Revista Mexicana de Física 31 Nº 4 (1985) 743-758

ORTHONORMALIZATION ON THE PLANE: A GEOMETRIC APPROACH

C. E. Solivérez†
Centro Atómico Bariloche# and Instituto Balseiro§
csoliverez@gmail.com

and

E. R. Gagliano† (deceased)

Instituto de Desarrollo Tecnológico para la Industria Química

(recibido junio 4 de 1984; aceptado marzo 8, 1985)

ABSTRACT

When confininig oneself to the two-dimensional case it is possible to give simple geometric interpretations to the properties of alternative orthonormalization schemes. The symmetry, proximity and localization properties of the symmetric orthonormalization, and the delocalization property of the canonical scheme, among others, become thus evident at first sight.

---

† Affiliated to CONICET.

Comisión Nacional de Energía Atómica.

§ Comisión Nacional de Energía Atómica and Universidad Nacional de Cuyo.

RESUMEN

Al restringirse al caso bidimensional es posible dar interpretaciones geométricas sencillas de los diferentes esquemas de ortonormalización utilizadas en física y química. Las propiedades de simetría, proximidad y localización de la ortonormalización simétrica, y la de delocalización del esquema canónico, entre otras, resultan entonces evidentes a primera vista.

1. INTRODUCTION

Orthonormal bases are customarily used when solving physical and chemical problemas, because then the mathematical formulation usually becomes simpler. There are, nevertheless, situations where the natural bases are nonorthonormal. That is, for instance, the case of the basis vectors of crystallography, or the atomic orbitals used in molecular and solid-state calculations. This is so in the last case because different orbitals, although orthogonal when belonging to the same atom, have nonvanishing overlap when centered on different atoms.

As textbooks on mathematics discuss solely the Gram-Schmidt procedure[1] one may get the impression that this is the only known orthonormalization scheme. There are, in fact, infinitely may such schemes, the Gram-Schimdt one being probably the most cumbersome of all.

While the properties of all alternative orthonormalization schemes are long since well known[2], the necessity of working in a many-dimensional vectorial space precludes an easy visualization of the mathematical results. If one starts by considering the two-dimensional case the orthonormalization problem becomes instead a simple geometrical problem on the plane which may be solved almost by inspection. It is then easy to obtain a good grasp of the many important differences between the available procedures. As the two-dimensional case has an interest of its own (remember molecular orbitals) and the better understanding thus obtained is valuable for the general many-dimensional case, a detailed discussion is well warranted.

2. ORTHONORMALIZATION ON THE PLANE

Let ϕ_1 and ϕ_2 be two real, nonorthogonal, linearly independent and normalized wavefunctions:

$$\begin{aligned} <\phi_1|\phi_1> &= <\phi_2|\phi_2> = 1, \\ <\phi_1|\phi_2> &= <\phi_2|\phi_1> = S, \end{aligned} \tag{1}$$

where S is the overlap. If we take them to be vectors on the plane we may write, instead, the scalar product as

$$<\phi_j\,|\phi_k> \;= \mathrm{g}_{jk} = \vec{\phi}_j \bullet \vec{\phi}_k, \tag{2}$$

g_{jk} being an element of the symmetric metric matrix

$$\mathbf{g} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}. \tag{3}$$

Linear independence implies that $\vec{\phi}_1$ and $\vec{\phi}_2$ are not parallel, that is (see Fig. 1)

$$\begin{aligned} &-1 < S = \vec{\phi}_1 \bullet \vec{\phi}_2 = \cos\gamma < 1, \\ &\text{Det }\mathbf{g} = \begin{vmatrix} 1 & S \\ S & 1 \end{vmatrix} = 1 - S^2 \neq 0, \end{aligned} \tag{4}$$

thus showing $\mathbf{g}$ to be a nonsingular matrix with an inverse.

Our problem is finding a real transformation matrix $\mathbf{O}$ such that the new vectors

$$\vec{\phi}'_j = \sum_l \mathrm{O}_{lj}\,\vec{\phi}_l \tag{5}$$

are orthonormal. That is

$$\vec{\phi}'_j \bullet \vec{\phi}'_k = \delta_{jk}, \quad \mathbf{O}^{\mathrm{t}} \bullet \mathbf{g} \bullet \mathbf{O} = \mathbf{1}, \tag{6}$$

where δ_{jk} is Kronecker's delta, $\mathbf{O}^{\mathrm{t}}$ the transpose of $\mathbf{O}$ and $\mathbf{1}$ the unit matrix.

