



Short Course on Density Functional Theory and Applications

IV. Problems and Insights

Samuel B. Trickey

©Sept. 2008

Quantum Theory Project

Dept. of Physics and Dept. of Chemistry

trickey@qtp.ufl.edu



KS Eigenvalues

Spectroscopy causes great interest in one-electron energies. Because of Koopmans' theorem, these often are interpreted in terms of HF eigenvalues.

- Until now we've paid almost no attention to the KS eigenvalues, except for proving the Slater-Janak theorem (Lect. II-13)

$$\frac{\partial E_{v_{\text{ext}}}}{\partial n_j} = \epsilon_j$$

- Even passing familiarity with Koopmans' theorem should raise the expectation that interpretation of KS and HF eigenvalues differs. Koopmans' theorem is as follows. Suppose a neutral system $N_e = Z_{\text{tot}}$. Then the "unrelaxed" estimates of the first ionization potential (energy to remove one electron) and electron affinity (energy to add one electron) are

$$I := E_{\text{tot}, N_e = Z_{\text{tot}} - 1} - E_{\text{tot}, N_e = Z_{\text{tot}}} \approx \Delta E_{\text{unrel}, I} = -\epsilon_{\text{HOMO}}^{\text{HF}}$$

$$A := E_{\text{tot}, N_e = Z_{\text{tot}}} - E_{\text{tot}, N_e = Z_{\text{tot}} + 1} \approx \Delta E_{\text{unrel}, A} = -\epsilon_{\text{LUMO}}^{\text{HF}}$$

HOMO = Highest Occupied Molecular Orbital

LUMO = Lowest Unoccupied Molecular Orbital

Koopmans' theorem: integer change in electron number

Slater-Janak theorem: differential change in electron number



KS Eigenvalues and the Bandgap Problem

- Here is a table of atomic I values calculated from HOMO energies and total energy differences (all in eV) [adapted from SBT, Phys. Rev. Lett. 56, 881 (1986); HF values from C. Froese Fischer, *The Hartree-Fock method for Atoms*]

Atom	$-\epsilon_{HOMO}^{LSDA}$	ΔE^{LSDA}	$-\epsilon_{HOMO}^{HF}$	Expt.
H	7.6	13.3	13.6	13.595
He	15.6	24.2	25.0	24.6
Li	3.4	5.7	5.3	5.4
N	6.5	12.3	15.4	14.54
Ne	13.3	22.7	23.1	21.6
Na	3.3	5.6	5.0	5.14
P	6.5	10.7	10.7	10.55
Ar	10.6	16.1	16.1	15.76
Cr	4.2	7.7	6.5	6.67

Notice that LSDA total energy differences

(i) are NOT approximated by the negative of the HOMO energy. This behavior is consistent with the difference between Koopmans' and Slater-Janak theorems.

(ii) are reasonably good approximations to I

This disparity between homo eigenvalues and total energy differences is one manifestation of the “bandgap problem”.

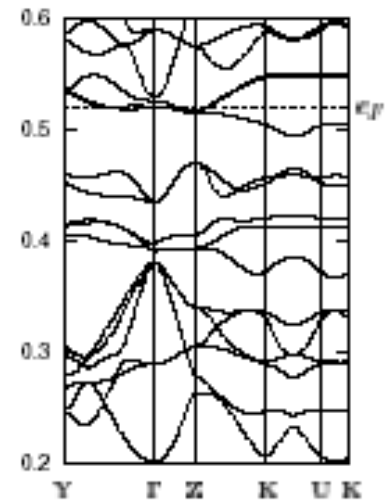
KS Eigenvalues and the Bandgap Problem

- For insulating solids, the LSDA bandgap (i.e. from “bare” KS eigenvalues) is too small, by 30-50%. [Table adapted from M. Hybertsen and S. Louie, *Adv. Quantum Chem.* **21**, 155 (1990) and SBT, Green, and Averill, *Phys. Rev. B* **8**, 4822 (1973)]

Element	ϵ_g^{LSDA}	ϵ_g^{EXPT}
C (dia.)	3.9	5.48
Si	0.52	1.17
Ar	8.4	14.29
Ge	< 0	0.744
LiCl	6.0	9.4

- As shown with Ge above, sometimes the bare KS eigenvalues even give a vanishing bandgap in a system known to be an insulator. Here →
are the LSDA bands for FeO, which has an experimental gap of 2.4 eV

[E. Engel, Chap. 2, *A Primer in Density Functional Theory*, Fig. 2.2]
Yet, the very same LSDA calculations typically give a good account of structure and reasonable cohesive energy.



