A unified view of the orthogonalization methods

2000 J. Phys. A: Math. Gen. 33 6219

(http://iopscience.iop.org/0305-4470/33/35/310)

J. Phys. A: Math. Gen. **33** (2000) 6219–6222. Printed in the UK PII: S0305-4470(00)12958-0

A unified view of the orthogonalization methods

Vipin Srivastava
School of Physics, University of Hyderabad, Hyderabad 500 046, India
E-mail: vpssp@uohyd.ernet.in

Received 29 March 2000

Abstract. A geometrical view of all known orthogonalization procedures is taken in order to understand their distinctive features and the inter-connections between them. Useful new information is gained. Its possible application to certain cognitive phenomena is also indicated.

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 a long time. It finds wide applications in mathematics, physics and chemistry. The methods involve a straightforward and simple mathematics and differ in the special conditions they satisfy. We present a simple analysis of the orthogonalization methods in terms of the projections of the given vectors on the orthogonalized bases vectors and find that the methods possess some powerful properties hitherto unknown. The discovered properties are not only curious but also have algorithmic value in that they enable geometrical constructions of the orthogonal basis sets and the eigenvalues of the metric matrix of the given vector set. These properties also underline the scenarios where the different methods should be applicable. We expect that the present insight should also provide clues to understand certain aspects of cognitive memory, e.g. along the lines of [1].

2. Orthogonalization methods—sequential and non-sequential

The Gram–Schmidt orthogonalization procedure, the oldest and most commonly used in mathematics (also applied recently to spin-glass like neural networks [1]), gives an orthonormal set depending on the sequence in which the given vectors are chosen. The work on sequence-independent orthogonalization, in a particular context of physical problems, was initiated by Landshoff [2] and Wannier [3], and later developed comprehensively by Löwdin [4]. Two notable methods of this class are Löwdin's symmetric [5] and canonical [6] orthogonalizations. The symmetric orthogonalization contains the results of Landshoff and Wannier [4]. The canonical orthogonalization is generally known in the physics literature (see [7] for references) as Schweinler–Wigner [8] (hereafter referred to as SW). Schweinler and Wigner in fact introduced a versatile parameter 'm' that highlighted an extremal property of Löwdin's canonical orthogonalization. Earlier Löwdin [9] had talked about an optimum principle obeyed by the canonical orthogonalization which we find in a sense resembles the SW extremal property. Recently Chaturvedi *et al* [7] have given a new orthonormal basis that has another extremal property in terms of the SW parameter m.

0305-4470/00/356219+04\$30.00

2.1. General framework

Let $\boldsymbol{V}$ represent a set of linearly independent vectors, $\vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_N$, in a 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 $\boldsymbol{Z}$:

$$\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{A}. \tag{1}$$

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

$$\langle \boldsymbol{Z}|\boldsymbol{Z}\rangle = \langle \boldsymbol{V}\boldsymbol{A}|\boldsymbol{V}\boldsymbol{A}\rangle = \boldsymbol{A}^\dagger\langle \boldsymbol{V}|\boldsymbol{V}\rangle\boldsymbol{A} = \boldsymbol{A}^\dagger\boldsymbol{M}\boldsymbol{A} = \boldsymbol{I} \tag{2}$$

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

$$\boldsymbol{A} = \boldsymbol{M}^{-1/2}\boldsymbol{B} \tag{3}$$

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, $\boldsymbol{Z} \equiv \boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{M}^{-1/2}$, while $\boldsymbol{B} = \boldsymbol{U}$, where $\boldsymbol{U}$ diagonalizes $\boldsymbol{M}$,

$$\boldsymbol{U}^\dagger\boldsymbol{M}\boldsymbol{U} = \boldsymbol{d} \tag{4}$$

gives the canonical orthogonalization, $\boldsymbol{Z} \equiv \boldsymbol{\Lambda} = \boldsymbol{V}\boldsymbol{U}\boldsymbol{d}^{-1/2}$. The orthonormal basis due to Chaturvedi *et al* [7], denoted here by $\boldsymbol{\Gamma}$, is related to $\boldsymbol{\Lambda}$, and coincides with $\boldsymbol{\Phi}$ in a particular circumstance indicated below.

2.2. Projections of $\boldsymbol{V}$ on $\boldsymbol{Z}$

Taking our cue from Schweinler and Wigner [8] we consider the matrix

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

of projection squares of the given vectors $\{\vec{v}_k\}$ on a basis set $\{\vec{z}_\kappa\}$. The elements in a row corresponding to a particular $\vec{v}_k$ add up to $|\vec{v}_k|^2$:

$$\sum_\kappa |(\vec{v}_k, \vec{z}_\kappa)|^2 = |\vec{v}_k|^2 \qquad k = 1, \ldots, N \tag{5}$$

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

$$\sum_k |(\vec{v}_k, \vec{z}_\kappa)|^2 = (\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}\boldsymbol{A}^\dagger)_{\kappa\kappa} = (\boldsymbol{B}\boldsymbol{M}\boldsymbol{B}^\dagger)_{\kappa\kappa} = c_\kappa \qquad \kappa = 1, \ldots, N. \tag{6}$$

