Some Questions PhL 10.26.06

****

###

Question: Assume that operators A and B commute. Are the eigenstates of B also eigenstates of A?

Answer ahead of time: if you choose the eigenstates of B properly such that any degenerate subspaces are diagonalized, then yes, the eigenstates of B are also eigenstates of A.

Think of this as a matrix question in an N-dimensional space, so we are talking eigenvectors.

$A|a_i\rangle = a_i|a_i\rangle$ // the eigenvectors and eigenvalues of A
$B|b_i\rangle = b_i|b_i\rangle$ // the eigenvectors and eigenvalues of B

First, consider this situation. Suppose someone tells you that $B|f\rangle = b_i|f\rangle$. If the eigenvalue b_i has multiplicity 1 (is non-degenerate), then you know that $|f\rangle = k_i|b_i\rangle$, which says that $|f\rangle$ is an eigenvector of B, and if f is normalized, then $k_i = b_i$. Otherwise you know that

$|f\rangle = \sum_n k_n |b_i,n\rangle$ where $|b_i,n\rangle$ are the degenerate eigenvectors.

That is, $|f\rangle$ is a vector in the degenerate subspace. So in this case one could not conclude that " $|f\rangle$ is an eigenvector of B", but rather that "$|f\rangle$ is a linear combination of eigenvectors of B in the subspace of b_i ".

The relevance of this situation is as follows: Since A and B commute, we can say

$B\,[A|b_i\rangle] = A\,B|b_i\rangle = A\,b_i\,|b_i\rangle = b_i\,[A|b_i\rangle]$ for all i

Thus, the vector $[A|b_i\rangle]$ is like the $|f\rangle$ discussed above. If b_i is nondegenerate, then $A|b_i\rangle = k_i\,|b_i\rangle$, so we conclude that the eigenstates of B are also eigenstates of A. If b_i is degenerate, then we conclude that:

$A|b_i\rangle = \sum_n k_n |b_i,n\rangle$

Using G-S orthogonalization in the subspace, you can select the states $|b_i,n\rangle$ to remove the summation. Then you can conclude that in this basis, the eigenstates if B are in fact eigenstates of A.

In our H_2+ molecule situation, the symmetric ground state is non-degenerate so in this particular case, since [H,P] = 0, the solutions MUST (as claimed above) have definite parity. When dealing with higher states, there is going to be degeneracy and you have to deal with the GS issue.

In our matrix notes, we do claim this Theorem 19E:

"If two Hermitian matrices commute, then *you can find* a unitary transformation that diagonalizes both matrices simultaneously. Conversely, if you can find such a unitary transformation, then the matrices must commute".

The conversely part is easy to prove, and this appears on page 330 of M&M. The other direction is harder because of the degenerate subspace issue. This is what we discussed above in operator notation, and is more clearly discussed here on the web:

http://galileo.phys.virginia.edu/classes/751.mf1i.fall02/751LinearAlgebra.htm (now reaped)

Corollary A: Suppose A and B commute, and suppose the eigenvectors of B are all non-degenerate (but we know nothing about the degeneracy of the A eigenvectors). Then suppose $B|f\rangle = b_i|f\rangle$. We may conclude that $|f\rangle = k_i\ |b_i\rangle$ by inserting a simple expansion and knowing that the B states are non-degenerate, as outlined above. Now consider:

$$B\ [A|b_i\rangle] = A\ B|b_i\rangle = A\ b_i\ |b_i\rangle = b_i\ [A|b_i\rangle] \qquad \text{for all i}$$

We now think of $|f\rangle = [A|b_i\rangle]$ and we conclude that $[A|b_i\rangle] = k_i\ |b_i\rangle$. Thus, we conclude that $|b_i\rangle$, which we assumed was an eigenstate of B, is also an eigenstate of A. The eigenvalue is $a_i = k_i$. The eigenstates of A may have degeneracy, but we don't care about that.

Corollary B: Suppose as above, but all we know is that state $|b_i\rangle$ is non-degenerate for some particular i. The conclusion still follows, but only for that one state. That is to say, if $|b_i\rangle$ is an eigenstate of B, then it is also an eigenstate of A, but only for that value of i.

Application: Suppose H and P commute, where H is a Hamiltonian and P is the parity operator. The spectrum of operator P is { +1, -1} with no degeneracy. Therefore, any eigenstate of P is also an eigenstate of H. This does not say, however, that an eigenstate of H is necessarily an eigenstate of P, unless we are talking about an eigenstate of H that happens to be non-degenerate, such as perhaps a ground state of some system.

Application: A nucleus has a Hamiltonian H and is assumed to be in its ground state of energy E_n which is non-degenerate. For all interactions assumed significant for the nucleus, parity is conserved so HP=PH. In this case, the ground state must have a definite parity -- the eigenstate of H is also an eigenstate of P. So we can write such a ground state as $|n=0,p\rangle$ where $p = \pm 1$.

