by one **5** new patterns. Quite surprisingly it turns out that the $\mathbf{2}^{\boldsymbol{nd}}$ canonical pattern carries the largest eigenvalue and therefore captures the maximum projection of all the raw patterns. As of now we can not offer an explanation as to why the principal component does not change as new patterns are added but further studies are on. Figure 7.7 and 7.8 show how each of **5** patterns changes as the sixth pattern is added and they are subjected to canonical and symmetric orthogonalizations respectively. The changes are generally vigorous in the case of canonical orthogonalization (even signs change) whereas in the case of symmetric orthogonalization the changes caused by adding the $\mathbf{6}^{\boldsymbol{th}}$ pattern are moderate (the sign is not changed in any of the examples).

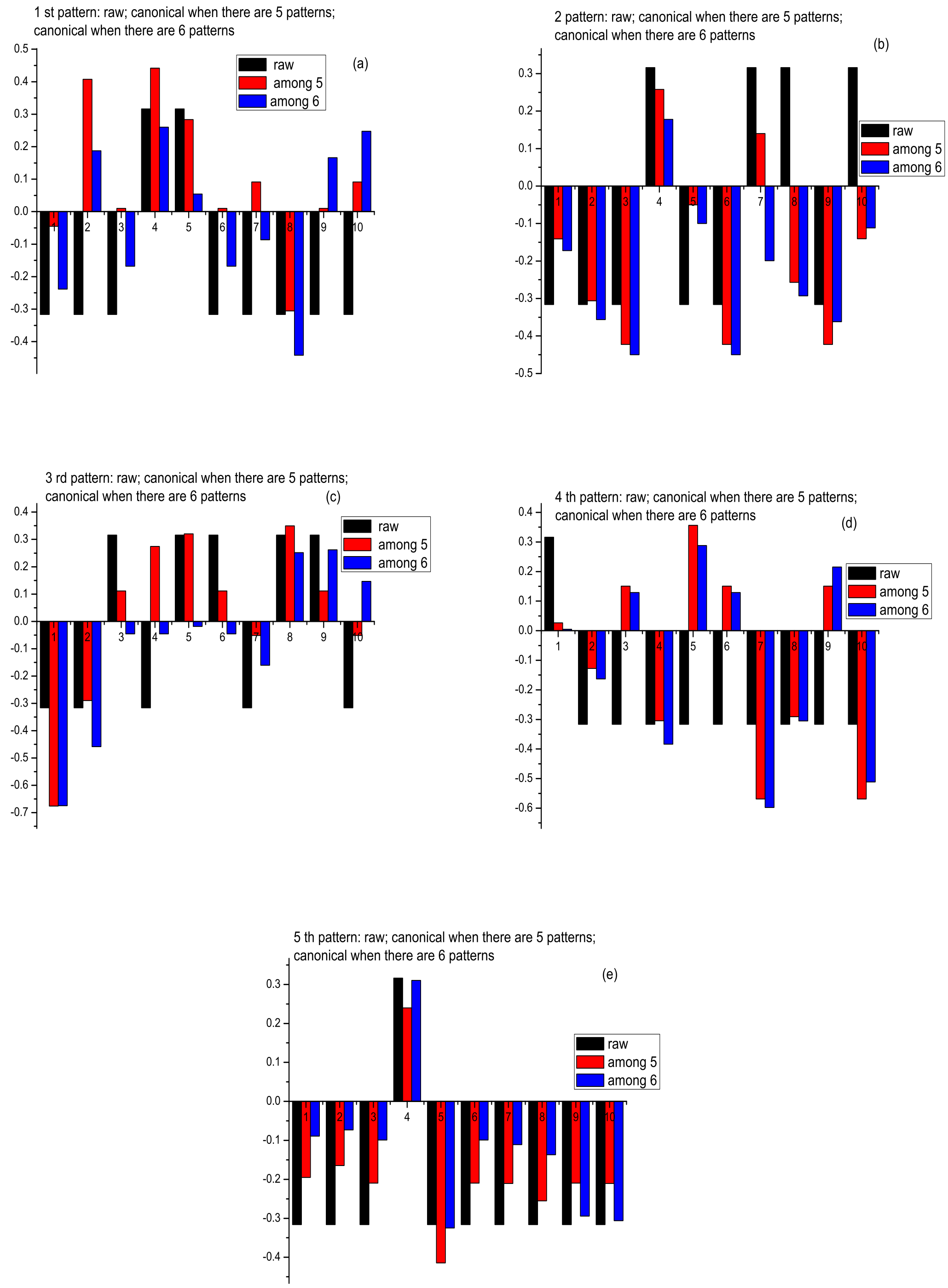


Figure 7.7: Five raw patterns (black) together with their canonically orthogonalized patterns (red) and what they become (blue) when a $\mathbf{6}^{th}$ new pattern is added and all **6** are canonically orthogonalized.