Note that we have the identity

$$\sum_{\kappa=1}^{N} c_\kappa = \sum_{k=1}^{N} |\vec{v}_k|^2 \qquad \text{a constant for a given set } \boldsymbol{V}. \tag{7a}$$

Further, we have

$$\sum_\kappa c_\kappa^2 = m \qquad \text{the SW parameter.} \tag{7b}$$

A basis set $\boldsymbol{Z}$ will satisfy the set of simultaneous equations (6) with the positive real numbers c_κ obeying the identity (7*a*)†. Specific bases will satisfy additional conditions on the values of c_κ either through (7*b*) or otherwise. For instance, if $c_\kappa = |\vec{v}_k|^2$ with

† The set of equations (6) will have many more solutions than those that form orthonormal bases. Many of these may not even be bases. A N-dimensional orthonormal basis set will require $N(N-1)/2$ conditions to be fulfilled.

$\kappa = k$ (i.e. $\boldsymbol{B} = \boldsymbol{I}$ in (6)), we will get the *symmetric* basis $\boldsymbol{Z} = \boldsymbol{\Phi}$; $m = m_{\max}$, which will arise for the maximally lop-sided distribution of the c_κ (satisfying (7*a*)), will give the *canonical* basis $\boldsymbol{Z} = \boldsymbol{\Lambda}$; and $m = m_{\min}$, which will correspond to an average distribution, $c_1 = c_2 = \cdots = c_N = (c_1 + c_2 + \cdots + c_N)/N$, will give the basis $\boldsymbol{Z} = \boldsymbol{\Gamma}$ of Chaturvedi *et al* [7]. For normalized $\vec{v}_k$ the basis $\boldsymbol{\Gamma}$ and the symmetric basis $\boldsymbol{\Phi}$ become the same.

2.3. New information

The 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 $\vec{v}_k$ on a $\vec{\phi}_\kappa$, say $\vec{\phi}_l$, is equal to the sum of squared projections of all the $\vec{\phi}_\kappa$ on the $\vec{v}_k$ with $k = l$, 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 [4], $\langle \boldsymbol{Z} - \boldsymbol{V} | \boldsymbol{Z} - \boldsymbol{V} \rangle$, has its smallest value when $\boldsymbol{B} = \boldsymbol{I}$ or $\boldsymbol{Z} = \boldsymbol{\Phi}$. If viewed slightly differently, the above symmetry interestingly implies that the squared projections of all the $\vec{v}_k$ on a $\vec{\phi}_l$ add up to the squared length of $\vec{v}_l$, the feature that can be used to geometrically generate the symmetric basis set.

The last property above turns into a stricter condition in the case of $m = m_{\min}$, where $\boldsymbol{Z} = \boldsymbol{\Gamma}$: the basis vectors $\vec{\gamma}_\kappa$ are arranged such that the sum of squared projections of all the $\vec{v}_k$ on each $\vec{\gamma}_\kappa$ is the same—equal to the average of $|\vec{v}_k|^2$—irrespective of how the $\vec{v}_k$ are arranged. In effect, the set $\boldsymbol{\Gamma}$ is arranged so as to cancel the effects of inhomogeneity in the distribution of $\vec{v}_k$.

On the other hand, in the $m = m_{\max}$ case, with $\boldsymbol{Z} = \boldsymbol{\Lambda}$, the basis vectors $\vec{\boldsymbol{\lambda}}_\kappa$ must be oriented such that they sample those directions in which bunches of $\vec{v}_k$ tend to be oriented. In order to attain $m = m_{\max}$ the canonical basis set is arranged in such an optimal fashion that the sum of squared projections of all the $\vec{v}_k$ on one of the $\vec{\lambda}_\kappa$, say $\vec{\lambda}_1$ (i.e. c_1) is the largest. After the $\vec{\lambda}_1$ is fixed, the rest of the set, orthogonal to $\vec{\lambda}_1$, is oriented such that another $\vec{\lambda}_\kappa$, say $\vec{\lambda}_2$, is able to maximize the total squared projection of all the $\vec{v}_k$ on it. This will be c_2 and will be smaller than c_1. All the $\vec{\lambda}_\kappa$ are arranged according to this optimum-principle that ensures for the given set $\{\vec{v}_k\}$ the most lop-sided distribution of c_κ with $c_1 > c_2 > \cdots > c_N$. This particular set of c_κ in fact comprises the eigenvalues of $\boldsymbol{M}$ in the basis $\boldsymbol{\Lambda}$ (take $\boldsymbol{B} = \boldsymbol{U}$ in equation (6)) which are generically non-degenerate. The arrangement of $\vec{\lambda}_\kappa$ that yields the SW condition of $m = m_{\max}$ manifests Löwdin's optimal property [4, 9] of the canonical orthogonalization. Note that just as the c_κ gains its largest values $(\boldsymbol{d})_{\kappa\kappa}$ for $\boldsymbol{B} = \boldsymbol{U}$, the quantity