Application: Suppose as above we write a state as $|n,p\rangle$. Suppose V is a vector operator under parity so we know that $P^{-1}\mathbf{V}\ P = -\mathbf{V}$. (**V** is not a pseudovector). Then it follows that $\langle n,p\ |\ \mathbf{V}\ |\ n,p\rangle = 0$ as a one line proof shows, since $p^2 = 1$. This is why a ground state nucleus has no electric dipole moment, since the operator in question there is an axial vector related to **r (** integral of ρ x **r**). In contrast, the magnetic moment is a pseudovector so this argument does not apply, and a nuclear ground state *can* have a magnetic moment. Think of magnetic moment coming from a ring of current centered at an origin. Under parity, that ring maps into itself (both r and v are regular vectors so invert) and B does not change so M does not change. B is also a pseudovector. Of course to the extent that a nucleus includes weak interactions, there may be some small parity violation, and then there could be some small nuclear electric dipole moment. [By the same argument, an electron has no electric dipole moment.]

****

###

Question: Derive the rule that [H+][OH-] = 10^{-14} . Why is this true? Why does the equation have the form shown, and why is the number on the right the one shown?

Answer: Consider reaction A + B → C + D. It seems reasonable to assume that the reaction rate to the right is proportional to [A] at least in a low molar solution. So rate to the right is k_r [A][B]. The rate to the left is k_l [C][D]. In equilibrium, these rates are equal, so we can write:

$$\frac{[C][D]}{[A][B]} = K, \text{ some constant that is a function of temperature}$$

The tradition is to put the right side molarities on the top.

Now suppose the reaction were really 2A → C + D. Just treat this as A + A → C + D, and you end up with a derivation of the power rule where instead of [A][B] we have $[A]^2$. If you double the concentration of A, the reaction rate to the right goes up by 4X. There is really no trick to this result. So this leads to the general formula with powers that you see, for example, in the red book page 244. These rules apply of course to "single stage" reactions where you need everything to collide at once, and typically this is not the situation for complicated reactions!

So, for the reaction $H_2O \rightarrow H^+ + OH^-$ our general rule says

$$\frac{[H^+][OH^-]}{[H_2O]} = K, \text{ some constant that is a function of temperature}$$

But $[H_2O]$ is huge and constant in such a weak solution and it basically always $[H_2O]$ = 1000gm/18gm = 55.6 molar. There are 55.6 moles of 18 g each in a liter of water which is 1000 g. So we just absorb this constant into K to get $[H^+][OH^-] = Kw$. At 25 C, *it happens* that this number is Kw = 1.011 x 10^{-14} , but everyone treats it as if it were exactly 1 x 10^{-14} because we almost always use the logarithm of such umbers. Note that

$$\log\ 1.011 \times 10^{-14} = -14 + \log 1.001 = -14 + .00043 = -13.99996$$

Web quote: "This constant, Kw, is called the water autoprotolysis constant or water autoionization constant. (Sometimes the prefix auto is dropped, as was done in the title of this section.) It can be determined by experiment and has the value 1.011 x 10^{-14} at 25 °C. Generally, a value of 1.0 x 10^{-14} is used. "

In neutral water, we have $[H^+] = [OH^-]$ so then $[H^+]^2 = 1.011 \times 10^{-14}$ and $[H^+] = 1.005 \times 10^{-7}$ and then pH = -log 1.005 x 10^{-7} = 7 - log1.005 = 7 - .002 = 6.998. In general, pH is not measured to this high level of accuracy, so the error at 25C can be ignored.

The constant Kw does very with temperature substantially, see page D-79 of the CRC book. For example, at 0C the 14 number becomes about 15, and at 60C it is about 13. It just happens that it comes out very close to 1 at 25C, and our whole pH scale is then based on this. The human body temperature is 37C where Kw is about 13.6 instead of 14.0.

Just another comment about temperature. When for example, you have the A+B collision to make something, there is likely an activation potential to get over. In the distribution of A and B molecules, some will be in the high tail of the Poisson or whatever thermal distribution it is, and the reaction then might only go for those molecule pairs which collide physically *and* for which E(A) + A(B) ≥ activation.

As temperature increases, the there are more pairs for which the second condition is true, so we expect reaction rates to increase with temperature.

****

Question: Do a quick review of BTU, calories, running, windows and sun coming in. ###

The BTU is an amount of energy, like joules. Here are some conversions:

1 BTU = 1060 Joules = 252 calories = .25 kcal = .25 Cal
1 Cal = 4187 Joules
1 Joule = .00024 Cal
1 lb fat = 3500 Cals = 3500*4187 J = 14,654,500 Joules
50 watts = 50 Joules/sec
1 session = 50 watts for 30 minutes = 50*60*30 = 90,000 Joules = 90000*.00024 = 21.6 Cals

On the last line, only about 20% of exercise biking goes into mechanical work, so you get to multiply by 5 to get the total calorie burn. Thus, the above 30 minute session really burns about 100 Cals.

"If we take a group of cyclists, or a group of rowers and perform sub maximal testing on them to determine how much energy they consume when performing a standard sub maximal workload, we find that overall work efficiency will range between about **17 and 26%,** with an average somewhere in the middle of that range. In other words for every 100 Calories of energy burned, we manage to convert 20 Calories of that energy to useful work on the pedals of the ergometer, or as pulling power on the rowing machine. Now, if your goal is to lose body fat during exercise, then I suppose it pays to be inefficient, since it is Calories burned that matter. However, if your goal is to move your body faster than the other guy, than being 25% efficient is way better than 18%! So, what are the sources of inefficiency and what, if anything can we do about them?"

