

Energy derivatives in variational calculations using Slater-type and Gauss-type orbitals

A. Kolakowska and J. D. Talman

Department of Applied Mathematics and Department of Physics and Centre for Chemical Physics, University of Western Ontario, London, Ontario, Canada N6A 5B7

(Received 10 October 1995)

A method is described for computing the first and second derivatives with respect to the nonlinear parameters of the energy expectation value in atomic and molecular calculations involving Slater-type or Gauss-type orbitals. The approach is illustrated by computing the H atom 1s ground-state wave function in terms of Gaussian orbitals.

PACS number(s): 02.70.-c, 31.15.-p, 02.60.Pn

A great many calculations in atomic and molecular physics involve the expansion of single-particle wave functions in terms of Slater-type (STO) or Gaussian-type (GTO) orbitals. These are functions of the form

$$\psi_S(\mathbf{r}) = r^n e^{-\alpha r} Y_{lm}(\hat{r}), \quad (1)$$

$$\psi_G(\mathbf{r}) = r^{2n} e^{-\alpha r^2} Y_{lm}(\hat{r}), \quad (2)$$

respectively. The parameter α occurring in these functions is usually referred to as a nonlinear or scaling parameter. Most such calculations involve computation and diagonalization of the Hamiltonian matrix in a basis of STO's or GTO's to find approximate energy eigenvalues for the ground state or some of the low-lying excited states. These calculations, which are applying the so-called Rayleigh-Ritz method, are variational in nature, in that the coefficients for expanding the wave function in terms of the basis orbitals can be viewed as the solution of the problem of minimizing the expectation value of the Hamiltonian.

In such calculations, some attempt is also made to minimize the energy with respect to the nonlinear parameters as well. The problem of minimizing with respect to nonlinear parameters is, however, much more difficult than the linear eigenvalue problem presented by the variational calculation for the expansion coefficients. In a recent article Hill [1] has studied in detail the case in which the basis functions are scaled by a single nonlinear parameter. In this work the dependence of the convergence of the energy on increasing dimension of the basis set was analyzed; from this an optimum value of the parameter can be estimated.

There is a vast literature concerned with problems of nonlinear optimization [2], and it is safe to say that there is no single best way to approach these problems. However, it is surely true that the solution is greatly facilitated if the gradient vector of the derivatives of the object function with respect to the nonlinear parameters can be obtained explicitly, and even further facilitated if the Hessian matrix of the second derivatives is known.

It seems rather surprising that the problem of finding the derivatives of the eigenvalues of the Hamiltonian matrix with respect to the nonlinear parameters in STO's

and GTO's has not been addressed much earlier. This contrasts with the problem of computing the derivatives of molecular energies with respect to nuclear coordinates which has been exhaustively studied. We have not, however, found such a discussion in the literature, and the object of this note is to show that these derivatives can be obtained in a fairly simple way.

In this discussion, we consider only the variational problem for the single-particle Schrödinger equation. The much more interesting cases are many-particle calculations, in particular Hartree-Fock molecular-orbital linear combinations of atomic orbitals (MO-LCAO) calculations, in which the elements of the Hamiltonian matrix are quartic functionals of the basis orbitals. The approach described here can also be used in such calculations, with some additional complications, and has indeed been used successfully in configuration-interaction (CI) calculations for He-like ions using the Dirac Hamiltonian [3].

As an example, we take the problem of expanding the hydrogenic ground-state wave function in GTO's. The variational problem is then to minimize the variational functional

$$I[\psi] = \int_0^\infty \left[[\psi'(r)]^2 - \frac{2}{r} [\psi(r)]^2 \right] r^2 dr \quad (3)$$

subject to the condition

$$\int_0^\infty \psi(r)^2 r^2 dr = 1. \quad (4)$$

The wave function $\psi(r)$ is to be approximated in the form

$$\psi(r) = \sum_{i=1}^N c_i e^{-\alpha_i r^2} \quad (5)$$

where the c_i and α_i are variational parameters. Solutions for this problem were given many years ago by Huzinaga [4] for $N = 1, 2, \dots, 10$.

The variational problem can be expressed in general in the form of the generalized eigenvalue problem

$$H\mathbf{x}_i = \lambda_i B\mathbf{x}_i \quad (6)$$

where H is the Hamiltonian matrix with elements

$$h_{ij} = \langle \psi_i | \hat{H} | \psi_j \rangle \quad (7)$$

and B is the overlap matrix with elements

$$b_{ij} = \langle \psi_i | \psi_j \rangle. \quad (8)$$

We view H and B as being functions of a set of N parameters α_m .