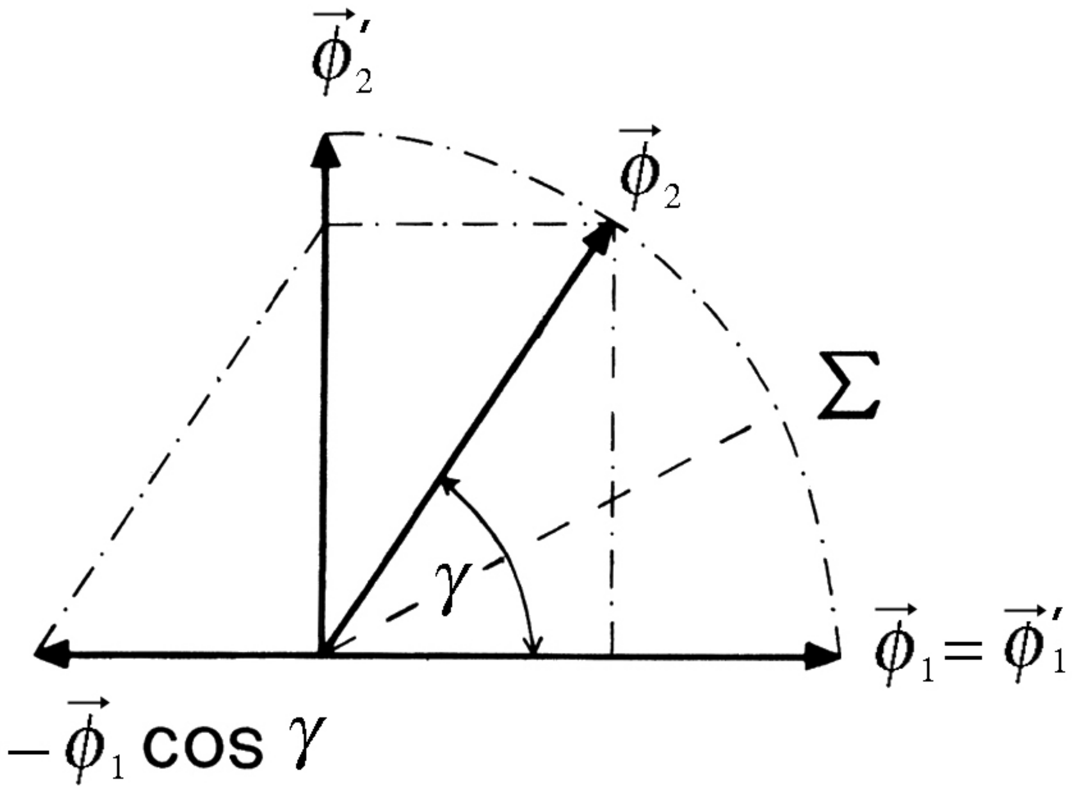

Fig. 1. Two unit vectors $\vec{\phi}_1$ and $\vec{\phi}_2$ are linearly independent if the angle γ between them is different from 0, π. A vector $\vec{\phi}_2'$ normal to $\vec{\phi}_1$ may then be obtained by the Gram–Schmidt procedure Eq. (8). The dashed line Σ is a mirror line.

The general solution of the second Eq. (6) is

$$\mathbf{O} = \mathbf{g}^{-\frac{1}{2}} \bullet \mathbf{U}, \quad \mathbf{U}^{t} = \mathbf{U}^{-1}, \tag{7}$$

where **U** is an arbitrary orthogonal matrix. In the general case of complex vectors the hermitian adjoint should be taken instead of the transpose, and **U** should be taken to be unitary.

The best known orthonormalization method is the Gram–Schmidt process[1] where, starting with a given vector $\vec{\phi}_1$, a normal vector $\vec{\phi}_2'$ is obtained from

$$\vec{\phi}_2' = \frac{\left(\vec{\phi}_2 - \cos\gamma\,\vec{\phi}_1\right)}{\left\|\vec{\phi}_2 - \cos\gamma\,\vec{\phi}_1\right\|}. \tag{8}$$

It is clear that the Gram-Schmidt method gives the leading role to the initial vector, thus destroying the symmetry properties of the primitive set. Assume, for instance, that $\vec{\phi}_1$ and $\vec{\phi}_2$ are related by the mirror line Σ in Fig. 1,

$$\mathbf{\Sigma} \bullet \vec{\phi}_1 = \vec{\phi}_2, \quad \mathbf{\Sigma} \bullet \vec{\phi}_2 = \vec{\phi}_1, \tag{9}$$

where $\mathbf{\Sigma}$ is the diadic performing the reflection. It is immediately seen that the reflection symmetry is not preserved for the new basis, this being a consequence of the unequal weight given to each primitive vector in the construction of the respective new one, as measured by the projection $\vec{\phi}_j \bullet \vec{\phi}_j'$.

An alternative procedure comes to mind at once where the new orthogonal vectors are symmetrically rotated respect to the old ones, as shown in Fig. 2. It may be seen that the reflection symmetry is now preserved for the new vectors. As a matter of fact all unitary relationships between basis vectors are preserved in this orthonormalization process[3] which corresponds to Löwdin's symmetric orthonormalization

$$\mathbf{O} = \mathbf{g}^{-\frac{1}{2}}. \tag{10}$$

We will now find $\mathbf{g}^{-\frac{1}{2}}$. It should first be noticed that any square root of g (there are four of them, but only one is appropiate) must verify the equation

$$\mathbf{g}^{\frac{1}{2}} \bullet \mathbf{g}^{\frac{1}{2}} = \mathbf{g}, \quad \text{where} \quad \mathbf{g}^{\frac{1}{2}} = \begin{pmatrix} p & q \\ q & p \end{pmatrix}. \tag{11}$$

From here we obtain

$$\begin{aligned} &p^2 + q^2 = 1, \quad 2pq = \cos\gamma, \\ &p = \cos\alpha, \qquad q = \sin\alpha, \end{aligned} \tag{12}$$