The body is said to consume about 100 watts at rest, so 2400 watt-hours each day.

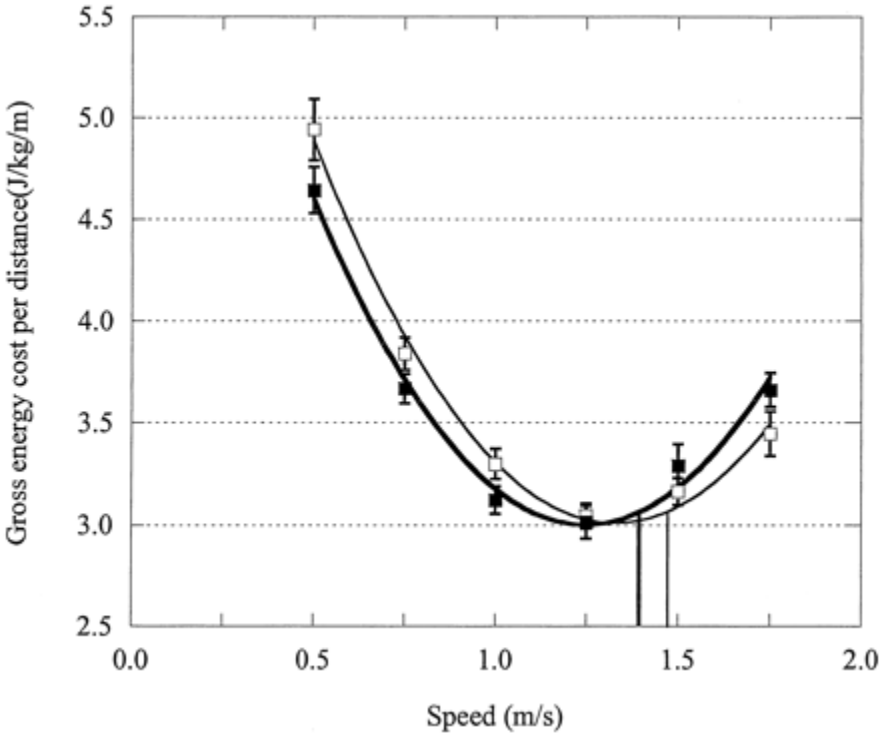
1 watt-hour = 1 Joule/sec * 3600 sec = 3600 Joules = 3600*.00024 = 0.864 Cals
2400 watt-hours = 2400* 0.864 = 2073.6 Cals.

Is this really true? Another site says 120 watts and about 2500 Cals/day to support it. So during the 30 minutes of bike ride up hill, you burn the 120 watts normal plus another 100 watts.

1 Cal/hour = 4187 Joules / 3600 sec = 1.16 watts
100 Cal/hour = 116 watts

So by doing nothing but activities of daily life, one is burning about 100 Cal/hour or 116 watts. Exercise adds more above this.

Here is an interesting graph that I have wondered about:



When people walk, they tend to walk at a speed the minimizes their energy consumption per distance walked! Normal people walk 1.33 msec (thin bar), while fat people walk a little slower (fat bar). Suppose I weigh 190 lbs/2.2 = 86 kg and walk 1 mile = 5280 feet/3 = 1760 meters roughly. Then according to the above graph, if I walk at the natural speed, I burn about 3*86*1760 = 454080 Joules /4187 = 108.45 Cals. This is a famous number, 100 calories for walking a mile. This is "gross" total energy burned.

So when you exercise, you burn EXTRA calories at the margin, and this might control whether you gain or lose weight. General conditioning is the real benefit, however. Back to this subject later.

Now about windows. (see notes in Misc Phys binder). A 1/8" piece of glass conducts 300*ΔT(K) watts/m^2, and if there were no convection layer, that would be it. So imagine a mere 70-30 = 40F *5/9 = 22C = ΔT. Then you would conduct about 300*22 = 6600 watts/m^2 out. That is much less than S coming in during full-on sun! Luckily, this is not how windows really work.

Suppose your window has R = 1 in British units. That means R(mks) = .2 m^2-K/W. So if you have a 1 m^2 window, and ΔT = 1 degree K, then flow = 1*1/.2 = 5 watts flows through it. So if you have 22C across the window, maybe you have 100 watts flowing out, and then full sunlight brings 1000 watts in.

This subject is now on hold, I should buy a library card for $50 as a Rimrock expense.

****

###

Question: If a nucleus has an even number of protons and an even number of neutrons, why does the ground state have spin S = 0 ?

System is all fermions so must be overall AS. Probably the ground state in some sense has orbital L = 0 which is Sym, so iso x spin must be AS. Imagine a two-proton nucleus, Helium 2 (does not exist, diproton). The isospin state is |pp> which is one of three possible states of the I=1 system. Under particle exchange, such a state is S (Symmetric) for isospin. Thus, spin must be AS which means spin = 0. Same would apply to a 4-proton nucleus. What about a state |nn,pp> ? Well, basically to answer this question, you should consider a system of N spin-1/2 particles. You should then construct the ket eigenfunctions for each total spin state using CG coefficients. Then you write down a particular state of interest like |nnpp> and see what isospin symmetry it has. But I am then confused because you probably have to superpose states like |npnp>, so you then have to clarify the quantum ket notation and do it all properly. There are lots of exchange operators to think about. I guess I will defer doing this problem for now, but it should not be too hard.