Equation (6) has a set of N eigenvalues and eigenvectors; the latter satisfy the orthogonality relation

$$(\mathbf{x}_i, B\mathbf{x}_j) = \delta_{ij}. \quad (9)$$

(Matrix elements in the physical Hilbert space will be denoted by the Dirac bracket notation, and in the N -dimensional subspace by the standard linear algebra inner product notation.) Differentiating Eq. (6) with respect to a parameter α_m gives the identity

$$(H - \lambda_i B) \frac{\partial \mathbf{x}_i}{\partial \alpha_m} + \left(\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right) \mathbf{x}_i = \frac{\partial \lambda_i}{\partial \alpha_m} B\mathbf{x}_i. \quad (10)$$

Taking the inner product of Eq. (10) with respect to \mathbf{x}_i

gives the result

$$\frac{\partial \lambda_i}{\partial \alpha_m} = \left(\mathbf{x}_i, \left[\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right] \mathbf{x}_i \right). \quad (11)$$

This relation is the same as first order perturbation theory. It is also essentially the same as the Hellmann-Feynman relation [5] although the context is different; in the Hellmann-Feynman relation the matrix is an operator in Hilbert space depending on a physical parameter such as an internuclear spacing or an electric field strength.

Taking the inner product with \mathbf{x}_j , $j \neq i$, gives

$$\left(\mathbf{x}_j, B \frac{\partial \mathbf{x}_i}{\partial \alpha_m} \right) = \frac{1}{\lambda_i - \lambda_j} \left(\mathbf{x}_j, \left[\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right] \mathbf{x}_i \right). \quad (12)$$

Again, this result is familiar from perturbation theory. If Eq. (9) is differentiated with respect to α_m in the case $i = j$, it is found that

$$\left(\mathbf{x}_i, B \frac{\partial \mathbf{x}_i}{\partial \alpha_m} \right) = -\frac{1}{2} \left(\mathbf{x}_i, \frac{\partial B}{\partial \alpha_m} \mathbf{x}_i \right). \quad (13)$$

These results can be combined to give the expansion

$$\begin{aligned} \frac{\partial \mathbf{x}_i}{\partial \alpha_m} &= \sum_{j=1}^N \left(\mathbf{x}_j, B \frac{\partial \mathbf{x}_i}{\partial \alpha_m} \right) \mathbf{x}_j \\ &= -\frac{1}{2} \left(\mathbf{x}_i, \frac{\partial B}{\partial \alpha_m} \mathbf{x}_i \right) \mathbf{x}_i + \sum_{j=1, j \neq i}^N \frac{1}{\lambda_i - \lambda_j} \left(\mathbf{x}_j, \left[\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right] \mathbf{x}_i \right) \mathbf{x}_j. \end{aligned} \quad (14)$$

Differentiating Eq. (11) with respect to α_n gives

$$\frac{\partial^2 \lambda_i}{\partial \alpha_n \partial \alpha_m} = \left(\mathbf{x}_i, \left[\frac{\partial^2 H}{\partial \alpha_n \partial \alpha_m} - \lambda_i \frac{\partial^2 B}{\partial \alpha_n \partial \alpha_m} \right] \mathbf{x}_i \right) + 2 \left(\frac{\partial \mathbf{x}_i}{\partial \alpha_n}, \left[\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right] \mathbf{x}_i \right) - \frac{\partial \lambda_i}{\partial \alpha_n} \left(\mathbf{x}_i, \frac{\partial B}{\partial \alpha_m} \mathbf{x}_i \right). \quad (15)$$

This can be reexpressed using Eq. (14) as

$$\begin{aligned} \frac{\partial^2 \lambda_i}{\partial \alpha_n \partial \alpha_m} &= \left(\mathbf{x}_i, \left[\frac{\partial^2 H}{\partial \alpha_n \partial \alpha_m} - \lambda_i \frac{\partial^2 B}{\partial \alpha_n \partial \alpha_m} \right] \mathbf{x}_i \right) \\ &+ 2 \sum_{j=1, j \neq i}^N \frac{1}{\lambda_i - \lambda_j} \left(\mathbf{x}_j, \left[\frac{\partial H}{\partial \alpha_m} - \lambda_i \frac{\partial B}{\partial \alpha_m} \right] \mathbf{x}_i \right) \left(\mathbf{x}_j, \left[\frac{\partial H}{\partial \alpha_n} - \lambda_i \frac{\partial B}{\partial \alpha_n} \right] \mathbf{x}_i \right) \\ &- \frac{\partial \lambda_i}{\partial \alpha_n} \left(\mathbf{x}_i, \frac{\partial B}{\partial \alpha_m} \mathbf{x}_i \right) - \frac{\partial \lambda_i}{\partial \alpha_m} \left(\mathbf{x}_i, \frac{\partial B}{\partial \alpha_n} \mathbf{x}_i \right). \end{aligned} \quad (16)$$