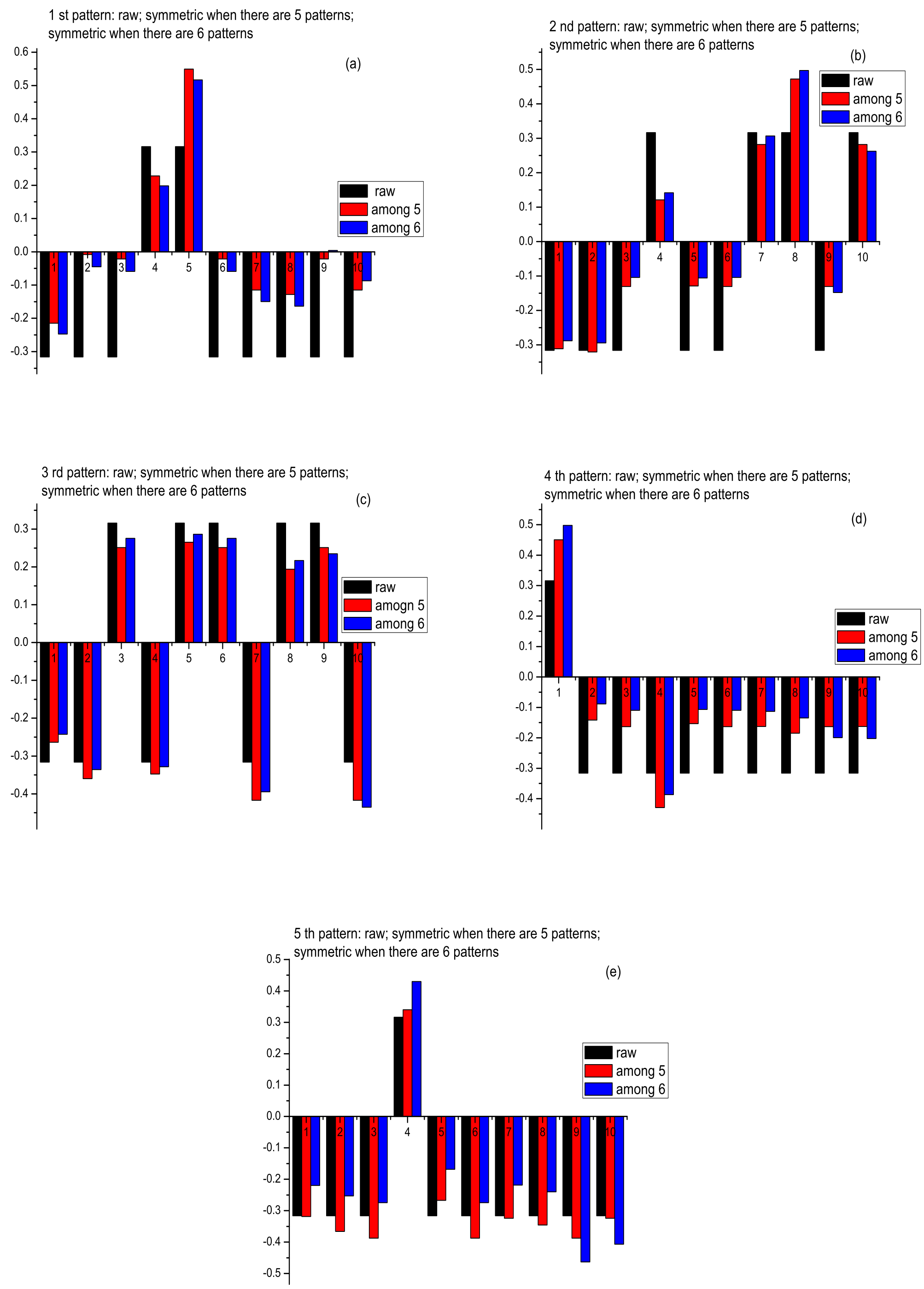


Figure 7.8: Five raw patterns (black) together with their symmetrically orthogonalized patterns (red) and what they become (blue) when a $\mathbf{6}^{th}$ new pattern is added and all **6** are symmetrically orthogonalized.

Chapter 8

Conclusion and Outlook

We have done an extensive study of orthogonalization procedures to generate orthonormal basis sets from a given set of linearly independent vectors. This was done with the aim of applying these procedures to a variety of problems in areas ranging from mathematics to biology.

The orthogonalization procedures come in two categories, *sequential* and *democratic* depending on whether the given vectors have to be orthogonalized in a one-after-the-other manner, or by considering the entire lot in one go. We concentrated mainly on the latter type because their applications have not been explored much except in the domain of quantum chemistry.

Fascinated and motivated by the remarkable geometrical properties that the democratic orthogonalization schemes are known to possess, we planned to study their relevance and implications in a variety of contexts. Our discussions are built around the so called *canonical* and *symmetric* orthogonalization schemes due to Löwdin [1]. However, we have found that the canonical orthogonalization was in fact invented independently by several people from time to time with different names. We have established their equivalence. In the process we also worked out ways of using the orthogonalization schemes for arbitrarily large number of linearly dependent vectors.