****

Question: do the math for combining two spin=1/2 particles. ###

States are $|+1/2,+1/2\rangle$, $(1/\sqrt{2})\,[\,|+1/2,-1/2\rangle + |-1/2,+1/2\rangle]$, $|-1/2,-1/2\rangle$ for S = 1

$(1/\sqrt{2})\,[\,|+1/2,-1/2\rangle - |-1/2,+1/2\rangle]$ for S=0

The first three states are symmetric, the S = 0 state is antisymmetric under particle switch. So, if you have a state as noted above which is | proton, proton>, the isospin state is symmetric I = 1, and if $\ell = 0$, then the spin state must be antisymmetric and must therefore be S = 0.

****

Question: How do you derive the classical Larmor precession frequency? ###

Answer: Put a magnetic moment $\boldsymbol{\mu}$ into a constant B field. The torque on the magnet is $\boldsymbol{\mu}$ x **B.** The magnetic moment is related to the angular momentum **J** by equation $\boldsymbol{\mu} = \gamma \mathbf{J}$. Torque is also $\mathbf{N} = d\mathbf{J}/dt$. This is all the information you need. So we have $d\mathbf{J}/dt = \boldsymbol{\mu} \times \mathbf{B} = \gamma\, \mathbf{J} \times \mathbf{B}$ where B is assumed a constant vector, so put it in the z direction. If we write $\mathbf{J} = \mathbf{J_L} + \mathbf{J_T}$ we find that $d\mathbf{J_L}/dt = 0$ and

$$d\mathbf{J_T}/dt = \gamma\, \mathbf{J}_T \times \mathbf{B} = (\gamma B)\, \mathbf{J}_T \times \hat{\mathbf{z}} = (-\gamma B)\, \hat{\mathbf{z}} \times \mathbf{J}_T.$$

If you draw a little circle picture here, you see that, if $\gamma < 0$, $d\mathbf{J_T}$ goes along the circumference in a CCW direction so you have normal circular motion in the right-hand-rule sense. From the trivial geometry of this picture you compute that $\delta\theta \approx [(-\gamma B) J_T\, dt] / J_T$ so $\omega = d\theta/dt = -\gamma B$, QED. The minus sign is important because it says that if $\gamma > 0$, the Larmor $\mathbf{J}_T$ rotates clockwise which is backwards from the RHR sense.

****

Question: How is the quadrupole moment of a potential Φ(r) related to the electric field E(r) ? ###

Answer: We know that $\mathbf{E} = -\nabla\Phi$, and we know how to expand Φ in spherical harmonics. In my tensor harmonics notes, I know how to the right hand side like so:

$$-\nabla\Phi = -\Sigma_{\ell m}\{\partial_r\Phi_{\ell m}(r)\,\mathbf{Y}_{\ell m} + \Phi_{\ell m}(r)/r\;\mathbf{Z}_{\ell m}\}$$

where we are using here the *vector* harmonics (each of which is a function of θ and ϕ only):

$$\mathbf{X}_{\ell m} = \mathbf{r} \times \nabla\, Y_{\ell m} \qquad \mathbf{Y}_{\ell m} = \hat{\mathbf{r}}\, Y_{\ell m} \qquad \mathbf{Z}_{\ell m} = r\nabla\, Y_{\ell m}$$

We can expand the left hand side like so:

$$\mathbf{E} = \Sigma_{\ell m}\,[\,E_{x\ell m}(r)\,\mathbf{X}_{\ell m} + E_{y\ell m}(r)\,\mathbf{Y}_{\ell m} + E_{z\ell m}(r)\,\mathbf{Z}_{\ell m}\,]$$

If we dot each side of $\mathbf{E} = -\nabla\Phi$ with a vector harmonic and integrated $d\Omega$, I think we arrive at these conclusions:

$$E_{x\ell m}(r) = 0 \qquad E_{y\ell m}(r) = -\partial_r\Phi_{\ell m}(r) \qquad E_{z\ell m}(r) = -\Phi_{\ell m}(r)/r$$

So at least we can relate the multiple E field to the multiple Φ field in this way. ^

****

Question: If you expand a function $\Phi(\mathbf{r})$ as a Taylor series about the origin, you get this: ###

$$\Phi(\mathbf{r}) = \Phi(\mathbf{0}) + r \bullet \nabla\Phi + r \bullet \nabla\nabla\Phi \bullet r + \text{etc} \qquad \nabla\nabla = \text{Hessian, not } \nabla^2$$

The second term is just the double sum $r_i r_j\, \partial_i\partial_j\, \Phi$ so involves second derivatives of the potential. It is easy then to write the general form for any term in this series. Now the question is this: **how do the terms in the above expansion relate to the multipole moments $\Phi_{\ell m}(\mathbf{r})$** ?