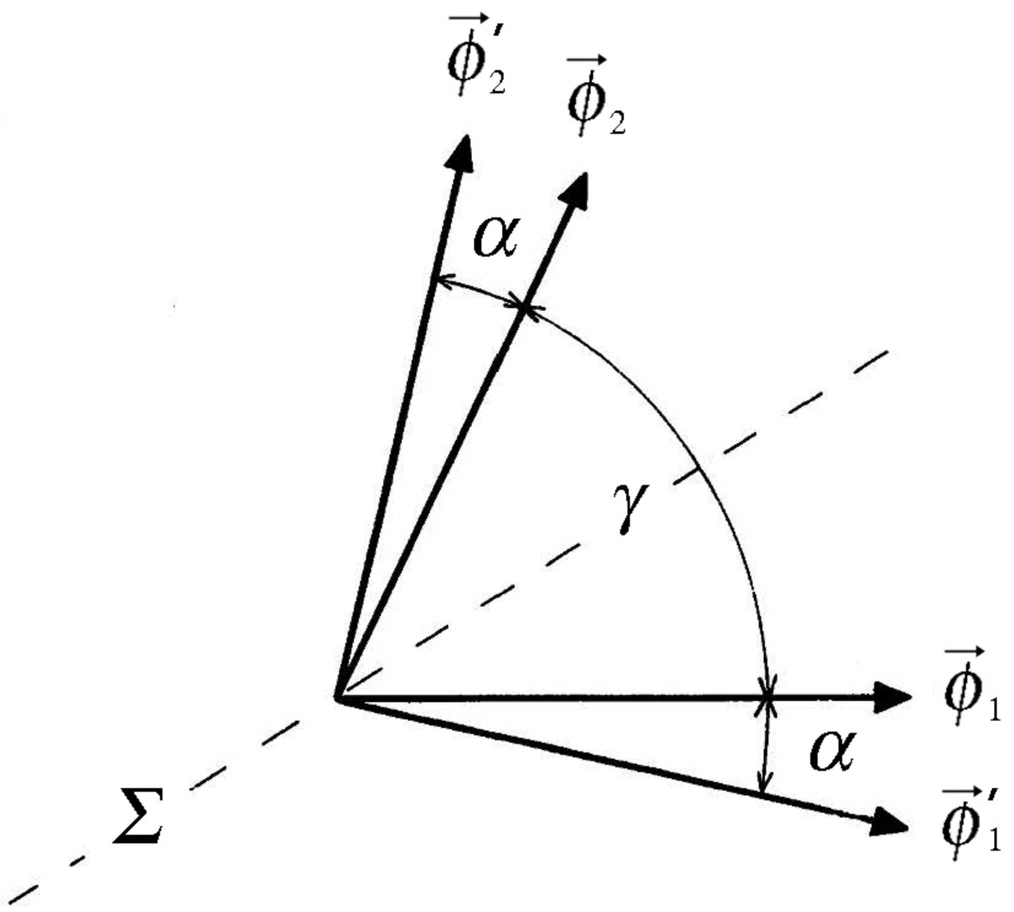

Fig. 2. The symmetric orthonormalization process generates new vectors $\vec{\phi}'_1$, $\vec{\phi}'_2$ which are symmetrically rotated respect to the original ones, thus preserving orthogonal relationships such as the mirror line Σ.

$$\sin 2\alpha = \cos\gamma. \tag{13}$$

Because $\gamma \neq 0, \pi$ it is easily seen from Fig. 3 that Eq. (13) always has four different solutions in the interval $(-\pi;\pi)$, as given by

$$\alpha = \frac{\pi}{4} + \frac{\gamma}{2} + n\pi, \tag{14}$$

where the choice of the integer n depends on γ.

The inverse matrix $\mathbf{g}^{-\frac{1}{2}}$ may now be written

$$\mathbf{g}^{-\frac{1}{2}} = \begin{pmatrix} \dfrac{\cos\alpha}{\cos 2\alpha} & \dfrac{-\sin\alpha}{\cos 2\alpha} \\ \dfrac{-\sin\alpha}{\cos 2\alpha} & \dfrac{\cos\alpha}{\cos 2\alpha} \end{pmatrix}. \tag{15}$$

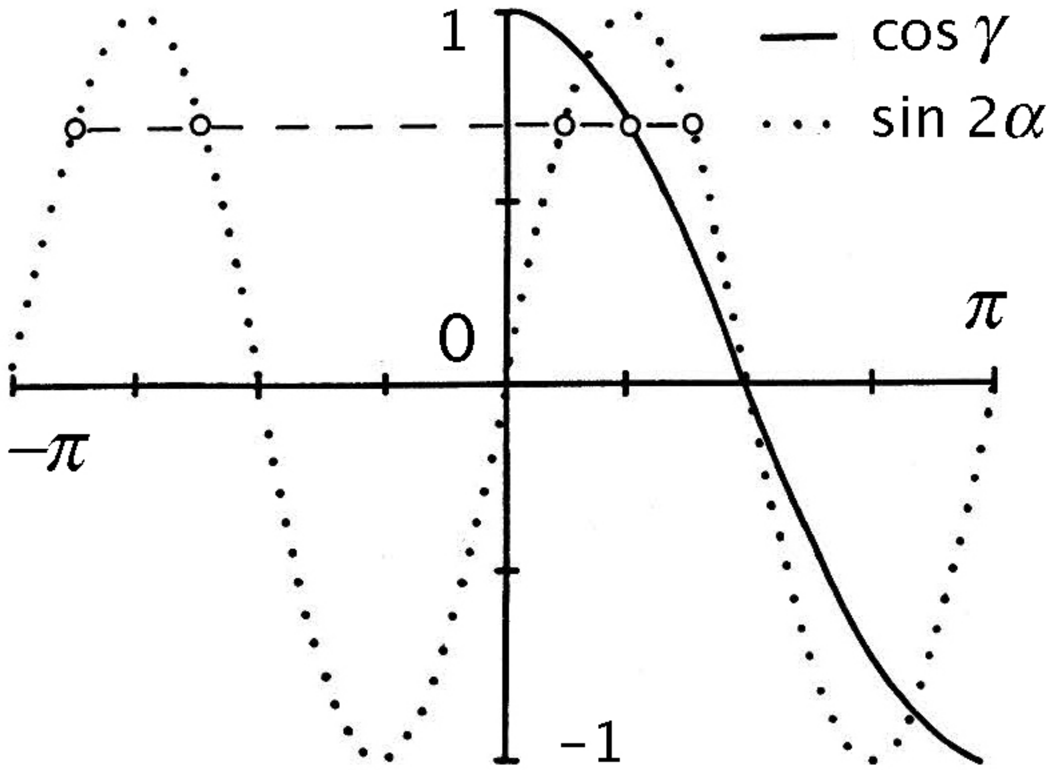

Fig. 3. For any given value of γ in (0; π) ($\gamma \neq$ 0, π) there are always four differente solutions to the equation $\sin 2\alpha = \cos\gamma$ which determines $\mathbf{g}^{-½}$, here indicated by circles.