KS Eigenvalues and the Bandgap Problem

- **Bad news – the problem also shows up in GGAs, with essentially the same magnitudes, just differences in detail.**
- **Remark – This presentation is not in historical order. By the time GGAs were introduced, the causes of the band gap problem were well-explored and reasonably well-understood.**

KS Eigenvalues – the Ionization Potential Theorem

• **Theorem:** For the *exact* E_{xc} functional, the KS HOMO eigenvalue is exactly the negative of the ionization potential. [M. Levy, J.P. Perdew, and V. Sahni, Phys. Rev. A **30**, 2745 (1984)] Here is a somewhat “hand-waving” demonstration of the physics of the argument.

• From the KS construction (Lect. II-3) $\frac{\delta E_{v_{ext}}}{\delta n} = \mu$

But μ enforces the particle number N_e . So, we can prove (not proved here) $\mu = \frac{\partial E}{\partial N_e} \equiv \varepsilon_F = -I$

But, the Slater-Janak theorem says that the least energy cost to change electron number is the HOMO eigenvalue $\frac{\partial E_{v_{ext}}}{\partial n_{HOMO}} = \varepsilon_{HOMO}$

Therefore, $\varepsilon_{HOMO} = -I$

▪ Since this is a theorem about the *exact* E_{xc} and neither LSDA nor GGAs behave in keeping with the theorem, the obvious conclusion is that there is something wrong with the approximate functionals.

▪ But that may not be the whole story because a bandgap is not just I :

$$\varepsilon_g := I - A$$

Self-interaction Error: the H Atom

Recall Lect. III-4. Here's the only line of the table of LSDA results needed:

Atom	$E_{ext}[n]$	$T_S[n]$	$E_{ee}[n]$	$E_x[n]$	$E_c[n]$	$E_{tot}[n]$
H	-26.8	13.2	8.4	-7.2	-0.6	-13.0

And here are the relevant equations

$$E_{v_{ext}}[n] = T_S[n] + E_{ee}[n] + E_{xc}[n_\alpha, 0] + E_{ext}[n]$$

$$T_S[n] \equiv -\frac{1}{2} \int d\mathbf{r}_1 \phi_{1\alpha}(\mathbf{r}) \nabla^2 \phi_{1\alpha}(\mathbf{r})$$

$$E_{ee}[n] = \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

$$E_{ext} = \int d\mathbf{r} v_{ext}(\mathbf{r}) n(\mathbf{r}); \quad n(\mathbf{r}) \equiv n_\alpha(\mathbf{r}) = |\phi_{1\alpha}(\mathbf{r})|^2$$

In atomic H, the electron has a
Coulomb self-repulsion:
completely spurious

In atomic H, E_{xc} must cancel that self-repulsion (“self-interaction”) exactly.

But it is a one-electron system, so $E_c \equiv 0$ and therefore $E_x[n_\alpha, 0] = -E_{ee}[n]$

As the table shows clearly, the LSDA does not achieve this cancellation by 1.2 eV.

Even adding the also spurious E_c does not achieve cancellation by 0.6 eV.