Answer: First, the orthogonality of the Y's says this:

$$\int d\Omega\; Y_{\ell' m'}^*(\theta,\phi)\; Y_{\ell m}(\theta,\phi) = \delta_{\ell',\ell}\;\delta_{m',m}$$

We could then conclude that, if we expand $\Phi(\mathbf{r}) = \Sigma_{\ell m}\Phi_{\ell m}(r)\; Y_{\ell m}(\theta,\phi)$, then

$$\Phi_{\ell m}(r) = \int d\Omega\; Y_{\ell m}^*(\theta,\phi)\;\Phi(\mathbf{r}) = \int d\Omega\; Y_{\ell m}^*(\theta,\phi)\;[\,\Phi(\mathbf{0}) + r \bullet \nabla\Phi + r \bullet \nabla\nabla\Phi \bullet r + \text{etc}\,]$$

Now the first term gives $\Phi(\mathbf{0})\;\delta_{0,\ell}\;\delta_{0,m}$ a numerical constant so only contributes to the 00 multipole. Remember that all the gradients shown are evaluated at r=0 so are not functions of r. The second term

therefore gives $\nabla\Phi(0) \bullet [\int d\Omega \; Y_{\ell m}^*(\theta,\phi)\, \mathbf{r}]$. Think of the bracket as $\sim <\ell m| \, r \, | 00>$. Since r is an ℓ=1 type operator, we know that $1 \otimes 0 = 1$, so the only non-zero term will have $\ell = 1$. Thus, the second term can only contribute to the 1m multipoles. The third term involves $[\int d\Omega \; Y_{\ell m}^*(\theta,\phi)\, r_i\, r_j]$ which involves a Cartesian rank-2 tensor. This is no doubt a mixture of operators having ℓ = 2,1 and 0 and I would guess only ℓ = 2 and 0 due to parity somehow. So I would guess that our third term contributes to the 00 and 2m multipoles. And the progression would continue.

Conclusion #1: the 2m multipoles of $\Phi_{\ell m}(r)$ get contributions only from the $r \bullet \nabla\nabla\Phi \bullet r$ and higher terms in the expansion which we have ignored. So if we can ignore these higher terms in the potential expansion (probably the odd terms give nothing anyway), then we can make this claim:

"The ℓm = 2m components of the potential $\Phi_{\ell m}(r)$ arise from the second derivatives of the potential $\Phi(r)$, which means they arise from first derivatives of the electric field components like $\partial_i E_j$. In making this claim, we ignore the amplitude of fourth, six, etc derivatives of the potential Φ. "

****

###

Question: **What is the "secular approximation" ? Also, what is the Rotating Wave Approximation?**

Answer: Not very much general stuff on this. But here:

"The outcome of this analysis is perhaps surprising: the result which one obtains by making the rotating wave approximation in a quantum mechanical formalism (the Jaynes-Cummings model),"

".... invariance and vanishes when the **rotating-wave-approximation** is applied. ... the rotating wave approximation (also known as the secular approximation); ..."

So another name is the RWA method. That is the name it is mostly known under! See wiki here:

http://en.wikipedia.org/wiki/Rotating_wave_approximation

After reading Schiff, I think I am now prepared to understand this wiki page and will annotate it here.

Lemma #1: Can write $H = \sum_k E_k \, |k><k|$ in any picture, energy representation.

Proof: Consider H acting on an arbitrary state $|\psi>$. We have

$$H|\psi> \; = H \sum_k \; |k><k|\psi> \; = \sum_k E_k \, |k><k|\psi>$$

Since this is true for any vector ψ in the Hilbert space, it must be true as an operator equation, QED.

A shorter proof is this (all done in just using operators, and now the implied sum)

$H = H\ |k><k|\ = E_k\ |k><k|$

Application: spin system with ground state |g> having energy $E_g = 0$ and excited state |e> having energy $E_e = \hbar\omega$. Then, ignoring interaction H terms, we can say $H_0 = \sum_k E_k\ |k><k|\psi> = \hbar\omega_o\ |e><e|$, exactly as stated on the wiki page.

Now consider an interaction or perturbing Hamiltonian $H_I = -\mathbf{d}\bullet\mathbf{E}$ where E is an applied EM field and where d is the dipole moment operator in Hilbert Space. Because both states have definite parity, the diagonal dipole moment matrix elements vanish. The off-diagonals do not vanish, however. This is very easy to show on one line. So define $\mathbf{d_{eg}} = <e|\mathbf{d}|g>$ which is a simple vector (whereas **d** is an operator).

Now can write the d operator as follows:

$\mathbf{d} = |n><n|\mathbf{d}|m><m| = |e><e|\mathbf{d}|g><g| + |g><g|\mathbf{d}|e><e| = \mathbf{d_{eg}}\ |e><g| + \mathbf{d_{eg}}^*\ |g><e|$

Next, write the interaction Hamiltonian as follows, where the E field is real, but written as the sum of a complex term + CC:

$$\begin{aligned} H_I &= -\vec{d}\cdot\vec{E} \\ &= -\left(\vec{d}_{\rm eg}|{\rm e}\rangle\langle{\rm g}| + \vec{d}^*_{\rm eg}|{\rm g}\rangle\langle{\rm e}|\right)\cdot\left(\vec{E}_0 e^{-i\omega_L t} + \vec{E}_0^* e^{i\omega_L t}\right) \\ &= -\left(\vec{d}_{\rm eg}\cdot\vec{E}_0 e^{-i\omega_L t} + \vec{d}_{\rm eg}\cdot\vec{E}_0^* e^{i\omega_L t}\right)|{\rm e}\rangle\langle{\rm g}| - \left(\vec{d}^*_{\rm eg}\cdot\vec{E}_0 e^{-i\omega_L t} + \vec{d}^*_{\rm eg}\cdot\vec{E}_0^* e^{i\omega_L t}\right)|{\rm g}\rangle\langle{\rm e}| \\ &= -\hbar\left(\Omega e^{-i\omega_L t} + \tilde{\Omega} e^{i\omega_L t}\right)|{\rm e}\rangle\langle{\rm g}| - \hbar\left(\tilde{\Omega}^* e^{-i\omega_L t} + \Omega^* e^{i\omega_L t}\right)|{\rm g}\rangle\langle{\rm e}|, \end{aligned}$$