The four different values of α correspond to the four different combinations of signs of the eigenvalues of $\mathbf{g}^{-½}$ which are obtained from the secular equation

$$\left(\frac{\cos\alpha}{\cos 2\alpha} - \lambda\right)^2 - \left(\frac{\sin\alpha}{\cos 2\alpha}\right)^2 = 0, \tag{16}$$

namely

$$\lambda^{\pm} = \frac{\cos\alpha \mp \sin\alpha}{\cos 2\alpha} = \frac{\sin\left(\alpha \pm \dfrac{\pi}{4}\right)}{\sqrt{2}\cos 2\alpha}, \tag{17}$$

eigenvalues which are plotted in Fig. 4. Replacing (10) and (15) into (5) we obtain

$$\vec{\phi}_1' = \frac{\cos\alpha}{\cos 2\alpha}\vec{\phi}_1 - \frac{\sin\alpha}{\cos 2\alpha}\vec{\phi}_2,$$

$$\vec{\phi}_2' = -\frac{\sin\alpha}{\cos 2\alpha}\vec{\phi}_1 + \frac{\cos\alpha}{\cos 2\alpha}\vec{\phi}_2. \tag{18}$$

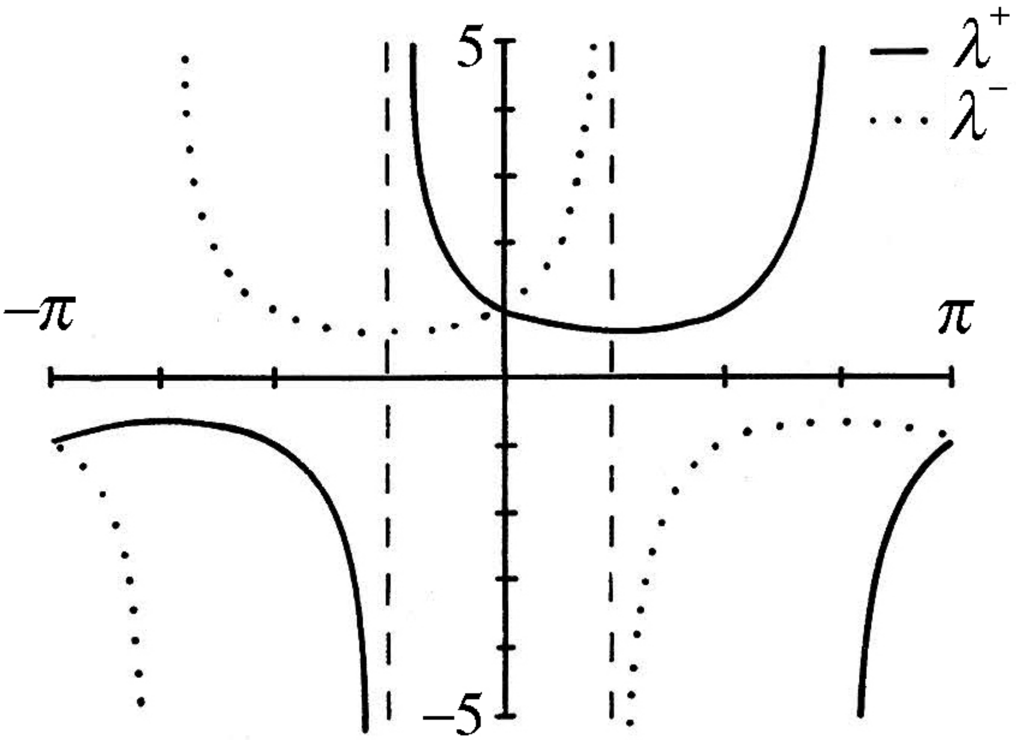

Fig. 4. Eigenvalues of $\lambda^{\pm}$ of $\mathbf{g}^{-1/2}$ as a function of α. A comparison with Fig. 3 shows the branch $-\pi/4 \le \alpha \le \pi/4$ to be the only one which makes $\mathbf{g}^{-1/2}$ positive definite.

The new vectors are easily verified to be orthonormalized and to correspond to a rotation in α of the primitive vectors, as may be seen from

$$\vec{\phi}_1 \bullet \vec{\phi}_1' = \vec{\phi}_2 \bullet \vec{\phi}_2' = \frac{\cos\alpha - \sin\alpha\cdot\cos\gamma}{\cos 2\alpha} = \cos\alpha, \tag{19}$$

where use has been made of Eq. (13). We may see from Fig. 4 that both eigenvalues are positive when

$$-\frac{\pi}{4} \le \alpha \le \frac{\pi}{4}, \quad \alpha = \frac{\pi}{4} - \frac{\gamma}{2}, \tag{20}$$

this being the only branch in Fig. 3 which makes $\mathbf{g}^{-1/2}$ positive definite[4]. This solution, depicted in Fig. 2, has the unique property of giving orthonormalized vectors which are the closest to the old ones in the sense that

$$\Delta^2 = \left\|\vec{\phi}_1' - \vec{\phi}_1\right\|^2 + \left\|\vec{\phi}_2' - \vec{\phi}_2\right\|^2 \tag{21}$$

is a minimum. This property, which holds for the general many-dimensional case[5], may be easily proved here. To that end consider, as in Fig. 5, an arbitrary pair of orthonormal vectors where

$$\left\|\vec{\phi}_j' - \vec{\phi}_j\right\| = \left|2\sin(\alpha_j/2)\right|, \quad \gamma = \alpha_1 + \alpha_2 + \frac{\pi}{2}. \tag{22}$$

