Chem 3502/4502 **Physical Chemistry II (Quantum Mechanics)** 3 Credits

Spring Semester 2006

Christopher J. Cramer

Lecture 32, April 14, 2006

(Some material in this lecture has been adapted from Cramer, C. J. *Essentials of Computational Chemistry*, Wiley, Chichester: 2002; pp. 40-45; 181-182.)

Solved Homework

In cartesian coordinates (Å), our water geometry placed oxygen at position (0.000, 0.000, 0.116), one hydrogen at (0.000, 0.751, –0.465) and the other hydrogen at (0.000, –0.751, –0.465). To work in a.u., we need to convert these coordinates to bohr using 1 bohr = 0.529 Å. In that case, the coordinates for water are oxygen at position (0.000, 0.000, 0.219), one hydrogen at (0.000, 1.420, –0.879) and the other hydrogen at (0.000, –1.420, –0.879).

From Mulliken population analysis, we had charges of q = –0.372 on oxygen and 0.186 on each hydrogen

Thus, we have

$$\langle x \rangle = \sum_{i=1}^{3} x_i q_i = (0.0 \bullet -0.372) + (0.0 \bullet 0.186) + (0.0 \bullet 0.186) = 0.00$$

$$\langle y \rangle = \sum_{i=1}^{3} y_i q_i = (0.0 \bullet -0.372) + (1.420 \bullet 0.186) + (-1.420 \bullet 0.186) = 0.00$$

$$\langle z \rangle = \sum_{i=1}^{3} z_i q_i = (0.219 \bullet -0.372) + (-1.420 \bullet 0.186) + (-1.420 \bullet 0.186) = -0.61$$

Recalling

$$\langle \mu \rangle = \sqrt{\langle x \rangle^2 + \langle y \rangle^2 + \langle z \rangle^2}$$

we see that <μ> = 0.61 a.u. From lecture 14, we know that 1 a.u. of dipole moment is 2.542 debye. So, the dipole moment from the Mulliken charges in units of debye is 2.542 x 0.61 = 1.55 D. This result is not too far from the experimental value of 1.8 D, suggesting that Mulliken charges have physically realistic magnitudes, at least based on our water molecule example.

One might go on, then, to make chemical interpretations based on variations in Mulliken charges. For instance, if we compute methanol and acetic acid to have oxygen-bound proton Mulliken charges of 0.125 and 0.441, respectively, we might interpret this to imply that methanol will be a weaker acid than water and acetic acid a stronger one. Such a simple predictive tool might be quite useful in judging relative acidity for less obvious cases.

Other Computed Properties—Spatial Extent

The dipole moment (and higher moments) describe directional polarizations in the charge distribution of a molecule. Such polarizations are very important in understanding how molecules interact with external electric fields, to include playing a role in absorption spectroscopy, as we've previously seen.

The dipole involves the expectation values of cartesian coordinates x, y, and z to first power, so it reflects an oriented distribution, positive vs. negative. The quadrupole moment, on the other hand, involves contributions from terms including x^2, y^2, and z^2, and these measure displacement from a center along cartesian axes rather than bias from one side to another.

Note that if we take the sum of the three cartesian displacements squared we have

$$\left\langle r^2 \right\rangle = \left\langle x^2 \right\rangle + \left\langle y^2 \right\rangle + \left\langle z^2 \right\rangle \tag{32-1}$$

If we evaluate $\langle r^2 \rangle$ only for the electronic part of the wave function, this defines what is called the "electronic spatial extent". That is, it is a measure of how "big" the molecule is, since it reports how far out the electronic density extends with significant probability. (Remember that the wave function formally goes out to infinity, but it becomes so close to zero at some point that it is not really meaningful to worry about the "tails".)

For water in our STO-3G calculation, $\langle r^2 \rangle$ = 17.74 a.u. The square root of this expectation value has atomic units of distance (bohr) and is a bit more than 4. That corresponds to the electron density being roughly equally distributed inside and outside a shell of 2.1 Å centered on the molecular origin. Such a distance is roughly consistent with the van der Waals radius of oxygen.

An interesting comparison can be made with the water radical anion. You may recall we considered this species in the homework to lecture 31. For this negatively charged molecule, $\langle r^2 \rangle$ = 29.75 a.u., a dramatic increase. This large change tells us that the extra electron is held rather weakly to the water molecule. Loosely held electrons localize at large distances from the nuclei and thus contribute to large values for $\langle r^2 \rangle$.