We derived two distinct classes of orthogonal polynomials using canonical and symmetric schemes of Löwdin. These are unique, in that they possess very special properties and are looking for applications.

A significant application of the two orthogonalization schemes was done to organise and interpret huge sets of data on gene-expression profiles. The use of canonical orthogonalization quite predictably yielded results identical to those obtained earlier using SVD though we got some new insights due to the special geometrical property of canonical orthogonalization underlined by us. The results using symmetric orthogonalization scheme are however novel. But there are many loose ends to be tied, which require more work in collaboration with biologists.

Our investigation of the famous neural networks model due to Hopfield by incorporating the two democratic orthogonalization schemes has paved the way for studying cognitive learning and memory under special circumstances. The comparison with the earlier work on Hopfield network with Gram-Schmidt orthogonalization incorporated shows significant departure in the results. In the broad terms this shows that the brain might handle sequential learning through Gram-Schmidt, and the storage of information thus acquired, very differently from the way it might process the information acquired in bunches (i.e. simultaneous intake of a set of disparate information), for example in episodic memories. A lot of work needs to be done and can be done in close association with neuroscientists to unravel some mysteries of how the brain functions.

In sum we have done an in-depth analysis of almost all orthogonalization procedures in literature. We have demonstrated in this thesis that these orthogonalization techniques, which were originally invented for applications in mathematics and physical problems, can lead to invaluable insight into biological phenomena as well. Our work shows some new directions for further research in the areas of theoretical biology.

Appendix A

Vector Spaces and Matrix Algebra

A.1 Introduction

In this appendix we have briefly summarized the preliminaries of vector spaces and matrix algebra that are used in different contexts in our work with the help of certain text books [6, 14, 15]. The concepts of vector spaces, linear vector spaces, linear dependence, linear independence, basis, inner products, orthogonality, orthonormality, finite dimensional vector spaces and their properties, linear transformations, matrix algebra and few orthogonalization methods are discussed concisely. These basic concepts are recapitulated in order to have uniform notations and to make the thesis somewhat self contained. We have given the properties of some special matrices which appear where ever necessary.

A.2 Vector Spaces

The space formed by a given set of vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$ is known as a *vector space*, denoted by $\mathbf{V}$. If the vectors in a vector space are of degree *unity* then the vector space is called a *linear vector space.* The number of elements in a vector is called its *dimension* or *order* of a vector. The number of vectors in a vector space is called its *dimension* or *order* of a vector space. The vector space formed by a given set of vectors may be *real* or *complex* depending upon whether the vectors are *real* or *complex*, denoted by $\mathbf{V}^{\mathbb{R}}$ and $\mathbf{V}^{\mathbb{C}}$ respectively.

A.3 Linear Dependence and Independence

Let $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$ be a given set of vectors which forms a vector space $\mathbf{V}$. If one vector can be expressed in terms of any other vectors, then the vectors are said to be *linearly dependent* on each other.

For linearly dependent vectors,

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}, \tag{A.1}$$

such that $c_1 \neq c_2 \neq \cdots \neq c_n \neq 0$, where $c_1, c_2, \cdots, c_n$ are constants. That is, at least one of the constants should not be equal to zero.

In the case of linear dependence, the number of vectors has no limit and it does not depend on the number of dimensions of the space. In 3-D space, we have any number of vectors which satisfy the linear dependence condition.

In a vector space, if one vector cannot be expressed in terms of any other vectors then the vectors are said to be *linearly independent* of each other.

For linearly independent vectors,

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}, \tag{A.2}$$

such that $c_1 = c_2 = \cdots = c_n = 0$.

The maximum number of linearly independent vectors in a space cannot be more than that of the number of dimensions of the space. In a 3-D space, we can not have more than three independent vectors.

A.4 Basis for a Space

Let $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$ be n linearly independent vectors in an n-dimensional space. The set of n linear independent vectors is known as a *basis* for the given space. Geometrically, it is a set of coordinate axes.

$$Basis = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}.$$

Now, any other vector in the space can be expressed in terms of the basis vectors for the space. For a given space, the basis is not unique. It can be selected in an infinite number of ways.

A.5 Inner Product

Let $\vec{v}_1 = (v_{11}\ v_{12}\ \cdots, v_{1n})$ and $\vec{v}_2 = (v_{21}\ v_{22}\ \cdots, v_{2n})$ be two geometric vectors in $\mathbb{R}^n$ space. The inner product of $\vec{v}_1,\ \vec{v}_2$ and norm of a vector in $\mathbb{R}^n$ are defined as

$$< \vec{v}_1\ \vec{v}_2 > \; = \; \vec{v}_1.\vec{v}_2 = (v_{11}\ v_{21} + v_{12}\ v_{22} + \cdots + v_{1n}\ v_{2n}), \tag{A.3}$$

$$\| \vec{v}_1 \| \; = \; \sqrt{< \vec{v}_1\ \vec{v}_1 >} = \sqrt{{v_{11}}^2 + {v_{12}}^2 + \cdots + {v_{1n}}^2}. \tag{A.4}$$

A.6 Orthogonalization

The mathematical transformation whereby a set of vectors is converted into a set of mutually orthogonal vectors is called as *Orthogonalization*.