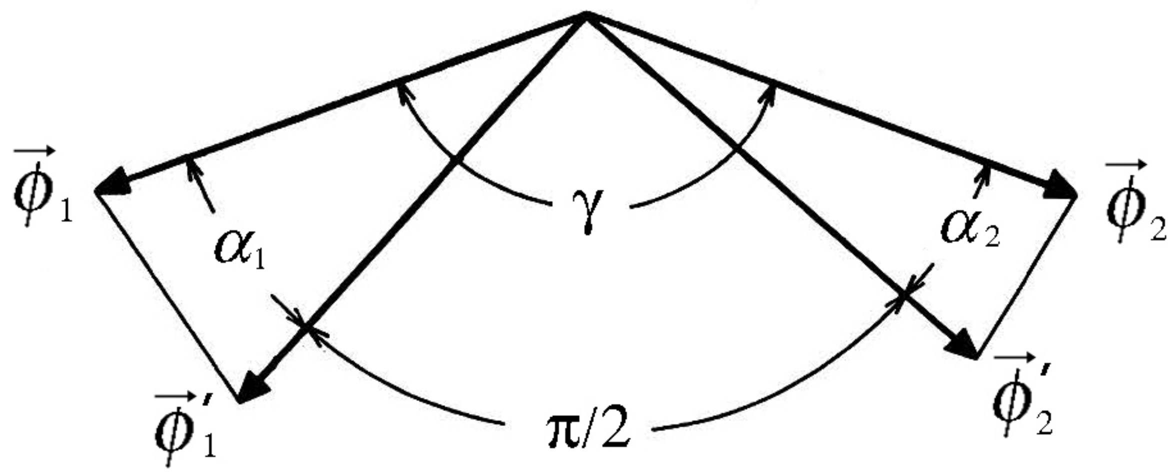

Fig. 5. For a general orthonormalized pair $\vec{\phi}_1'$, $\vec{\phi}_2'$ the angles α_1 and α_2. will be different. The case is here illustrated where $\gamma > \pi/2$ for which it is proved in the text that Δ^2 Eq. (21) is a minimum when $\alpha_1 = |\alpha_2|$.

Therefore

$$\begin{aligned}(\Delta/2)^2 &= \sin^2(\alpha_1/2) + \sin^2(\alpha_2/2)\\ &= \sin^2(\alpha_1/2) + \sin^2([\gamma - \alpha_1 - \pi/2]/2)\\ &= 1 - \frac{1}{2}\cos\alpha_1 - \frac{1}{2}\cos(\gamma - \alpha_1 - \pi/2),\end{aligned} \tag{23}$$

which is easily verified to be a minimum if

$$\alpha_1 = -\pi/4 + \gamma/2 = |\alpha_2|. \tag{24}$$

A similar proof holds if $\gamma > \pi/2$ showing that it should be $|\alpha_1| = |\alpha_2|$. The other three solutions for α yield orthonormalized vectors which do not have the proximity property and are related to $\vec{\phi}_1^{'}$ and $\vec{\phi}_2^{'}$ Eq. (20) by

$$\begin{aligned}&\vec{\phi}_1^{''} = -\vec{\phi}_1^{'}, \quad \vec{\phi}_2^{''} = -\vec{\phi}_2^{'},\\ &\vec{\phi}_1^{'''} = \vec{\phi}_2^{'}, \quad \vec{\phi}_2^{'''} = \vec{\phi}_1^{'},\\ &\vec{\phi}_1^{''''} = -\vec{\phi}_2^{'}, \quad \vec{\phi}_2^{''''} = -\vec{\phi}_1^{'}.\end{aligned} \tag{25}$$

In Fig. 6 a plot is given of the values of the coefficients in Eq. (18) corresponding to the positive-definite branch Eq. (20). They are seen to blow-up in the linearly dependent case $S = \cos\gamma = \pm 1$. The reason for this behaviour may be understood in the following fashion. Let us first define

$$\vec{\phi}^{\pm} = \frac{1}{2}(\vec{\phi}_1 \pm \vec{\phi}_2), \tag{26}$$

from which it follows that

$$\begin{aligned}\vec{\phi}_1^{'} &= \frac{\vec{\phi}^{+}}{\sqrt{2}\cos(\gamma/2)} + \frac{\vec{\phi}^{-}}{\sqrt{2}\sin(\gamma/2)},\\ \vec{\phi}_2^{'} &= \frac{\vec{\phi}^{+}}{\sqrt{2}\cos(\gamma/2)} - \frac{\vec{\phi}^{-}}{\sqrt{2}\sin(\gamma/2)},\end{aligned} \tag{27}$$

This is the linear combination of the normal vectors $\vec{\phi}^{+}$, $\vec{\phi}^{-}$ which gives the symmetrically orthonormalized vectors shown in Fig. 7. When $\vec{\phi}_2$ approaches $-\vec{\phi}_1$ (that is, $\cos\gamma = S$ approaches -1), $\vec{\phi}^{+}$

becomes smaller, thus requiring a larger coefficient in order to make $\vec{\phi}_j'$ turn further away from $\vec{\phi}_j$. In the limit $\vec{\phi}_2 = -\vec{\phi}_1$ an infinitely large coefficient would be required. A similar analysis holds for the limit in which $\vec{\phi}_2$ becomes parallel to $\vec{\phi}_1$. From a mathematical point of view the divergence appears because one of the eigenvalues of **g** vanishes, causing a divergence of one eigenvalue of $\mathbf{g}^{-½}$.