Self-interaction Error

- The fact that the LSDA energy differences ΔE^{LSDA} are reasonable approximations for the ionization potentials (slide 3) while the eigenvalues are not, suggests that the energy differences benefit from error-cancellation by subtraction.
- Such cancellation behavior also is consistent with self-interaction error: spurious repulsion pushes the orbital eigenvalues up, whereas taking a difference removes the total energy error approximately:

$$\Delta E^{LSDA} = E_{v_{ext},N-1}^{LSDA} - E_{v_{ext},N}^{LSDA} \approx E_{v_{ext},N-1} + \delta_{SI,N-1}^{LSDA} - \left(E_{v_{ext},N} + \delta_{SI,N}^{LSDA} \right)$$

$$\delta_{SI,N-1}^{LSDA} \approx \delta_{SI,N}^{LSDA}$$

- The physics of the second equation is that most of the self-interaction comes from the highly localized states, i.e. core states, which have relatively little relaxation upon ionization.
- A version of the cancellation argument is consistent with LSDA and GGA atomization energies and cohesive energies being too large in magnitude (systems are too bound):

$$E_{cohesive}^{LSDA} = E_{v_{ext},aggregate}^{LSDA} - \sum_{atoms} E_{v_{ext},atom}^{LSDA} \approx E_{v_{ext},aggregate} + \delta_{SI,aggregate}^{LSDA} - \sum_{atoms} \left(E_{v_{ext},atom} + \delta_{SI,atom}^{LSDA} \right)$$

$$\delta_{SI,aggregate}^{LSDA} < \sum_{atom} \delta_{SI,atom}^{LSDA} \Rightarrow \left| E_{cohesive}^{LSDA} \right| > \left| E_{cohesive} \right|$$

Negative Ion Problem

- Electron affinities by E^{LSDA} in LSDA and GGA do not come out well and, in some cases, the negative ions don't converge at all. How bad is the problem?
- Consider F^- .
 - In central field, this should be a Ne configuration $1s^2 2s^2 2p^6$
 - For v_{xLSDA} , the *most* negative configuration achievable with an all-numerical code is $1s^2 2s^2 2p^{5.76}$
[N. Rösch and SBT, J. Chem. Phys. 106, 8940 (1997).]
 - This is a clear sign of self-repulsion and/or a potential that decays too fast (hence, cannot support diffuse bound states).
- So how is it that some workers get good values of A from simple functionals?
 - (1) J.M. Galbraith and H.F. Schaefer III, J. Chem. Phys. 105, 862 (1996)
 - (2) J.P. Perdew et al., Phys. Rev. B 46, 6671 (1992).
 - ✓ Answer (1): a finite basis compels unbound, non-normalizable orbitals to be normalizable!
 - ✓ Answer (2): the input density was HF, which is self-interaction free!

Negative Ion Problem - Insights

- In addition to self-interaction, another contributor to the difficulty is that LSDA and GGA potentials have the wrong asymptotic behavior. The difficulty is easiest to see in LSDA X for a single atom. At large distance from the atom, the true density decay is hydrogen-like:

$$n(\mathbf{r}) \sim \exp(-2^{3/2} I |\mathbf{r}|)$$
$$\Rightarrow v_{XKSLDA}(\mathbf{r}) = -\left(\frac{3}{\pi}\right)^{1/3} n^{1/3}(\mathbf{r}) \sim -\left(\frac{3}{\pi}\right)^{1/3} \exp\left[-(2^{3/2}/3) I |\mathbf{r}|\right]$$

- The true density decay given above is a consequence of the asymptotic behavior of the correct physical potential. For that potential, BOTH the Coulomb and X contributions go asymptotically as $v_X(\mathbf{r}) \sim -1/r$, $v_{ee} \sim 1/r$

(Hint: think about E_x in atomic H again – it is just a self-interaction correction, so it must be coulombic, hence its functional derivative must look coulombic.)

- Deduction – the LSDA and GGA densities must decay too fast. This over-localization can only worsen the self-interaction error.

Derivative Discontinuity

- “But wait, there’s more!” (in the language of TV ads).
- Think about DFT, not Spin-DFT. (After all, the HK theorems guarantee that the ground state can be parameterized by the density alone.)
 - Consider H_2^+ as the bond length R grows arbitrarily large.
At each R the system has a single Fermi level.
Thus, for LDA, GGA, MGGA, and other XC functionals that are not explicitly orbital-dependent, the large R limit is $1/2$ an electron centered on each nucleus, not H plus H^+
 - Now consider a heteronuclear diatomic, ab , again as R grows arbitrarily large
By a similar argument, the large R limit is $a^+ + b^-$, not the proper neutral a and b atoms.
- What is going on?