The two vectors $\vec{v}_1$ and $\vec{v}_2$ are said to be *orthogonal* if and only if their inner product is zero, i.e.

$$\vec{v}_1.\vec{v}_2 = (\vec{v}_1, \vec{v}_2) =< \vec{v}_1 \ \vec{v}_2 >= 0. \tag{A.5}$$

A.7 Orthonormalization

The two vectors $\vec{v}_1$ and $\vec{v}_2$ are said to be orthonormal if and only if they are *orthogonal* and *normalized to unity.* The orthonormalized vectors are orthogonal with their length equal to unity.

Mathematically, the orthonormalized vector can be expressed as:

$$\vec{v}_1.\vec{v}_2 = (\vec{v}_1, \vec{v}_2) =< \vec{v}_1 \ \vec{v}_2 >= 0 \text{ and } \vec{v}_1 = \frac{\vec{v}_1}{\| \vec{v}_1 \|}. \tag{A.6}$$

A.8 Finite Dimensional Vector Spaces

A.8.1 Construction of Linear Vector Spaces

For $\mathbf{V}$ to be a linear vector space in $\mathbb{R}^n$, it has to satisfy the following properties.

1. Vector Addition: $\forall \ \vec{v}_1, \vec{v}_2 \in \mathbb{R}^n : \vec{v}_1 + \vec{v}_2 = (v_{11}v_{21} + v_{12}v_{22} + \cdots + v_{1n}v_{2n})^T$.

2. Scalar Multiplication: $\forall \ \vec{v}_1 \in \mathbb{R}^n, \alpha \in \mathbb{R} : \alpha \ \vec{v}_1 = (\alpha v_{11}, \alpha v_{12}, \cdots, \alpha v_{1n})^T$.

3. Null Vector: $\vec{0} = (0, 0, \cdots, 0)^T \in \mathbb{R}^n$.

4. Norm: $\forall \ \vec{v}_1 \in \mathbb{R}^n : \|\vec{v}_1\| = \sqrt{v_{11}^2 + v_{12}^2 + \cdots + v_{1n}^2}$.

5. Inner Product: $\forall\, \vec{v}_1, \vec{v}_2 \in \mathbb{R}^n :< \vec{v}_1\ \vec{v}_2 >= \ (v_{11}v_{21}+v_{12}v_{22}+\cdots+v_{1n}v_{2n}).$

(a) $< \vec{v}_1\ \vec{v}_1 > = \parallel \vec{v}_1 \parallel^2 \geq 0.$ (b) $< \vec{v}_1\ \vec{v}_2 > = < \vec{v}_2\ \vec{v}_1 > .$

(c) $< \vec{v}_1\ \alpha\vec{v}_2 + \beta\vec{v}_3 > = \ \alpha < \vec{v}_1\ \vec{v}_2 > +\beta < \vec{v}_1\ \vec{v}_3 > .$

6. Dimension: There exists a basis of vectors $\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ such that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0} \in \mathbb{R}^n, c_1 = c_2 = \cdots = c_n = 0.$$

7. Orthonormal Basis: For an orthonormal basis $\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$,

$$\begin{aligned} < \vec{e}_i\ \vec{e}_j > = \ \delta_{i,j} &= 0,\ i \neq j, \\ &= 1,\ i = j. \end{aligned}$$

A.8.2 Properties of Projections in Orthonormal Basis

$\forall\, \vec{v}_1 \in \mathbb{R}^n : \vec{v}_1 = v_{11}\vec{e}_1 + \cdots + v_{1n}\vec{e}_n,$ where $< \vec{e}_i\ \vec{e}_j > = \ \delta_{i,j}, v_{i,j} = (\vec{e}_j, \vec{v}_1).$

1. Invariance of inner products: $< \vec{v}_1\ \vec{v}_2 > = \ (v_{11}\ v_{21}+v_{12}\ v_{22}+\cdots+v_{1n}\ v_{2n}).$

2. Parseval's equality: $\parallel \vec{v}_1 \parallel^2 =< \vec{v}_1\ \vec{v}_1 > = \ v_{11}^2 + v_{12}^2 + \cdots + v_{1n}^2.$

3. Bessel's inequality: $\forall\, m \leq n : v_{11}^2 + v_{12}^2 + \cdots + v_{1n}^2 \leq \vec{v}_1\ \vec{v}_1 > .$

4. Schwarz's inequality: $< \vec{v}_1\ \vec{v}_2 >^2 \leq < \vec{v}_1\ \vec{v}_1 > < \vec{v}_2\ \vec{v}_2 > .$

5. Traingle inequality: $\parallel \vec{v}_1 + \vec{v}_2 \parallel \leq \parallel \vec{v}_1 \parallel + \parallel \vec{v}_2 \parallel .$

A.9 Linear Transformation

The totality of all vectors of order n is known as an n-dimensional vector space. The mathematical operation of matrix on a vector is another vector, which changes its direction and length. Hence the result of operating on a vector space with a matrix is to map the space on to itself. Such a mapping is known as a *linear transformation*. Hence a linear transformation is a function between two vector spaces that preserves the operations of vector addition and scalar multiplication. Once a vector space has been transformed by a linear transformation, it may be transformed again by another or the same linear transformation.