So, $\langle r^2 \rangle$ can be considered to be a simple, single, quantum descriptor associated with molecular size.

Other Computed Properties—Ionization Potentials and Electron Affinities

We have previously noted that the total energy for a HF Slater determinant may be computed as

$$E_{\mathrm{HF}} = \left\langle \Psi^{\mathrm{HF}} \middle| H \middle| \Psi^{\mathrm{HF}} \right\rangle = 2 \sum_{i}^{\text{occupied}} \varepsilon_i - \sum_{i,j}^{\text{occupied}} \left(2J_{ij} - K_{ij} \right) \qquad (32\text{-}2)$$

where ε_i is the energy of a single electron in orbital i, and J and K are the usual Coulomb and exchange integrals, which would otherwise be double counted in the first sum when computing the total energy expression.

Consider now the process of removing one electron from the molecule. The minimum energy for that removal is called the ionization potential (IP). By its definition, this makes the ionization potential the negative of the binding affinity for the least tightly held electron (i.e., one of the electrons in the HOMO).

A Dutch economist (of all things...) named Koopmans was the first to prove that if we use only the *occupied* molecular orbitals of the neutral molecule as a basis set from which to form molecular orbitals for the radical cation, then the *optimal* MOs for the radical cation put exactly two electrons in the identical MOs as for the neutral and one electron in the identical HOMO as for the neutral. In other words, in the absence of other basis functions with which to work, the HF orbitals do *not* relax following ionization.

In that case, if we evaluate the energy change as the difference in energy between the neutral and the radical cation, we will find it to be precisely $\varepsilon_{\mathrm{HOMO}}$ for the neutral. This result is known as Koopmans' theorem. Note the simplicity implicit in Koopmans' theorem: one need not do any kind of calculation on the radical cation itself, one need only inspect $\varepsilon_{\mathrm{HOMO}}$ for the neutral to predict the ionization potential.

Koopmans also showed that the same reasoning applies to ionization of more tightly bound electrons (i.e., those in lower energy orbitals): as long as we don't permit electrons in higher energy orbitals to drop down into the low-energy "hole" that is created, then the difference in energy between the radical cation and the neutral can be estimated as the orbital energy of the electron being ionized in the neutral.

Finally, Koopmans showed that if an electron is *attached* to a molecule to make a radical anion, then the electron affinity (EA, which is the negative of the binding energy for the new electron) should be the negative of the LUMO orbital energy from the HF calculation on the neutral molecule.

Koopmans' theorem equates the energy of the HOMO with the negative of the IP. However, this approximation ignores the effect of electronic relaxation in the ionized product, i.e., the degree to which the remaining electrons redistribute themselves

following the detachment of one from the HOMO (remember, as noted above, Koopmans result derives from forcing the radical cation to build a wave function using only the occupied MOs of the neutral as a basis set; if we give it additional flexibility, it will build other, better MOs after ionization). If we instead calculate the IP as the difference in HF energies for the closed-shell neutral and the open-shell product, we obtain the so-called ΔSCF IP

$$\mathrm{IP}_{\Delta\mathrm{SCF}} = E_{\mathrm{HF}}\left(\mathrm{A}^{+\bullet}\right) - E_{\mathrm{HF}}(\mathrm{A}) \tag{32-3}$$

where orbital relaxation *is* included. Including relaxation results in a smaller predicted IP, since relaxation lowers the energy of the cation radical relative to the neutral (the HOMO energy used in Koopmans' theorem derives from orbitals already fully relaxed for the neutral). Note, however, that the neutral species has one more electron than the radical cation, and thus there will be a larger error from ignoring electron correlation. By ignoring these effects through the use of HF theory, we destabilize the neutral more than the radical cation, and too small an IP would be expected in any case. Thus, Koopmans' theorem benefits from a cancellation of errors: the orbital relaxation and the electron correlation effects offset one another. In practice, the cancellation can be remarkably good; Koopmans' theorem IPs are often within 0.3 eV or so of experiment provided basis sets of good quality are used in the HF calculation. However, this favorable cancellation begins to break down if IPs are computed for orbitals other then the HOMO. As more tightly held electrons are ionized, particularly core electrons, the relaxation effects are much larger than the correlation effects, and Koopmans' approximation should not be used as it leads to significant errors (instead, a ΔSCF should be computed where possible).