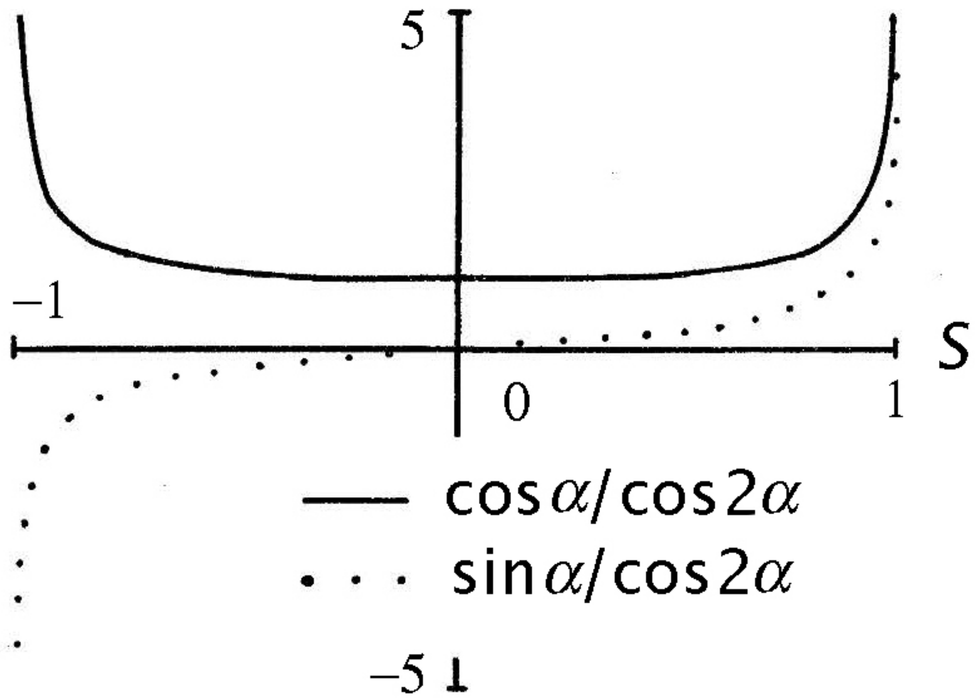

Fig. 6. Coefficients of the symmetrically orthonormalized vectors Eq. (18) as functions of the overlap $S = \cos\gamma$. The coefficients diverge for the linearly dependent case $S = 1$.

As $\vec{\phi}^+$ and $\vec{\phi}^-$ are always orthogonal, one may wonder to which particular transformation **O** in Eq. 7 they correspond. To that end we will first consider the matrix **S** which diagonalizes $\mathbf{g}^{-½}$. It may be easily verified that

$$\mathbf{S}^t \cdot \mathbf{g}^{-\frac{1}{2}} \cdot \mathbf{S} = \begin{pmatrix} \lambda^+ & 0 \\ 0 & \lambda^- \end{pmatrix}, \tag{28}$$

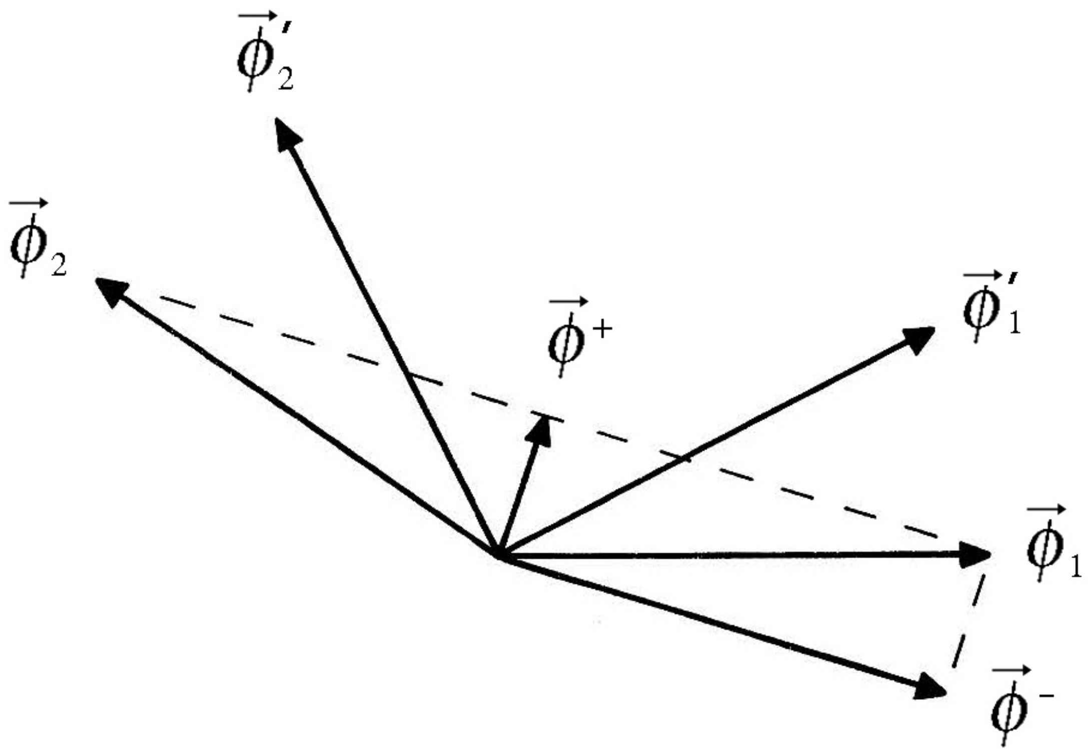

Fig. 7. Forming an appropriate linear combination of $\vec{\phi}^+$ and $\vec{\phi}^-$ in order to obtain a symmetrically orthonormalized pair $\vec{\phi}'_1$, $\vec{\phi}'_2$, requires giving more and more weight to $\vec{\phi}^+$ as $\vec{\phi}_2$ becomes antiparallel to $\vec{\phi}_1$.