Consider two co-ordinate systems and suppose we need to transform the components of a given vector in one co-ordinate system into those in another system. When the components of a vector in one co-ordinate system can be expressed as a linear combination of those in another co-ordinate system, the transformation is a *linear transformation*. When the origin of the two co-ordinate systems is the same, the transformation is said to be *homogeneous*, otherwise it is *inhomogeneous*. Most of the time, we deal with homogeneous transformations.

A.10 Matrix Algebra

A.10.1 Types of Matrices

1. **Symmetric matrix**: A matrix $\mathbf{V}$ is *symmetric* if the transpose of $\mathbf{V}$ is equal to $\mathbf{V}$.

2. **Skew-symmetric matrix**: A matrix $\mathbf{V}$ is *skew-symmetric* if the transpose of $\mathbf{V}$ is equal to $-\mathbf{V}$.

3. **Hermitian metric matrix**: A matrix $\mathbf{V}$ is Hermitian if transpose of its complex conjugate is equal to $\mathbf{V}$.

4. **Skew-Hermitian metric matrix**: A matrix $\mathbf{V}$ is Skew-Hermitian if transpose of its complex conjugate is equal to $-\mathbf{V}$.

5. **Singular matrix**: If the determinant of $\mathbf{V}$ is equal to zero then $\mathbf{V}$ is a singular matrix.

6. **Non-singular matrix**: If the determinant of $\mathbf{V}$ is not equal to zero then $\mathbf{V}$ is a non-singular matrix.

7. **Positive definite matrix**: A real symmetric matrix $\mathbf{V}$ is positive definite if the quadratic from $\boldsymbol{x}'\mathbf{V}\boldsymbol{x}$ is positive for all vectors $\boldsymbol{x}$.

8. **Positive semi-definite matrix**: A real symmetric matrix $\mathbf{V}$ is positive semi-definite if the quadratic from $\boldsymbol{x}'\mathbf{V}\boldsymbol{x} \geq \mathbf{0}$ for all vectors $\boldsymbol{x}$ and $\boldsymbol{x}'\mathbf{V}\boldsymbol{x}$ is zero for atleast one $\boldsymbol{x}$.

9. **Negative definite matrix**: A real symmetric matrix $\boldsymbol{V}$ is positive definite if the quadratic from $\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x} < \mathbf{0}$ for all vectors $\boldsymbol{x}$.

10. **Negative semi-definite matrix**: A real symmetric matrix $\mathbf{V}$ is positive semi-definite if the quadratic from $\boldsymbol{x}'\mathbf{V}\boldsymbol{x} \leq \mathbf{0}$ for all vectors $\boldsymbol{x}$ and $\boldsymbol{x}'\mathbf{V}\boldsymbol{x}$ is zero for atleast one $\boldsymbol{x}$. The matrix $\mathbf{V}$ is *nonnegative definite* if it is *negative definite*.

11. **Unitary matrix**: A unitary matrix $\mathbf{U}$ is a complex matrix which satisfies the condition $\mathbf{U}^{\dagger}\ \mathbf{U} = \mathbf{U}\ \mathbf{U}^{\dagger} = \mathbf{I}$. Here $\mathbf{I}$ is the identity matrix and $\mathbf{U}^{\dagger}$ is the Hermitian adjoint of $\mathbf{U}$. This condition says that a matrix $\mathbf{U}$ is unitary if and only if it has an inverse which is equal to its conjugate transpose $\mathbf{U}^{\dagger}$, i.e., $\mathbf{U}^{-1} = \mathbf{U}^{\dagger}$.

12. **Orthogonal matrix**: Orthogonal matrix is an unitary matrix in which all entries are real.

A.10.2 Inverse of a Matrix

A matrix $\mathbf{V}$ is said to be *invertible* if there exists a matrix $\mathbf{W}$ such that $\mathbf{V}\ \mathbf{W} = \mathbf{W}\ \mathbf{V} = \mathbf{I}$. If this is the case, then the matrix $\mathbf{W}$ is uniquely determined by $\mathbf{V}$ and is called the inverse of $\mathbf{V}$, denoted by $\mathbf{V^{-1}}$.

If the matrix is a non-square matrix, then it does not have an inverse. A square matrix that is *not invertible* is called *singular* or *degenerate* matrix. A square matrix is *singular* if and only if its determinant is *zero*.