In the example being considered in this article, each orbital ψ_i is parametrized by a single parameter α_i . The derivatives of a matrix element are therefore given by

$$\begin{aligned} \frac{\partial h_{ij}}{\partial \alpha_m} &= \left\langle \frac{\partial \psi_i}{\partial \alpha_m} | \hat{H} | \psi_j \right\rangle + \left\langle \psi_i | \hat{H} | \frac{\partial \psi_j}{\partial \alpha_m} \right\rangle \\ &= \left\langle \frac{\partial \psi_m}{\partial \alpha_m} | \hat{H} | \psi_j \right\rangle \delta_{im} + \left\langle \psi_i | \hat{H} | \frac{\partial \psi_m}{\partial \alpha_m} \right\rangle \delta_{jm}. \end{aligned} \quad (17)$$

Similarly, the second derivatives are given by

$$\begin{aligned} \frac{\partial^2 h_{ij}}{\partial \alpha_m \partial \alpha_n} &= \left\langle \frac{\partial^2 \psi_m}{\partial \alpha_m^2} | \hat{H} | \psi_j \right\rangle \delta_{im} \delta_{mn} \\ &+ \left\langle \psi_i | \hat{H} | \frac{\partial^2 \psi_m}{\partial \alpha_m^2} \right\rangle \delta_{jn} \delta_{mn} \\ &+ \left\langle \frac{\partial \psi_n}{\partial \alpha_n} | \hat{H} | \frac{\partial \psi_m}{\partial \alpha_m} \right\rangle (\delta_{im} \delta_{jm} + \delta_{jm} \delta_{in}). \end{aligned} \quad (18)$$

Because of the nature of the basis functions, the derivatives of the matrix elements can be calculated in the same way as the original matrix elements simply by increasing the index n by 1 or 2.

If the energy functional were quadratic in the nonlinear parameters, the energy minimum would be given by the Newton formula

$$A(\delta\alpha) = -\mathbf{G} \quad (19)$$

where A is the Hessian matrix, $\delta\alpha$ is the required increment in the nonlinear parameters, and \mathbf{G} is the energy gradient. In practice, however, the functional is much more complicated and in most regions of the parameter space the Hessian matrix appears not to be positive definite. This is the complicating feature of nonlinear optimization problems referred to above. We have adopted the following strategy to deal with this problem. At a particular parameter point α , the Hessian matrix is transformed to principal axes, and the components of the gradient in these coordinates are computed. The Newton equation is then solved only in the directions in which the eigenvalues of A are positive, i.e., in the subspace in which A is positive definite. This is readily done by computing a "pseudoinverse" for A as

$$A_{ps}^{-1} = \sum_{\mu_i > 0} \frac{1}{\mu_i} \mathbf{u}_i \mathbf{u}_i^T \quad (20)$$

where μ_i and \mathbf{u}_i are the eigenvalues and eigenvectors of the Hessian matrix. The solution of Eq. (19), which is given by

$$\delta\alpha = - \sum_{\mu_i > 0} \frac{1}{\mu_i} (\mathbf{u}_i, \mathbf{G}) \mathbf{u}_i, \quad (21)$$

then gives a vector in the direction of the maximum rate of decrease of the energy. If the parameter vector is close to a minimum, A is positive definite, and iteration of Eq. (19) gives a quadratically convergent search for the minimum.

Because of the nonlinearity, success of this strategy is not guaranteed. In particular, the Hessian matrix may be close to being singular, in which case the parameter increments may be very large. This can happen, for example, at a step in which an eigenvalue of the Hessian matrix changes from negative to positive. This problem has been circumvented by limiting the maximum $|\delta\alpha_i/\alpha_i|$ to be 1/2 by scaling down $\delta\alpha$ by the appropriate factor if it exceeds 1/2. In addition, if at a particular step the energy increases, the calculation returns to the previous parameter set, and the parameter increments are halved. With this strategy, the energy in the calculations converged downwards, the parameters converged to a region in which the Hessian matrix was positive definite, and the convergence became quadratic. In the calculations reported here, the final results were obtained in the order of ten iterations.

Table I shows the results of applying the minimization procedure to the problem of solving the H atom ground state in terms of GTO's. The results are the same as those of Huzinaga [4], apart from small differences in the nonlinear parameters which probably arise because of a less efficient search procedure in Huzinaga's calculation. Huzinaga refers to the possibility of multiple local minima. We have seen no evidence for this in the present calculations; in each case, the final results appear to be independent of the initial parameter values.

