

Kinetic Balance in Relativistic Electronic Structure Methods

Joshua Goings

April 27, 2014

Kinetic balance was discovered when early attempts at relativistic SCF calculations failed to converge to a bound state. Often, the energy was too low because of the negative energy continuum. The crux of the problem was that the four components of the spinor in relativistic methods were each allowed to vary independently. In other words, scientists treated each component with its own, independent basis without regard to how the components depended on each other. This problem was eventually solved by paying attention to the non-relativistic limit of the kinetic energy, and noting that the small and large components of the four-spinor are not independent. There is, in fact, a coupling that we refer to as “kinetic balance”. I’ll show you how it works.

First, as you may have guessed, we only need to consider the one-electron operator. This is the matrix form of the Dirac equation, and it contains (in addition to other terms) the contributions to the kinetic energy. Written as a pair of coupled equations, we have

$$\begin{aligned} [\mathbf{V}^{LL} - E\mathbf{S}^{LL}] \mathbf{c}^L + c\mathbf{\Pi}^{LS} \mathbf{c}^S &= 0 \\ c\mathbf{\Pi}^{SL} \mathbf{c}^L + [\mathbf{V}^{SS} - (2mc^2 + E)\mathbf{S}^{SS}] \mathbf{c}^S &= 0 \end{aligned} \quad (1)$$

Where $\mathbf{V}^{LL} = \langle \chi^L | V | \chi^L \rangle$, $\mathbf{S}^{LL} = \langle \chi^L | \chi^L \rangle$, and $\mathbf{\Pi}^{LS} = \langle \chi^L | -i\hbar\boldsymbol{\sigma} \cdot \nabla | \chi^S \rangle$, and so on. These are the potential, overlap, and momentum terms of the Dirac equation in a basis $|\chi\rangle$.

Now, if we look at the second of the two paired equations, we note that for potentials of chemical interest (e.g. molecular or atomic), \mathbf{V}^{SS} is negative definite. We also know that when we solve the equation, we are looking for an energy above the negative energy continuum, which is to say we want $E > -2mc^2$. Since the overlap is positive definite, putting all of these constraints together mean that we have a nonsingular matrix (and therefore invertible!) matrix in the second of our coupled equations. We can rewrite the second equation as

$$\mathbf{c}^S = [(2mc^2 + E)\mathbf{S}^{SS} - \mathbf{V}^{SS}]^{-1} c\mathbf{\Pi}^{SL} \mathbf{c}^L \quad (2)$$

Substituting this expression back into the first (top) equation yields

$$[\mathbf{V}^{LL} - E\mathbf{S}^{LL}] \mathbf{c}^L + c\mathbf{\Pi}^{LS} [(2mc^2 + E)\mathbf{S}^{SS} - \mathbf{V}^{SS}]^{-1} c\mathbf{\Pi}^{SL} \mathbf{c}^L = 0 \quad (3)$$

This form is very useful for analysis. Now, we are going to use the matrix relation

$$(\mathbf{A} - \mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{A} - \mathbf{B})^{-1} \quad (4)$$

with $\mathbf{A} = 2mc^2\mathbf{S}^{SS}$ and $\mathbf{B} = \mathbf{V}^{SS} - E\mathbf{S}^{SS}$. This leads to the rather long expression

$$\begin{aligned} & \left[\mathbf{V}^{LL} - E\mathbf{S}^{LL} + \frac{1}{2m}\mathbf{\Pi}^{LS} [\mathbf{S}^{SS}]^{-1} \mathbf{\Pi}^{SL} \right] \mathbf{c}^L = \dots \\ \dots & = \frac{1}{2m}\mathbf{\Pi}^{LS} [\mathbf{S}^{SS}]^{-1} [\mathbf{V}^{SS} - E\mathbf{S}^{SS}] [\mathbf{V}^{SS} - (2mc^2 + E)\mathbf{S}^{SS}]^{-1} \mathbf{\Pi}^{SL} \mathbf{c}^L \end{aligned} \quad (5)$$