A.10.3 Diagonalization of a Real Symmetric Matrix

Real symmetric matrices are diagonalizable by orthogonal matrices. If $\mathbf{M}$ is a real symmetric matrix, and $\mathbf{U}$ is some orthogonal matrix then the diagonalized matrix of $\mathbf{M}$ is given by $\mathbf{d} = \mathbf{U^T}\ \mathbf{M}\ \mathbf{U}$.

More generally, matrices are diagonalizable by unitary matrices if and only if they are *normal matrices*. (A matrix is normal if it commutes with its conjugate transpose).

A.10.4 Square Root of a Matrix

The square root of a square matrix $\mathbf{M}$ that is not necessarily positive-definite can be determined by the process of diagonalization. A matrix $\mathbf{S}$ is said to be a square root of $\mathbf{M}$ if the matrix product $\mathbf{S} \cdot \mathbf{S}$ is equal to $\mathbf{M}$.

The square root of a diagonal matrix $\mathbf{d}$ is obtained by taking the square root of all the entries on the diagonal. For general matrices, a square matrix $\mathbf{M}$ is diagonalizable if there is a matrix $\mathbf{U}$ such that $\mathbf{d} = \mathbf{U^{-1}}\ \mathbf{M}\ \mathbf{U}$ is a diagonal matrix. This happens if and only if $\mathbf{M}$ has $\boldsymbol{n}$ eigenvectors which constitute a basis for an $\boldsymbol{n}$-dimensional complex vector space $\mathbb{C}^{\boldsymbol{n}}$. In this case, $\mathbf{U}$ can be chosen to be the matrix with the $\boldsymbol{n}$ eigenvectors as columns. Now, $\mathbf{M = U\ d\ U^{-1}.}$

Hence the square root of $\mathbf{M}$ becomes $\mathbf{M^{1/2} = U\ d^{1/2}\ U^{-1}.}$ This approach is valid only for diagonalizable matrices.

A.11 Orthogonalization Methods

Orthogonalization methods are basically classified into two categories viz., *sequential* and *non-sequential* or *democratic.*

In the following subsections, we have given the essence of few orthogonalization methods such as the Gram-Schmidt, the spectral decomposition, the principal component analysis and the singular value decomposition.

A.11.1 Gram-Schmidt Orthogonalization

Let $\mathbf{V}$ be a finite dimensional inner product space, which may be a complex vector space spanned by the basis vectors $\boldsymbol{\{|\vec{v}_1>, |\vec{v}_2>, \cdots, |\vec{v}_n>\}}$. According to the *Gram-Schmidt Orthogonalization Process*, an orthonormal set of basis vector functions $\boldsymbol{\{|\vec{o}_1>, |\vec{o}_2>, \cdots, |\vec{o}_n>\}}$ can be constructed *sequentially* from any given basis using the following steps.

(1) Initially there is no constraint on choosing the first orthonormal basis vector $\boldsymbol{\vec{o}_1}$. We can choose any of the given basis vectors as $\boldsymbol{\vec{o}_1}$. Let us take,

$$\boldsymbol{|\vec{o}_1> = |\vec{v}_1>.}$$

(2) The second basis vector, which is orthogonal to the first one, can be constructed by using the orthogonality condition:

$$<\vec{o}_2|\vec{o}_1>=0.$$

Pick arbitrarily a second vector, say, $|\vec{v}_2>$ and do the following.

$$|\vec{o}_2>=|\vec{v}_2> -c\,|\vec{o}_1> .$$

The unknown scalar constant c is determined by substituting $|\vec{o}_2>$ in the orthogonality condition, which gives

$$c=\frac{<\vec{v}_2|\vec{o}_1>}{<\vec{o}_1|\vec{o}_1>}.$$

Thus c is the projection of $|\vec{v}_2>$ on to $|\vec{o}_1>$.

(3) Continue the process with all the vectors $|\vec{v}_k>$'s picking them randomly,

$$|\vec{o}_k>=|\vec{v}_k> -c_1\,|\vec{o}_1> -c_2\,|\vec{o}_2> - \cdots - c_{k-1}\,|\vec{o}_{k-1}>,$$

where the constants $c_1, c_2, \cdots, c_{k-1}$ are determined by satisfying the orthogonality conditions:

$$<\vec{o}_k|\vec{o}_j>=0 \textit{ for } j=1,2,\cdots,k-1$$

Since $|\vec{o}_1>, |\vec{o}_2>, \cdots, |\vec{o}_{k-1}>$ are already orthogonal to each other, the scalar constant c_j can be written as:

$$c_j=\frac{<\vec{v}_k|\vec{o}_j>}{<\vec{o}_j|\vec{o}_j>},$$

and, we have the following general *Gram-Schmidt formula* for constructing the orthogonal basis vectors $|\vec{o}_1>, |\vec{o}_2>, \cdots, |\vec{o}_n>$:

$$|\vec{o}_k>=|\vec{v}_k> -\sum_{j=1}^{k-1}\frac{<\vec{v}_k|\vec{o}_j>}{<\vec{o}_j|\vec{o}_j>}|\vec{o}_j>, \textit{ for } k=1,2,\cdots,n.$$

The orthonormality is ensured by normalizing $|\vec{o}_k>$'s.