where

$$\mathbf{S} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{29}$$

and the eigenvalues $\lambda^{\pm}$ are given by Eq. (17). It should be noticed that $-\mathbf{S}$ is also an admissible solution, but we need not consider it separately. Choosing $\mathbf{U} = \mathbf{S}$ in Eq. (7)

$$\mathbf{O} = \mathbf{g}^{-\frac{1}{2}} \bullet \mathbf{S}, \tag{30}$$

from Eq. (5) we obtain

$$\vec{\phi}_1' = \sqrt{2}\,\lambda^+ \vec{\phi}^+, \quad \vec{\phi}_2' = -\sqrt{2}\,\lambda^- \vec{\phi}^-, \tag{31}$$

which correspond to Löwdin's canonical orthonormalization[2]. Its outstanding peculiarity is that it gives equal weight to both primitive vectors in the construction of each of the new ones, that is

$$\begin{aligned} \left|\vec{\phi}_1 \bullet \vec{\phi}_1'\right| = \left|\vec{\phi}_2 \bullet \vec{\phi}_1'\right| = \left|\frac{\lambda^+}{\sqrt{2}}\right|, \\ \left|\vec{\phi}_1 \bullet \vec{\phi}_2'\right| = \left|\vec{\phi}_2 \bullet \vec{\phi}_2'\right| = \left|\frac{\lambda^-}{\sqrt{2}}\right|. \end{aligned} \tag{32}$$

Due to its consequences for the localization, this property should be compared with that of the symmetric orthonormalization:

$$\vec{\phi}_1 \bullet \vec{\phi}_1' = \vec{\phi}_2 \bullet \vec{\phi}_2', \quad \vec{\phi}_1 \bullet \vec{\phi}_2' = \vec{\phi}_2 \bullet \vec{\phi}_1'. \tag{33}$$

When $\vec{\phi}_1$ and $\vec{\phi}_2$ are atomic orbitals centered on different sites, the property Eq. (33) implies that the symmetric orthonormalization is the one which incorporates into each new orbital $\vec{\phi}_j'$ the maximun amount of $\vec{\phi}_j$ that is compatible with the preservation of the point symmetries, this being a feature of great importance in molecular and crystalline calculations. One thus obtains the most localized symmetry preserving orthonormal basis. On the other hand, the canonical orthonormalization, by giving equal weight to both primitive vectors in the construction of each of the new ones, is the most delocalized of all possible orthonormal basis. These features are illustrated in Figs. 8 and 9 for two hydrogen 1s orbitals centered $2a_0$ apart, where a_0 is the Bohr radius and S = 0.586453. For the sake of completeness we give in Fig. 10 the overlap S and the angles γ and α corresponding to different separations R of the aforementioned orbitals, where S has been calculated using the standard formulas[6].

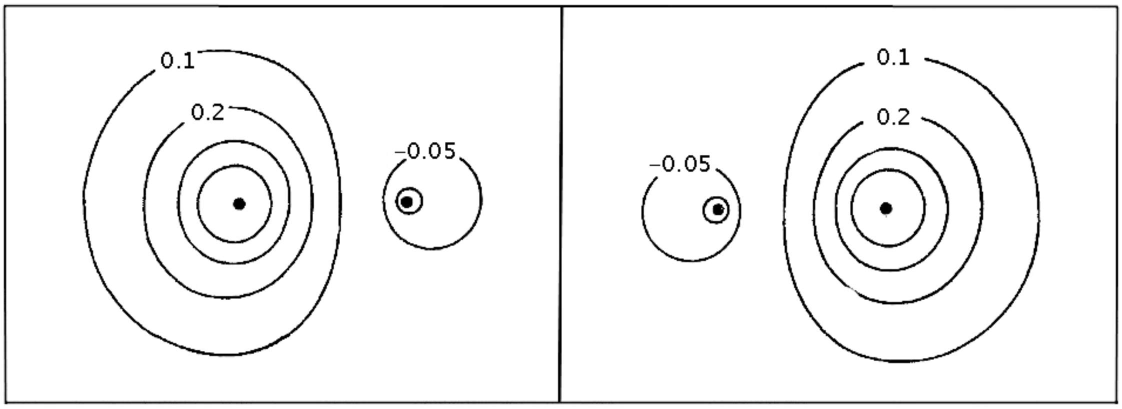

Fig. 8. Contours of the symmetrically orthonormalized orbitals corresponding to two hydrogen 1s orbitals whose nuclei (black dots) are $2a_0$ apart. The amplitude values are given in units of $a_0^{-3/2}$. The localization is the largest one compatible with the preservation of the reflection symmetry.