Derivative Discontinuity (cont'd.)

•From Lect. II-15, systems with non-integer electron number are well-defined in DFT by a two-state ensemble. Suppose fixed Z_{tot} and electron number = $N \pm \beta$ with $N = Z_{tot}$ and $0 \leq \beta \leq 1$. Then the lowest energy for the two systems and the corresponding densities are linear combinations of the ground state energies and densities for the integer electron number systems:

$$E_{N_e+\beta} = (1-\beta)E_{N_e,0} + \beta E_{N_e+1,0}; \quad 0 \leq \beta \leq 1$$

$$E_{N_e-\beta} = (1-\beta)E_{N_e,0} + \beta E_{N_e-1,0}; \quad 0 \leq \beta \leq 1$$

$$n_{N_e+\beta}(\mathbf{r}) = (1-\beta)n_{N_e,0}(\mathbf{r}) + \beta n_{N_e+1,0}(\mathbf{r})$$

$$n_{N_e-\beta}(\mathbf{r}) = (1-\beta)n_{N_e,0}(\mathbf{r}) + \beta n_{N_e-1,0}(\mathbf{r})$$

•Therefore, the chemical potentials are

$$\mu_{N_e+\beta} = \frac{\partial E}{\partial N} = \frac{\partial E_{N_e+\beta}}{\partial \beta} = E_{N_e+1,0} - E_{N_e,0} = -A_{N_e}; \quad 0 \leq \beta \leq 1$$

**μ is DISCONTINUOUS
at integer electron
number!**

$$\mu_{N_e-\beta} = \frac{\partial E}{\partial N} = -\frac{\partial E_{N_e-\beta}}{\partial \beta} = E_{N_e,0} - E_{N_e-1,0} = -I_{N_e}; \quad 0 \leq \beta \leq 1$$

Derivative Discontinuity (cont'd.)

Here is a vivid illustration of the difference between exact, discontinuous behavior and the continuous (inexact) behavior of most functionals.

The system is the carbon atom. [P. Mori-Sánchez, A.J. Cohen, and W. Yang, *J. Chem. Phys.* **125**, 201102 (2006)]