Why this ridiculous form? Look closely at each side and their dependence on the speed of light, c . The left hand side has the c^0 terms, and the right hand side has the c^{-2} terms. Since the non relativistic limit is found when the speed of light is infinite ($c \rightarrow \infty$), the whole right hand side goes to zero. This gives us

$$\left[\mathbf{V}^{LL} - E\mathbf{S}^{LL} + \frac{1}{2m}\mathbf{\Pi}^{LS} [\mathbf{S}^{SS}]^{-1} \mathbf{\Pi}^{SL} \right] \mathbf{c}^L = 0 \quad (6)$$

Now, if this is indeed the true non-relativistic limit, then we find that our kinetic energy term \mathbf{T}^{LL} is given by

$$\mathbf{T}^{LL} = \frac{1}{2m}\mathbf{\Pi}^{LS} [\mathbf{S}^{SS}]^{-1} \mathbf{\Pi}^{SL} \quad (7)$$

Or, more explicitly,

$$T_{\mu\nu}^{LL} = \frac{-\hbar^2}{2m} \sum_{\kappa\lambda} \langle \chi_\mu^L | \boldsymbol{\sigma} \cdot \nabla | \chi_\kappa^S \rangle \langle [\mathbf{S}^{SS}]_{\kappa\lambda}^{-1} \langle \chi_\lambda^S | \boldsymbol{\sigma} \cdot \nabla | \chi_\nu^L \rangle \quad (8)$$

Where that inner part, $|\chi_\kappa^S \rangle \langle [\mathbf{S}^{SS}]_{\kappa\lambda}^{-1} \langle \chi_\lambda^S |$, is an inner projection onto the small component basis space. Less formally, the small component is the ‘‘mathematical glue’’ that connects the two momentum operators. If the small component spans the same space as the momentum operators in the large component space, $\{\boldsymbol{\sigma} \cdot \nabla \chi_\mu^L\}$, then that inner projection just becomes the identity. This means that the expression becomes

$$T_{\mu\nu} = \frac{-\hbar^2}{2m} \langle \chi_\mu^L | (\boldsymbol{\sigma} \cdot \nabla) (\boldsymbol{\sigma} \cdot \nabla) | \chi_\nu^L \rangle = \frac{-\hbar^2}{2m} \langle \chi_\mu^L | \nabla^2 | \chi_\nu^L \rangle \quad (9)$$

Which is the kinetic energy term in the non-relativistic formulation! (N.B. We used the relation $(\boldsymbol{\sigma} \cdot \nabla) (\boldsymbol{\sigma} \cdot \nabla) = \nabla^2$).

So, when we set up our relativistic calculations, as long as we have the constraint that

$$\chi_\mu^S = (\boldsymbol{\sigma} \cdot \mathbf{p}) \chi_\mu^L \quad (10)$$

Then we will find that we recover the correct non-relativistic limit of our equations. The basis is called ‘‘kinetically balanced’’, and we won’t collapse to energies lower than $-2mc^2$.

A few stray observations before we finish. First, if we enforce the relation between small and large component basis functions, then we find that $\mathbf{\Pi}^{SL} = \mathbf{S}^{SS}$ and $\mathbf{\Pi}^{LS} = 2m\mathbf{T}^{LL} = 2m\mathbf{T}$. Second, this constraint actually maximizes kinetic energy, and any approximation that does not satisfy the kinetic balance condition will lower the energy. This was weird to me coming from the non relativistic Hartree-Fock background, where variationally, if you remove basis functions you raise the energy. The thing is that when doing relativistic calculations, you aren't bounded from below like their non-relativistic counterparts. While you can get variational stability, you are actually doing an "excited state" calculation (I am using the term "excited state" very loosely). Kind of odd, but the negative energy continuum does exist, and was a big factor in the predicting the existence of antimatter.

Finally, modern methods of relativistic electronic structure theory make use of the kinetic balance between large and small component basis functions to eliminate the small component completely. These are called "Dirac-exact" methods. One such example is NESC, or Normalized Elimination of the Small Component. In addition to reproducing the Dirac equation exactly, they have numerous computational benefits, as well as easily allowing for most (if not all) non-relativistic correlated methods to be applied directly. Thus after doing an NESC calculation, you get relativistic "orbitals" which can immediately be used in, say, a coupled cluster calculation with no computational modification.