As noted above, Koopmans' theorem is formally applicable to electron affinities (EAs) as well, i.e., the EA can be taken to be the negative of the orbital energy of the lowest unoccupied (virtual) orbital. Here, however, relaxation effects and correlation effects both favor the radical anion, so rather than canceling, the errors are additive, and Koopmans' theorem estimates will almost always underestimate the EA. It is thus generally a better idea to compute EAs from a ΔSCF approach whenever possible.

Geometry Optimization

One of the key motivations in early quantum modeling was the development of a protocol that would permit facile optimization of molecular geometries. While the energy of an arbitrary structure can be interesting, real molecules vibrate thermally about their equilibrium structures, so finding minimum energy structures is key to describing equilibrium constants, comparing to experiment, etc.

In its essence, geometry optimization is a problem in applied mathematics. How does one find a minimum in an arbitrary function of many variables? [Indeed, we have already discussed this exact problem in the context of variational optimization of

parametric wave functions. In the case of geometry optimization, we may be working with far fewer variables (the geometric degrees of freedom) and may moreover have analytic expressions for how the energy depends on the variables. The mathematical approach can thus be quite different.] As the problem is general, so too many of the details presented below will be general to any energy functional, although we care most about quantum mechanics.

For pedagogical purposes, let us begin by considering a case where we do not know how our energy depends on the geometric coordinates of our molecule. To optimize the geometry, all we can do is keep trying different geometries until we are reasonably sure that we have found the one with the lowest possible energy. How can one most efficiently survey different geometries?

It is easiest to proceed by considering a one dimensional case, i.e., a diatomic with only the bond length as a geometric degree of freedom. One selects a bond length, and computes the energy. One then changes the bond length, let us say by shortening it 0.2 Å, and again computes the energy. If the energy goes down, we want to continue moving the bond length in that direction, and we should take another step (which need not necessarily be of the same length). If the energy goes up, on the other hand, we are moving in the wrong direction, and we should take a step in the opposite direction. Ultimately, the process will provide three adjacent points where the one in the center is lower in energy than the other two. Three noncollinear points uniquely define a parabola, and in this case the parabola must have a minimum (since the central point was lower in energy than the other two). We next calculate the energy for the bond length corresponding to the parabolic minimum (the degree to which the computed energy agrees with that from the parabolic equation will be an indication of how nearly harmonic the local bond stretching coordinate is). We again step left and right on the bond stretching coordinate, this time with smaller steps (perhaps an order of magnitude smaller) and repeat the parabolic fitting process. This procedure can be repeated until we are satisfied that our step size falls below some arbitrary threshold we have established as defining convergence of the geometry. Note that one can certainly envision variations on this theme. One could use more than three points in order to fit to higher order polynomial equations, step sizes could be adjusted based on knowledge of previous points, etc.

In the multidimensional case, the simplest generalization of this procedure is to carry out the process iteratively. Thus, for LiOH, for example, we might first find a parabolic minimum for the OH bond, then for the LiO bond, then for the LiOH bond angle (in each case holding the other two degrees of freedom fixed), and then repeat the process to convergence. Of course, if there is strong coupling between the various degrees of freedom, this process will converge rather slowly.

What we really want to do at any given point in the multidimensional case is move not in the direction of a *single* coordinate, but rather in the direction of the greatest downward slope in the energy with respect to *all* coordinates. This direction is the opposite of the gradient vector, **g**, which is defined as

$$\mathbf{g}(\mathbf{q}) = \begin{bmatrix} \frac{\partial E}{\partial q_1} \\ \frac{\partial E}{\partial q_2} \\ \frac{\partial E}{\partial q_3} \\ \vdots \\ \frac{\partial E}{\partial q_n} \end{bmatrix} \tag{32-4}$$

where E is the energy (the expectation value of the Hamiltonian) and $\mathbf{q}$ is an n-dimensional coordinate vector ($n = 3N–6$ where N is the number of atoms if we are working in internal coordinates, $n = 3N$ if we are working in cartesian coordinates, etc.) When working in internal coordinates, one simply uses the chain rule to compute changes in energy with respect to, say, a bond length, in terms of changes in energy for x, y, and z displacements of the two atoms in the bond). If we cannot compute the partial derivatives that make up $\mathbf{g}$ analytically, we can do so numerically. However, that numerical evaluation requires at least one additional energy calculation for each degree of freedom. Thus, we would increase (or decrease) every degree of freedom by some step size, compute the slope of the resulting line derived from the energy of our initial structure and the perturbed structure, and use this slope as an estimate for the partial derivative. Such a "forward difference" estimation is typically not very accurate, and it would be better to take an additional point in the opposite direction for each degree of freedom, and then compute the "central difference" slope from the corresponding parabola. It should be obvious that, as the number of degrees of freedom increases, it can be particularly valuable to have an energy function for which the first derivative is known *analytically*. Happily, it turns out that the Hartree-Fock energy computed using *gaussian* basis functions admits to analytic derivatives—the time to compute the derivatives is about the same order of magnitude as the time to compute the energy times the number of degrees of freedom.