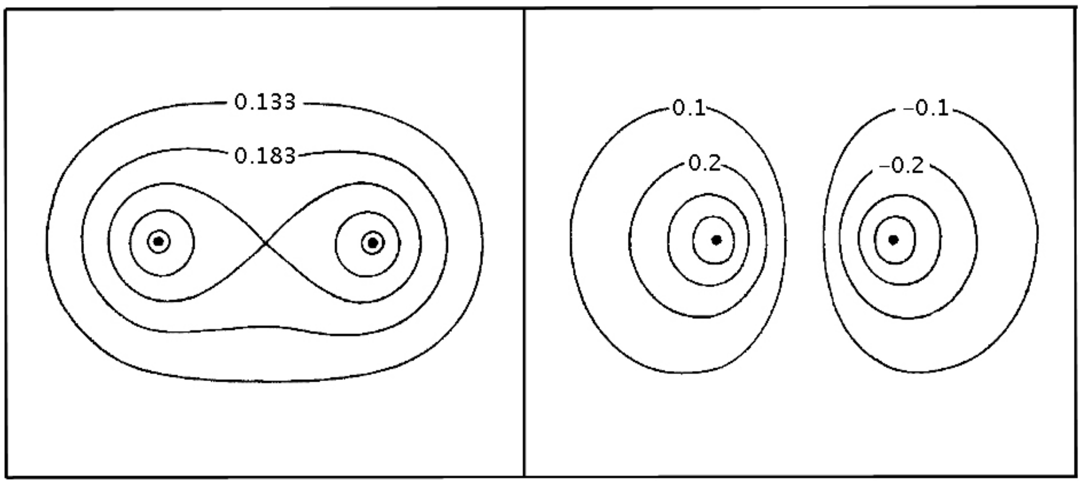

Fig. 9. Canonical orthonormalization of two hydrogen 1s orbitals $2a_0$ apart. All contours are equally spaced except the inner ones of the small lobes. Orbitals are as delocalized as is possible.

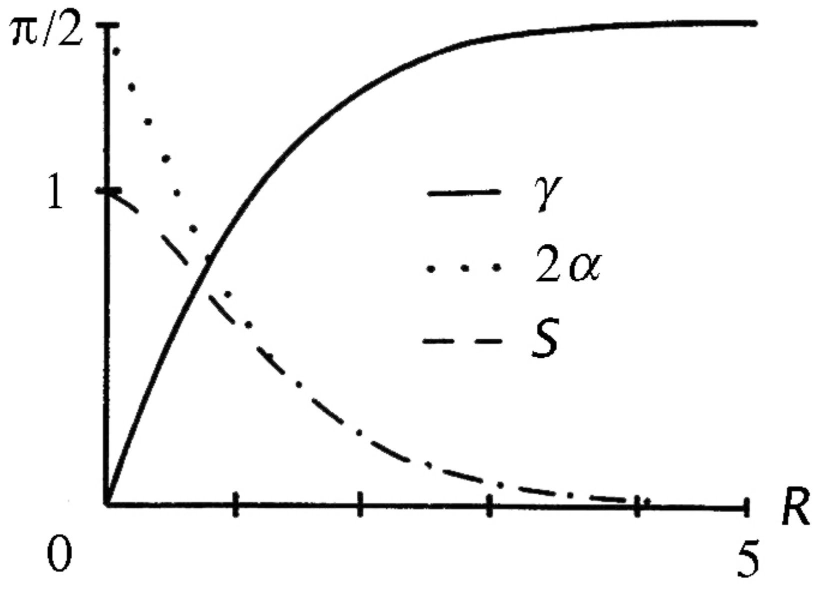

Fig. 10. Overlap $S = \cos\gamma$ and angles γ and $\alpha = \pi/4 - \gamma/2$ (symmetric orthonormalization) for two hydrogen 1s orbitals as a function of the distance R between centers. R is given in Ångstroms and the angles in radians.

3. CONCLUSIONS

When restricting oneself to the plane, it becomes easy to give simple geometric interpretations of the properties of the Gram-Schmidt, the symmetric and the canonical orthonormalization schemes. Taking the primitive nonorthogonal vectors to be of unit length, the quantum-mechanical overlap S is then the cosine of the angle between vectors.

It is thus shown that the only method which preserves orthogonal relationships between vectors (as illustrated by a mirror line) is the symmetric one. This solution is also the one that minimizes the "distance" between the old and the new vectors given by Eq. (21). It is also clearly seen why some coefficientes in the expansion of the latter in terms of the former blow up when approaching the limits $S = \pm 1$. This puts into sharper light the numerical problem of orthogonalizing quasi-linearly-dependent vectors.

The canonical orthonormalization turns out to be the one such that each new vector contains equal amounts of the old ones, as measured by the projection of the latter into the former. This corresponds to the maximum delocalization property of canonically orthonormalized orbitals, as opposed to the maximum localization of the symmetrically orthonormalized ones.

These easily grasped geometrical properties may be generalized to the n-dimensional case, thus providing a better understanding of the otherwise abstract orthonormalization problem.

REFERENCES

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, New York, Interscience (1953), pp. 50, 3.

2 P. O. Löwdin, *Arkiv. f. mat., astr.o.fysik*, **35A** (1947) 9; P. O. Löwdin, *J. Chem. Phys.* **18** (1950) 365; G. W. Pratt and S. F. Neustadter, *Phys. Rev.* **101** (1956) 1248; P. O. Löwdin, *Adv. in Phys.*, **5** (1956) 1; P. O. Löwdin, Quantum Chemistry Group for Research in Atomic, Molecular and Solid State Theory, *Preliminary Research Report Nº 213*, Uppsala (1968); P. O. Löwdin, *Adv. Quantum Chem.* **5** (1970) 185; J. G. Aiken and H. B. Jonassen, *J. Chem. Phys.* **62** (1975) 2745 and **63** (1975) 2772.

3 J. C. Slater and G. F. Koster, *Phys. Rev.* **94** (1954) 1498.

4 G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd. edition, New York, McGraw-Hill (1961) 420.

5 B. C. Carlson and J. M. Keller, *Phys. Rev.* **105** (1957) 102.

6 R. S. Mulliken, C. A. Rieke, O. Orloff and H. Orloff, *J. Chem. Phys.* **17** (1949) 1248.