By this construction, $\{|\vec{o}_1>, |\vec{o}_2>, \cdots, |\vec{o}_k>\}$ is an orthonormal set which spans a k-dimensional sub-space. When $k = n$, it spans the whole vector space.

Hence, the Gram-Schmidt orthogonalization procedure transforms any basis in an n-dimensional real vector space $\mathbb{R}^n$ (or a complex vector space $\mathbb{C}^n$) to an orthonormal basis in an n-dimensional real vector space $\mathbb{R}^n$ (or a complex vector space $\mathbb{C}^n$).

A.11.2 Spectral Decomposition

Let $\mathbf{M} = \mathbf{M}^T \in \mathbb{R}^{m\times m}$ be a real symmetric matrix in an m-dimensional real vector space. Then there exists an orthonormal matrix $\mathbf{U} \in \mathbb{R}^{m\times m}$, whose columns are orthonormal eigenvectors $\{|u_i>\}$ of $\mathbf{M}$, and eigenvalues $\{d_i\}$ such that the matrix $\mathbf{M}$ can be decomposed as

$$\mathbf{M} = \mathbf{U}\ \mathbf{d}\ \mathbf{U}^T = \mathbf{U}\ \mathbf{d}\ \mathbf{U}^\dagger = \sum_i d_i |u_i><u_i|.$$

This is called the *spectral decomposition* of $\mathbf{M}$.

A.11.3 Principal Component Analysis

Principal component analysis (PCA) is a mathematical technique which transforms a number of possibly correlated variables into a smaller number of uncorrelated variables. These are called *principal components.* The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. It is also sometimes referred to as the *KarhunenLoève transform* (KLT) or the *Hotelling transform.* PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each dimension.

PCA has been widely used in data analysis, data compression, dimension reduction, image processing, visualization, pattern recognition and time series analysis. PCA have also been applied to analyze the gene expression data.

PCA is essentially an orthogonal linear transformation which transforms the data to a new coordinate system such that the largest variance by any projection of the data lies on the first coordinate, called as the *first principal component*, the second largest variance on the second coordinate, called as the *second principal component* and so on.

The mathematical technique used in PCA is called *eigen analysis*. In this analysis, we solve for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix.

The given vectors can be written in matrix form by writing each vectors as a row of the matrix $\mathbf{X}$. Then the data matrix is centered by subtracting mean of each vector $\boldsymbol{\mu}$ from its components. Let it be $\mathbf{Y}$. The covariance matrix of sums of squared cross products (SSCP) is constructed using $\boldsymbol{[1/(n-1)]\, YY^T}$ and is solved for the eigenvalues and eigenvectors. The columns of eigenvector matrix constitute the direction of principal components.

For a data matrix, $\boldsymbol{X^T}$, with zero empirical mean, where each row represents a different repetition of the experiment, and each column gives the results from a particular probe, the PCA transformation is given by:

$$\mathbf{Y^T} = \mathbf{X^T W} = \mathbf{V\Sigma}, \tag{A.7}$$

where $\boldsymbol{V\Sigma W^T}$ is the singular value decomposition (SVD) of $\boldsymbol{X^T}$. Each column of $\boldsymbol{Y^T}$ represents a *principal component*. The components are called *principal factors* or *principal component scores*. The *first* and *second* principal components account for the maximum variability in the data set.

A.11.4 Singular Value Decomposition

A real $\boldsymbol{n} \times \boldsymbol{m}$ matrix $\mathbf{A}$ of rank $\boldsymbol{k}$ can be expressed as the product of three matrices which have a useful interpretation. This decomposition of $\mathbf{A}$ is referred to as a *singular value decomposition* and is given by

$$\mathbf{A} = \mathbf{U}\ \boldsymbol{\Sigma}\ \mathbf{V'} \tag{A.8}$$

where,

(a) $\boldsymbol{\Sigma}$ is a diagonal matrix with positive diagonal elements which are called *singular values* of $\boldsymbol{A}$ and are arranged in descending order.

(b) The $\boldsymbol{k}$ columns of $\mathbf{U}(\boldsymbol{n} \times \boldsymbol{k})$ are called *left singular vectors* of $\mathbf{A}$ and the $\boldsymbol{k}$ columns of $\mathbf{V}(\boldsymbol{m} \times \boldsymbol{k})$ are called the *right singular vectors* of $\mathbf{A}$.

(c) The matrices $\mathbf{U}$ and $\mathbf{V}$ have the property that $\mathbf{U'}\ \mathbf{U} = \mathbf{V'}\ \mathbf{V} = \mathbf{I}$. Hence the columns of $\mathbf{U}$ form an orthonormal basis for the columns of $\mathbf{A}$ in an $\boldsymbol{n}$-dimensional space while the columns of $\mathbf{V}$ form an orthonormal basis for the rows of $\mathbf{A}$ in an $\boldsymbol{m}$-dimensional space.