With $\mathbf{g}$ in hand, we can proceed in a fashion analogous to the one-dimensional case outlined above. We step along the direction defined by $–\mathbf{g}$ until we locate a minimum in the energy for this process; since we are taking points in a linear fashion, this movement is called a "line search" (even though we may identify our minimum by fitting our points to a polynomial curve). Then, we recompute $\mathbf{g}$ at the located minimum and repeat the process. Our new search direction is necessarily orthogonal to our last one, since we minimized E in the last direction. This particular feature of a steepest descent curve can lead to *very* slow convergence in unfavorable cases.

A more robust method is the so-called Newton-Raphson procedure. If we express the full energy as a multidimensional Taylor expansion in arbitrary coordinates (in matrix notation) and truncate at second order, we have

$$E\left(\mathbf{q}^{(k+1)}\right) = E\left(\mathbf{q}^{(k)}\right) + \left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right)\mathbf{g}^{(k)} + \frac{1}{2}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right)^{+}\mathbf{H}^{(k)}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right) \tag{32-5}$$

where the reference point is $\mathbf{q}^{(k)}$, $\mathbf{g}^{(k)}$ is the gradient vector for the reference point as defined by eq. 32-4, and $\mathbf{H}^{(k)}$ is the "Hessian" matrix for the reference point, whose elements are defined by

$$H_{ij}^{(k)} = \left.\frac{\partial^2 E}{\partial q_i \partial q_j}\right|_{\mathbf{q}=\mathbf{q}^{(k)}} \tag{32-6}$$

If we differentiate eq. 32-5 term by term with respect to the i'th coordinate of $\mathbf{q}^{(k+1)}$, noting that no term associated with point k has any dependence on a coordinate of point k+1 (and hence the relevant partial derivative will be 0), we obtain

$$\begin{aligned}\frac{\partial E\left(\mathbf{q}^{(k+1)}\right)}{\partial q_i^{k+1}} = \frac{\partial \mathbf{q}^{(k+1)}}{\partial q_i^{k+1}}\mathbf{g}^{(k)} + \frac{1}{2}\frac{\partial \mathbf{q}^{(k+1)^{+}}}{\partial q_i^{k+1}}\mathbf{H}^{(k)}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right) \\ + \frac{1}{2}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right)^{+}\mathbf{H}^{(k)}\frac{\partial \mathbf{q}^{(k+1)}}{\partial q_i^{k+1}}\end{aligned} \tag{32-7}$$

The l.h.s. of eq. 32-7 is the i'th element of the vector $\mathbf{g}^{(k+1)}$. On the r.h.s. of eq. 32-7, since the partial derivative of $\mathbf{q}$ with respect to its i'th coordinate is simply the unit vector in the i'th coordinate direction, the various matrix multiplications simply produce the i'th element of the multiplied vectors. Because mixed partial derivative values are independent of the order of differentiation, the Hessian matrix is Hermitian, and we may simplify eq. 32-7 as

$$g_i^{(k+1)} = g_i^{(k)} + \left[\mathbf{H}^{(k)}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right)\right]_i \tag{32-8}$$

where the notation $[]_i$ indicates the i'th element of the column matrix. The condition for a stationary point is that the l.h.s. of eq. 32-8 be 0 for *all* coordinates, or

$$\mathbf{0} = \mathbf{g}^{(k)} + \mathbf{H}^{(k)}\left(\mathbf{q}^{(k+1)} - \mathbf{q}^{(k)}\right) \tag{32-9}$$

which may be rearranged to

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} - \left(\mathbf{H}^{(k)}\right)^{-1}\mathbf{g}^{(k)} \tag{32-10}$$

This equation provides a prescription for the location of stationary points. In principle, starting from an arbitrary structure having coordinates $\mathbf{q}^{(k)}$, one would compute its

gradient vector **g** and its Hessian matrix **H**, and then select a new geometry $\mathbf{q}^{(k+1)}$ according to eq. 32-10. Eq. 32-9 shows that the gradient vector for this new structure will be the **0** vector, so we will have a stationary point.