where $\Omega = \mathbf{d_{eg}}\bullet\mathbf{E_0}/\hbar$ and $\tilde{\Omega} = \mathbf{d_{eg}}^*\bullet\mathbf{E_0}/\hbar$. Ω has the units of frequency and is called the **Rabi Frequency**. It is not the same as the ω in the ħω spacing of the two spin states. It seems that Ω could be complex since $\mathbf{E_0}$ can be arbitrary complex unrelated to $\mathbf{d_{eg}}$.

Now here is the next block from the wiki page: Notice that we must have $\bar{H} = UHU^\dagger$ to go along with the way U is shown to act on states.

The next stage is to find the Hamiltonian in the interaction picture, $\bar{H}$. The unitary operator required for the transformation is $U = e^{iH_0 t/\hbar}$, and an arbitrary state $|\psi\rangle$ transforms to $|\bar{\psi}\rangle = U|\psi\rangle$. The Schrödinger equation must still hold in this new picture, so

$$\bar{H}|\bar{\psi}\rangle = i\hbar\partial_t|\bar{\psi}\rangle = i\hbar\dot{U}|\psi\rangle + Ui\hbar\partial_t|\psi\rangle = \left(i\hbar\dot{U} + UH\right)|\psi\rangle = \left(i\hbar\dot{U}U^\dagger + UHU^\dagger\right)|\bar{\psi}\rangle,$$

where a dot denotes the time derivative. This shows that the new Hamiltonian is given by

$$\begin{aligned}
\bar{H} &= i\hbar\dot{U}U^\dagger + UHU^\dagger = i\hbar\left(\frac{i}{\hbar}UH_0\right)U^\dagger + U(H_0+H_I)U^\dagger = UH_IU^\dagger \\
&= -e^{i\omega_g t|\mathrm{g}\rangle\langle\mathrm{g}|+i\omega_e t|\mathrm{e}\rangle\langle\mathrm{e}|}\left(\hbar\left(\Omega e^{-i\omega_L t}+\tilde{\Omega}e^{i\omega_L t}\right)|\mathrm{e}\rangle\langle\mathrm{g}| + \hbar\left(\tilde{\Omega}^* e^{-i\omega_L t}+\Omega^* e^{i\omega_L t}\right)|\mathrm{g}\rangle\langle\mathrm{e}|\right)e^{-i\omega_g t|\mathrm{g}\rangle\langle\mathrm{g}|-i\omega_e t|\mathrm{e}\rangle\langle\mathrm{e}|} \\
&= -e^{i\,0\,t|\mathrm{g}\rangle\langle\mathrm{g}|+i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|}\left(\hbar\left(\Omega e^{-i\omega_L t}+\tilde{\Omega}e^{i\omega_L t}\right)|\mathrm{e}\rangle\langle\mathrm{g}| + \hbar\left(\tilde{\Omega}^* e^{-i\omega_L t}+\Omega^* e^{i\omega_L t}\right)|\mathrm{g}\rangle\langle\mathrm{e}|\right)e^{-i\,0\,t|\mathrm{g}\rangle\langle\mathrm{g}|-i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|} \\
&= -e^{i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|}\left(\hbar\left(\Omega e^{-i\omega_L t}+\tilde{\Omega}e^{i\omega_L t}\right)|\mathrm{e}\rangle\langle\mathrm{g}| + \hbar\left(\tilde{\Omega}^* e^{-i\omega_L t}+\Omega^* e^{i\omega_L t}\right)|\mathrm{g}\rangle\langle\mathrm{e}|\right)e^{-i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|} \\
&= -\hbar\left(\Omega e^{-i\omega_L t}+\tilde{\Omega}e^{i\omega_L t}\right)e^{i\omega_0 t}|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\left(\tilde{\Omega}^* e^{-i\omega_L t}+\Omega^* e^{i\omega_L t}\right)|\mathrm{g}\rangle\langle\mathrm{e}|e^{-i\omega_0 t} \\
&= -\hbar\left(\Omega e^{-i\Delta t}+\tilde{\Omega}e^{i(\omega_L+\omega_0)t}\right)|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\left(\tilde{\Omega}^* e^{-i(\omega_L+\omega_0)t}+\Omega^* e^{i\Delta t}\right)|\mathrm{g}\rangle\langle\mathrm{e}| \ .
\end{aligned}$$

where $\Delta t = \omega_L - \omega_0$ which is the distance from resonance. I agree with U moving to the interaction picture from the Schrodinger picture. Interaction picture objects have overbars here. The upper isolated line of algebra seems easy to follow. In the first line of the second group of equations, we use $U = \exp(iH_0t/\hbar)$ so that U-dot = $i/\hbar$ * U * H_0 where H_0 could be written on either side. On this first line, then, the first two terms cancel leaving just U H_I $U^\dagger$ which is really $\bar{H}_I$, the interaction Hamiltonian in the interaction picture. This says that in the interaction picture, the H_O piece is completely gone, we have tuned it out, and the entire Hamiltonian is just the interaction part $\bar{H}_I$. On the next three lines, he is just using our previous H_O expansion in the exponential, where one E_k value is 0. In the second last line, the expo operator is moved forward onto |e> on the left, and backward to <e| on the right, replacing the expo operator with a phasor number as shown. The final line is then obvious.