$$\sum_\alpha |A_{\alpha\kappa}|^2 = (\boldsymbol{A}^\dagger \boldsymbol{A})_{\kappa\kappa} = (\boldsymbol{B}^\dagger \boldsymbol{M}^{-1} \boldsymbol{B})_{\kappa\kappa} \tag{8}$$

has its smallest values for $\boldsymbol{B} = \boldsymbol{U}$, in which case it is $(\boldsymbol{d}^{-1})_{\kappa\kappa}$. If d_1^{-1} is the smallest value and the associated basis vector is $\vec{\lambda}_1$, Löwdin [9] showed that for all vectors orthogonal to $\vec{\lambda}_1$ the sum $\sum_\alpha |A_{\alpha\kappa}|^2$ has the smallest value $d_2^{-1} (>d_1^{-1})$ associated with $\vec{\lambda}_2$, and that one can go on in this manner to find the smallest d_κ^{-1} associated with the respective $\vec{\lambda}_\kappa$.

It should be noted that the value of $m_{\max}$ depends on the distribution of orientations of $\vec{v}_k$ relative to each other, whereas the value of $m_{\min}$ is independent of this. There can be any number of $\boldsymbol{V}$ satisfying a particular identity (7*a*) but differing in distributions of orientations of the vectors $\vec{v}_k$. Each of these will have a different $m_{\max}$ but the same $m_{\min}$. The inhomogeneity in the distribution of directions of a given set of vectors will decide the value of $m_{\max}$—the larger the inhomogeneity, the larger will be the value of $m_{\max}$.

Now we will examine the Gram–Schmidt orthogonalization (represented by $\boldsymbol{Z} = \boldsymbol{\Omega}$) in the framework of equation (6). The bases $\boldsymbol{\Omega}$ (=$\{\vec{\omega}_\kappa\}$), where for a predetermined sequence of $\vec{v}_k$ a sequence of $\vec{\omega}_\kappa$ is generated such that a $\vec{v}_k$ has non-zero projections only on those $\vec{\omega}_\kappa$

for which $\kappa \leqslant k$, satisfy the modified set of equation (6) with $|(\vec{v}_k, \vec{z}_\kappa)|^2 = 0$ for all $\kappa > k$ together with the identity $(7a)$. It is clear at the outset that none of the $\boldsymbol{\Omega}$ ($N!$ in number) can coincide with the symmetric basis $\boldsymbol{\Phi}$, since for any $\boldsymbol{\Omega}$ at least c_1 is necessarily greater than $|\vec{v}_1|^2$. But, an $\boldsymbol{\Omega}$ corresponding to a certain sequence of $\vec{v}_k$ can accidently coincide with the bases $\boldsymbol{\Gamma}$ or $\boldsymbol{\Lambda}$. The possibility of the coincidence with $\boldsymbol{\Lambda}$ will however be precluded if the $\vec{v}_k$ are normalized since then $\boldsymbol{\Lambda}$ coincides with $\boldsymbol{\Phi}$ as seen above.

3. Conclusion

In summary, we find that it is easy to visualize all the orthogonalization methods in terms of the projections of the given vectors on the orthogonal bases. This exercise elucidates interesting geometrical features of the orthogonalized bases hidden in their optimal, extremal and symmetry properties and also shows how these basis sets as well as the eigenvalues of the metric matrix of the given $\boldsymbol{V}$ can be constructed purely from geometrical considerations.

Finally we may point out that the recent proposal [1] that Schmidt's orthogonalization is a plausible cognitive function the brain might be employing to discriminate an object from another, is constrained by the orthogonalization procedure's 'sequence-dependence' which makes it applicable to phenomena like our thought process and use of language etc, that are sequential in time. However, there are a host of cognitive phenomena that are not sequential in time. For instance, detection of changes in a visual array, categorization according to numerosities, etc [10]. It is expected that the understanding developed here of the sequence-independent orthogonalizations will prove valuable in exploring such phenomena.

Acknowledgments

Thanks are due to Professor P-O Löwdin for correspondence and encouragement, and to Professor S Chaturvedi and Professor K Viswanath for discussions. Financial support was provided by the Ministry of Information Technology, Government of India.

References

[1] Vipin Srivastava and Edwards S F 2000 *Physica* A **276** 352
[2] Landshoff R 1936 *Z. Phys.* **102** 201
[3] Wannier G H 1937 *Phys. Rev.* **52** 191
[4] For a review, see Löwdin P O 1990 *Adv. Quantum. Chem.* 185
[5] Löwdin P-O 1947 *Ark. Mat. Astr. Fys.* A **35** 9
[6] Löwdon P-O 1956 *Adv. Phys.* **5** 1
[7] Chaturvedi S, Kapoor A K and Srinivasan V 1998 *J. Phys. A: Math. Gen.* **31** L367
[8] Schweinler H C and Wigner E P 1970 *J. Math. Phys.* **97** 1693
[9] Löwdin P-O 1955 *Phys. Rev.* **97** 1474
[10] Butterworth B 1999 *Science* **284** 928
See also Dehaene S *et al* 1999 *Science* **284** 970