Recall, however, that our derivation involved a truncation of the full Taylor expansion at second order. Thus, eq. 32-9 is only approximate, and $\mathbf{g}^{(k+1)}$ will not necessarily be **0**. However, it will probably be smaller than $\mathbf{g}^{(k)}$, so we can repeat the whole process to pick a point k+2. After a sufficient number of iterations, the gradient will hopefully become so small that structures $k+n$ and $k+n+1$ differ by a chemically insignificant amount, and we declare our geometry to be converged.

There are a few points with respect to this procedure that merit discussion. First, there is the Hessian matrix. With n^2 elements, where n is the number of coordinates in the molecular geometry vector, it can grow somewhat expensive to construct this matrix at every step even for functions that have analytical expressions for their second derivatives (as Hartree-Fock energies do!). Moreover, the matrix must be *inverted* at every step, and matrix inversion formally scales as n^3, where n is the dimensionality of the matrix. Thus, for purposes of efficiency (or in cases where analytic second derivatives are simply not available) approximate Hessian matrices are often used in the optimization process—after all, the truncation of the Taylor expansion renders the Newton-Raphson method *intrinsically* approximate. As an optimization progresses, second derivatives can be estimated reasonably well from finite differences in the analytic first derivatives over the last few steps. For the first step, however, this is not an option, and one typically either accepts the cost of computing an initial Hessian analytically for the level of theory in use, or one employs a Hessian obtained at a less expensive level of theory, when such levels are available. To speed up slowly convergent optimizations, it is often helpful to compute an analytic Hessian every few steps and replace the approximate one in use up to that point. For *really* tricky cases (e.g., where the PES is fairly flat in many directions) one is occasionally forced to compute an analytic Hessian for *every* step.

In any case, let us see how a good optimization algorithm does with our starting geometry for water. At the end of each SCF, it computes the energy, then the gradient, then takes a step and repeats the process until the gradient vector is reduced to sufficiently near zero. The results are shown below; it takes only 5 steps to reach the optimal geometry.

```
SCF Done:  E(RHF) =  -74.9617540762     A.U. after    7 cycles
SCF Done:  E(RHF) =  -74.9658121109     A.U. after    7 cycles
SCF Done:  E(RHF) =  -74.9658817799     A.U. after    6 cycles
SCF Done:  E(RHF) =  -74.9659006903     A.U. after    6 cycles
SCF Done:  E(RHF) =  -74.9659012170     A.U. after    5 cycles
```

Note that our initial guess was really rather good (the total drop in energy from beginning to end is only about 4 mE_h, or 2.5 kcal mol^{-1}). The final structure has bond lengths of 0.9894 Å and an internal bond angle of 100.03 deg, which is not very much different from where we started. In terms of other properties we looked at previously, the

dipole moment for the optimized structure is 1.71 D, $\langle r^2 \rangle$ = 18.27 (a bit more diffuse), and the Mulliken charges are –0.331 on O and 0.165 on H. So, the optimized structure is predicted to be a bit less polar than our guess structure.

Note that up to this point, we've been assuming that positions on the potential energy surface characterized by zero gradients must be minima. Recall, from calculus, that *any* stationary point has zero gradient. Minima are stationary points where the energy rises in every coordinate direction. Another chemically important stationary point on the potential energy surface is one where the energy rises in every coordinate direction except for *one*, and along that coordinate, the energy drops. Such a species defines a "transition-state structure". Transition-state structures are like passes through the mountains of the potential energy surface that separate one minimum from another. Reacting species can be thought of as flocks of birds flying in the valleys of minima. Chemical reactions occur when the birds fly over the lowest altitude saddle point (pass) separating one valley from another. We'll return to talk more about transition-state structures after Exam IV.

Homework

To be solved in class:

Now that we've done so much work with H_2O, consider a different molecule: H_2S. If we were to carry out an HF/STO-3G calculation for hydrogen sulfide instead of water, what aspects of the calculation would be similar to those for water? What aspects would be different? How would you expect the *results* from the calculations to differ? In particular, how would such optimized observables as total energy, bond lengths, bond angles, dipole moments, etc., be expected to differ? Explain the trend in each case.

To be turned in for possible grading Apr. 28: none.