The next step is to assume that the fast rotating terms average away over the time of an experiment, so we delete these two terms to get [where RWA means rotating-wave-approximation]

$$\bar{H}_{\mathrm{RWA}} = -\hbar\Omega e^{-i\Delta t}|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\Omega^* e^{i\Delta t}|\mathrm{g}\rangle\langle\mathrm{e}|.$$

The discarded terms have phase opposite the main terms (see bottom of equation block above), so were called "counter rotating wave" terms. The final step is to take this back from the interaction picture to the Schrodinger picture:

$$\begin{aligned}
H_{I,\mathrm{RWA}} = U^\dagger\bar{H}_{\mathrm{RWA}}U &= e^{-i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|}\left(-\hbar\Omega e^{-i\Delta t}|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\Omega^* e^{i\Delta t}|\mathrm{g}\rangle\langle\mathrm{e}|\right)e^{i\omega_0 t|\mathrm{e}\rangle\langle\mathrm{e}|} \\
&= -\hbar\Omega e^{-i\Delta t}e^{-i\omega_0 t}|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\Omega^* e^{i\Delta t}|\mathrm{g}\rangle\langle\mathrm{e}|e^{i\omega_0 t} \\
&= -\hbar\Omega e^{-i\omega_L t}|\mathrm{e}\rangle\langle\mathrm{g}| - \hbar\Omega^* e^{i\omega_L t}|\mathrm{g}\rangle\langle\mathrm{e}|.
\end{aligned}$$

which is similar to the previous algebra in the way things work. Note that ωL is the frequency of the applied E field and $\Omega = \mathbf{d_{eg}} \bullet \mathbf{E_0}/\hbar$ which is a number, possibly complex. The author then adds:

"A common first step beyond this is to remove the remaining time dependence in the Hamiltonian via another unitary transformation."

Now that I have done all this, I conclude that the RWA is NOT the same thing as the secular approximation Levitt is talking about, so I was misled by early web quotes above! Levitt's thing throws out off diagonal matrix elements of the interaction Hamiltonian. This seems a reasonable thing to do if you are applying photons that have only a small energy the size of the splits in the degeneracy levels he talks about.

Well OK, in Levitt's appendix 7.15 he is throwing out off diagonal interaction hamiltonian terms because they are widely separated in energy, which means the photon frequency separating these gaps is

large, so he is in effect "throwing out high frequency stuff" which is a little like what we did in the RWA above.

The better justification I think is this: you do systematic perturbation theory as outlined, for example, in Condon and Shortley which I have never read, but I dimly recall being taught the method in some class. You get expressions for things like transition rates in terms of sums where you have energy differences in the denominator. States in the sum which are close together will contribute most of the sum due to the small denominators, so you can approximate by throwing out contributions from far apart levels. This is what Levitt means by his "secular approximation".

In fact just this perturbation theory is discussed page 245 of Schiff and you see those denominators just mentioned. I do not plan to side track off on that right now.

Hmmm. I now see that the word "secular" has, as one of its meanings, the notion of something happening over a long time span -- something with a long period, as opposed to something fast with a short period. From OED:

L. *sæculaŽris,* f. *sæcul-um* generation, age, in Christian Latin 'the world', esp. as opposed to the church

Related to French derivative siecle which means a century. Not clear to me now this word got to mean worldly in Christian usage. Perhaps worldly means subject to the generations, whereas other-worldly means eternal. In the physics literature, you see the phrase "secular motion" used in reference to long-term motions, with fast motions filtered out. So that is the general idea of any **"secular approximation"**. You are filtering out high frequency information as your approximation. That is done in both Levitt's case, and in the RWA discussed above.

In fact, I think Levitt did the RWA implicitly on page 179 without calling it RWA.

****

Question: What is the Hamiltonian describing two interacting magnetic dipoles?

Answer: The result is given in Levitt page 204, a familiar result but I cannot quickly locate it in any of my books or on the web. So here is a quick derivation. Put dipole 1 at the origin $\mathbf{r}_1 = 0$, put dipole 2 at vector $\mathbf{r}_2$, and let point $\mathbf{r}$ be close to $\mathbf{r}_2$. Dipole 1 makes a magnetic field $\mathbf{B}_1$ at $\mathbf{r}_2$ which we shall calculate, and then the answer to the question is $H = -\boldsymbol{\mu}_2 \bullet \mathbf{B}_1$. The answer will be symmetric in 1 and 2, and each dipole feels the field of the other dipole and is "split" by it.