In Table II we show the results of the energy minimization procedure for trial wave functions of the form

$$\psi(r) = P_N(x^2)e^{-\alpha r^2} \quad (22)$$

where P_N is a polynomial of degree N and α is a single scaling parameter. (This is the basis considered by Hill [1].) The derivatives with respect to α can be obtained

TABLE I. H atom ground-state energy (in Ry) computed in a basis of GTO's. The exact value is 1. Also given is $\psi(0)$, the value of the approximate solution at the nucleus; the exact value is 2.

N	α_i	c_i	e	$\psi(0)$
1	0.282 94	0.980 14	-0.848 826 363	0.9801
2	0.201 53 1.332 50	0.624 07 0.859 81	-0.971 625 433	1.4836
3	0.151 38 0.681 29 4.500 36	0.397 12 0.772 77 0.550 16	-0.993 958 505	1.7200
4	0.121 95 0.444 54 1.962 26 13.010 70	0.261 26 0.657 91 0.577 90 0.340 65	-0.998 556 811	1.8377
5	0.103 07 0.327 23 1.164 66 5.123 57 34.061 34	0.177 15 0.549 87 0.571 87 0.387 36 0.214 63	-0.999 619 664	1.9009
6	0.089 97 0.258 59 0.800 01 2.833 05 12.452 42 82.921 77	0.123 17 0.455 91 0.547 80 0.415 46 0.255 93 0.138 69	-0.999 891 141	1.9370
7	0.080 32 0.213 98 0.598 54 1.841 25 6.509 59 28.605 31 190.687 83	0.087 44 0.376 82 0.513 87 0.429 88 0.287 95 0.170 87 0.091 81	-0.999 966 596	1.9586

TABLE II. H atom ground-state energy (in Ry) computed in the basis given in Eq. (23).

N	α	e
0	0.282 94	-0.848 826 363
2	0.401 72	-0.945 279 354
4	0.487 13	-0.973 215 341
6	0.559 44	-0.984 582 831
8	0.624 59	-0.990 178 500
10	0.685 07	-0.993 292 260
12	0.742 18	-0.995 180 257

immediately from the more general partial derivatives given in Eqs. (11) and (16). These results were obtained generally in fewer than five iterations. The values obtained for α agree with those given by Hill.

It is interesting to observe, as noted by Hill, that the energy and the nonlinear parameter do not change when N increases by 1 from an even value to an odd value; therefore the results for odd values of N are not given. The reason for this is as follows. If a normalized wave function ψ depends on a parameter α , it can be seen that

$$\frac{de}{d\alpha} = 2 \left\langle \frac{d\psi}{d\alpha} \left| \hat{H} \right| \psi \right\rangle, \quad (23)$$

$$\left\langle \frac{d\psi}{d\alpha} \left| \psi \right\rangle = 0. \quad (24)$$

Equations (7) and (8) show that at a minimum in α the matrices H and B have zero matrix elements be-

tween $\partial\psi/\partial\alpha$ and the variational wave function, and that adding $\partial\psi/\partial\alpha$ to the basis cannot affect the lowest eigenvalue. It is not hard to see that $r^{2N+2}e^{-\alpha r^2}$ is a linear combination of $\partial\psi/\partial\alpha$ and $r^{2n}e^{-\alpha r^2}$, $n = 0, \dots, N$. Therefore adding it to the basis cannot change the result since it is a linear combination of a function that cannot change the result and functions that are already in the basis.

It is also interesting to compare entries in the two tables. For a particular N , each entry depends on $2N$ parameters. It is seen that far greater accuracy, at least in the energy eigenvalue, is obtained from varying the nonlinear parameters than from increasing the dimension of a polynomial basis. This observation appears to be a well-known "folk theorem" in quantum chemistry.

This work has been supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

[1] R.N. Hill, Phys. Rev. A **51**, 4433 (1995).

[2] See, for example, Chap. 9 in D. Kahaner, C. Moler, and S. Nash, *Numerical Methods and Software* (Prentice-Hall, Englewood Cliffs, NJ, 1989) or Chap. 10 in W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992) for discussion and references.

[3] A.Kolakowska, J.D. Talman, and K. Aashamar, Phys. Rev. **53**, 168 (1996).

[4] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).

[5] H. Hellmann, *Einfuehren in die Quantenchemie* (Franz Deuticke, Leipzig, 1939); R.P. Feynman, Phys. Rev. **56**, 340 (1939).