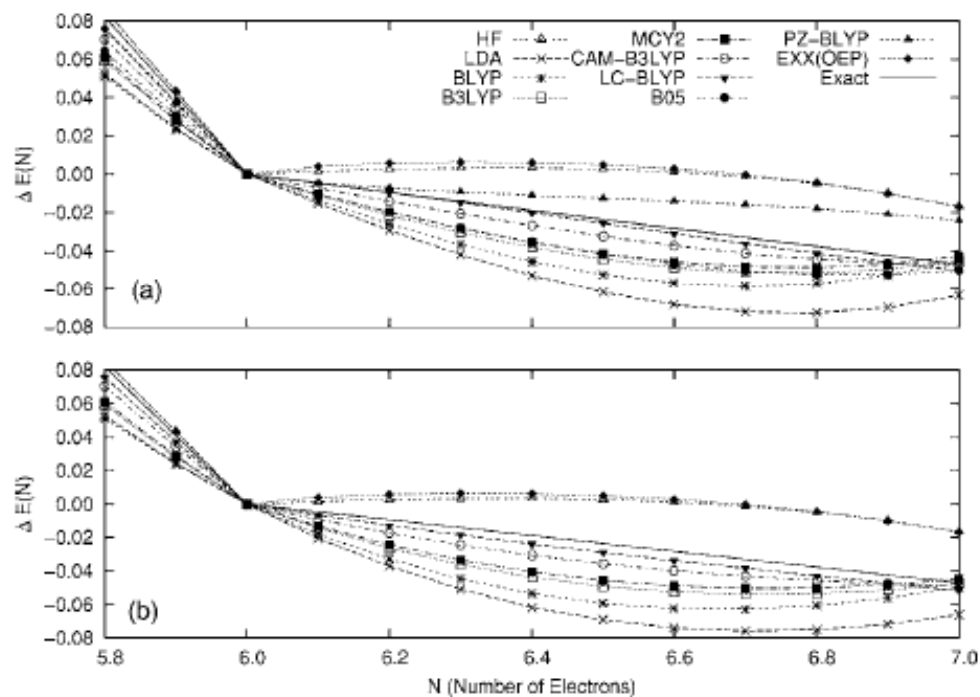


FIG. 1. The difference in energy of C with fractional numbers of electrons from the C atom, calculated in an unrestricted formalism using an aug-cc-pVTZ basis set with several different functionals (a) post-HF and (b) self-consistent.

Derivative Discontinuity (cont'd.)

- If now we keep in mind the definition of the exchange-correlation potential, the minimization result $\frac{\delta E_{v_{ext}}}{\delta n} = \mu_{N_e}$

(where the subscript on μ is added just to make the electron number dependence explicit), and the fact of the integer discontinuity, then

$$\left. \frac{\delta E_{xc}}{\delta n} \right|_{N_e+\beta} = v_{xc,N_e}(\mathbf{r}) + C_{N_e}; \quad 0 \leq \beta \leq 1$$

$$\left. \frac{\delta E_{xc}}{\delta n} \right|_{N_e-\beta} = v_{xc,N_e}(\mathbf{r}); \quad 0 \leq \beta \leq 1$$

- Thus the band gap in terms of KS eigenvalues is in the form of a Koopmans' difference *plus* a derivative discontinuity correction.

$$\mathcal{E}_g = \mathcal{E}_{LUMO,N_e}^{KS} - \mathcal{E}_{HOMO,N_e}^{KS} + C_{N_e}$$

Derivative Discontinuity and the Doppelgänger Problem

- The quantitative problem is to ascertain – or at least estimate – how much of the band gap error in bare KS eigenvalues is due to the derivative discontinuity and how much to omitted self-interaction error correction.

$$\mathcal{E}_g = \mathcal{E}_{LUMO, N_e}^{KS} - \mathcal{E}_{HOMO, N_e}^{KS} + C_{N_e} + C_{SIC}$$

Derivative Discontinuity – Back to the Stretched Diatomic Molecule

- So how does this resolve the two remote atoms dilemma (slide 11)? Again, consider the neutral ab molecule, stretched arbitrarily. Now shift δN_b of charge from a to b . The resulting energy change is

$$\Delta E = \mu_a(-\delta N_b) + \mu_b \delta N_b = -I_a(-\delta N_b) - A_b \delta N_b = (I_a - A_b) \delta N_b > 0$$

The last inequality is an experimental fact. The largest known $A = 3.62$ eV (Chlorine) is smaller than the smallest known $I = 3.89$ eV (Cesium).

So far as I know it has never been proved (e.g. as part of the stability of Coulombic matter).

In any event, the result is that the energy minimum is at integer electron numbers.

Notice that LSDA, GGA do not have a derivative discontinuity and therefore, the foregoing argument fails (naïvely, it gives $\Delta E = 0$), so the system can optimize to non-integer electron number on each center.

Editorial -The Systematic Improvement Issue

An oft-heard complaint, especially from “ab initio” quantum chemists, but also from formal many-body physicists, is that “DFT lacks a systematic means of improving the approximations used.”

The complaint is accurate only in a limited sense. It is true that DFT lacks a *mechanical recipe* for adding complexity to the approximations. In this sense, it is unlike many-body perturbation theory or the coupled-cluster method or configuration interaction.

***But*, adding complexity does not necessarily improve accuracy! The success of MBPT-2 compared to higher-order many-body perturbation theory is a clear example.**

What is true in DFT is that there are many theorems, limits, bounds, etc. that can and should be enforced on approximations. The process, however, requires thought and ingenuity. It is not mechanical.