Bibliography

[1] Löwdin P-O; *Ark. Mat. Astr. Fys.* A **35**, 9, **1947**.

[2] Vipin Srivastava and Edwards S F; *Physica* A, **276**, 352-358, **2000**.

[3] Löwdin P-O; *Adv. Phys.* **5**, 1, **1956**.

[4] Löwdin P-O; *Adv. Quantum. Chem.* 185-199, **1990**.

[5] Vipin Srivastava; *J. Phys.* A: *Math. Gen.* **33**, 6219-6222, **2000**.

[6] Jolliffe I T; *Principal Component Analysis*, (Springer-Verlag) **1986**.

[7] Searle S R, *Matrix Algebra Useful for Statistics*, See Chapter 13, (John Wiley & Sons, **1982**).

[8] When N and n are interchanged, the n vectors in N-dimension will be represented by $\boldsymbol{V^T}$. And $\vec{\boldsymbol{M}}$ for $\boldsymbol{V^T}$ will be $\boldsymbol{(V^T)^T(V^T)}$, i.e. $\boldsymbol{VV^T}$, i.e. $\underset{\to}{\mathrm{M}}$.

[9] Rojas R; *Neural Networks* (Springer), **1996**.

[10] R. Courant and D. Hilbert; *Methods of Mathematical Physics*; 3rd ed; Interscience; New York, Vol. 1; **1953**.

[11] Schweinler H C and Wigner E P; *J.Math. Phys.* **97**, 1693-1694, **1970**.

[12] A. Ramesh Naidu and Vipin Srivastava; *International Journal of Quantum Chemistry* **99**, 882-888, **2004**.

[13] Vipin Srivastava and A. Ramesh Naidu; *International Journal of Quantum Chemistry* **106**, 1258-1266, **2006**.

[14] Golub, Gene H., Van Loan, Charles F; *Matrix Computations* (3rd ed.), The johnhopkins University Press **1996**.

[15] R. A. Horn and C. R. Johnson; *Topics in Matrix Analysis*, Cambridge University Press, Cambridge **1991**.

[16] E. Domany; *Journal of Statistical Physics*, **110**, 1117-1139, **2003**.

[17] A. Schulze and J. Downward; *Nature Cell. Biol.* **3**, 190-195, **2001**.

[18] Holter N S et. al.; *PNAS* vol. **97** No. 15, 8409-8414, **2000**.

[19] Holter N S et. al.; *PNAS* vol. **98** No. 4, 1693-1698, **2001**.

[20] J. J. Hopfield; *Proc. Natl. Acad. Sci. USA* vol. **79**, 2554-2558, **1982**.

[21] J. J. Hopfield; *Proc. Natl. Acad. Sci. USA* vol. **81**, 3088-3092, **1984**.

[22] D. J. Amit; *Modelong Brain Functions*, Cambridge Univeristy Press, Cambridge, **1989**.

[23] J. Hertz, A. Krogh, R. G. Palmer; *Introduction to the Theory of Neural Networks*, Cambridge University Press, Cambridge, **1982**.

[24] P. Peretto; *An Introduction to the Modeling of Neural Networks*, Cambridge University Press, Cambridge, **1992**.

[25] The mathematical computation of this value is given in D. J. Amit, H. Gutfreund, H. Sompolinski; *Phys. Rev. Lett.* **55** 1530-1533, **1985**.

[26] R. M. French; *Catasrophic forgetting in connectionist networks*, Trends Cogn. Sci. **3**, 128-135, **1999**.

[27] R. Ratcliff; *Connectionist models of recognition memory: constraints imposed by learning and forgetting functions* Psychol. Rev. **97**, 285-308, **1990**.

[28] G. F. Marcus, S. Vijayan, S. Bandi Rao, P. M. Vishton; *Science* **283**, 77-80, **1999**.

[29] S. H. Lee, R. Blake; *Science* **284**, 1165-1168, **1999**.

[30] E. R. Wood, P. A. Dudehenco, H. Eichenbaum; *Nature* **397**, 613-616, **1999**.

[31] J. N. P. Rawlins; *Nature* **397**, 561-563, **1999**.

[32] A. R. McIntosh, M. Rajah, N. J. Lobaugh; *Science* **284**, 1531-1533, **1999**.

[33] V. Srivastava, D. J. Parker, S. F. Edwards; *Journal of Theoretical Biology* **253**, 514-517, **2008**.

Appendix B

List of Publications

1. **Löwdin's canonical orthogonalization: Getting round the restriction of linear independence**, *International Journal of Quantum Chemistry* **99** 882-888; A. Ramesh Naidu and Vipin Srivastava 2004.

2. **New classes of orthogonal polynomials**, *International Journal of Quantum Chemistry* **106** 1258-1266; Vipin Srivastava and A. Ramesh Naidu 2006.

3. **Deciphering gene expression profiles**, in preparation; Vipin Srivastava and A. Ramesh Naidu 2010.