According to B&B page 146, the vector potential from a dipole $\mathbf{m}_1$ at the origin is given by

$$\mathbf{A}_1 = (\mu_0/4\pi)\, \mathbf{m}_1 \times \mathbf{r} \,/\, r^3 \qquad \text{and} \qquad \mathbf{B}_1 = \nabla \times \mathbf{A}_1$$

Therefore $$\mathbf{B}_1 = (\mu_0/4\pi)\, \nabla \times [\, \mathbf{m}_1 \times (\mathbf{r} \,/\, r^3)\,]$$

We use the vector identity: $$\nabla \times (\, \mathbf{a} \times \mathbf{b}\,) = \mathbf{a}\,(\nabla \bullet \mathbf{b}) - \mathbf{b}\,(\nabla \bullet \mathbf{a}) + (\mathbf{b} \bullet \nabla)\,\mathbf{a} - (\mathbf{a} \bullet \nabla)\,\mathbf{b}$$

with $\mathbf{a} = \mathbf{m}_1$ and $\mathbf{b} = (\mathbf{r}/r^3)$. Since $\mathbf{m}_1$ is not r-dependent, the middle two terms vanish so we have

$$\mathbf{B}_1 = (\mu_0/4\pi)\, [\, \mathbf{a}\,(\nabla \bullet \mathbf{b}) - (\mathbf{a} \bullet \nabla)\,\mathbf{b}] = (\mu_0/4\pi)\, [\, \mathbf{m}_1\,(\nabla \bullet (\mathbf{r}/r^3) - (\mathbf{m}_1 \bullet \nabla)\,(\mathbf{r}/r^3)\,] = 1 + 2$$

Notice that the first term involves a divergence and the second a gradient of a vector. It seems easiest to write this out in components, so

$$\mathbf{B_1} = (\mu_0/4\pi)\ [\ \mathbf{m_1}\ \partial_i\,(r_i/r^3)\ -\ m_{1i}\,\partial_i\,(\ \mathbf{r}/r^3)\ ]\quad = 1 + 2$$

We now compute

$$\partial_i\,(r_i/r^3) = [\ r^3\ (\partial_i r_i) - r_i\ \partial_i(r^3)\ ]/r^6$$

But: $(\partial_i r_i) = 3$ and $(\partial_i r) = 1/2r\ *\ 2r_i = r_i/r$ so $\partial_i(r^3) = 3r^2(\partial_i r) = 3r^2(r_i/r) = 3rr_i$

Therefore: $\partial_i\,(r_i/r^3) = [\ r^3\ (\partial_i r_i) - r_i\ \partial_i(r^3)\ ]/r^6 = [\ r^3\ (3) - r_i\ (\ 3rr_i)\ ]/r^6 = [\ r^3\ (3) - 3r^3)\ ]/r^6 = 0$

Thus, our first term vanishes because $\partial_i\,(r_i/r^3) = 0$, the vector $\mathbf{r}/r^3$ has no divergence. This is true only away from the origin, and it has $\delta(\mathbf{r} - \mathbf{0})$ behavior at the origin, but we only care about $\mathbf{r}$ near $\mathbf{r_2}$.

We next compute

$$\partial_i\,(\ \mathbf{r}/r^3) = [\ r^3\ \ (\partial_i \mathbf{r}) - \mathbf{r}\ \partial_i(r^3)\ ]/r^6$$

But: $(\partial_i \mathbf{r}) = \hat{\mathbf{r}}_\mathbf{i}$ // for example, $\hat{\mathbf{r}}_\mathbf{1} = \hat{\mathbf{x}}$ and $\partial_i(r^3) = 3rr_i$ // from above

so: $\partial_i\,(\ \mathbf{r}/r^3) = [\ r^3\ \ (\partial_i \mathbf{r}) - \mathbf{r}\ \partial_i(r^3)\ ]/r^6 = [\ r^3\ \hat{\mathbf{r}}_\mathbf{i}\ - \mathbf{r}\ 3rr_i\ \ ]/r^6 = [\ \ \hat{\mathbf{r}}_\mathbf{i}\ \mathbf{-\ 3}\ \hat{\mathbf{r}}\ r_i/r\ ]\ /\ r^3$

In this last result, $\hat{\mathbf{r}}_\mathbf{i}$ refers to something like $\hat{\mathbf{x}}$, but $\hat{\mathbf{r}}$ is a unit vector pointing toward point $\mathbf{r}$. So:

$$\mathbf{B_1} = (\mu_0/4\pi)\ [\ 0\ -\ \ m_{1i}\,\partial_i\,(\ \mathbf{r}/r^3)] = -\ (\mu_0/4\pi)\ m_{1i}\ \ [\ \ \hat{\mathbf{r}}_\mathbf{i}\ \mathbf{-\ 3}\ \hat{\mathbf{r}}\ r_i/r\ ]\ /\ r^3$$
$$= -\ \ (\mu_0/4\pi)\ [\ \mathbf{m_1} - 3\ (\mathbf{m_1}\bullet\hat{\mathbf{r}})\ \hat{\mathbf{r}}\ ]\ /\ r^3$$

This then is the magnetic field generated by the dipole 1 at the origin, as viewed from any location $\mathbf{r}$. We now imagine setting $\mathbf{r} = \mathbf{r_2}$, the location of dipole 2, but we will still refer to the unit vector pointing from dipole 1 to dipole 2 (toward $\mathbf{r_2}$) to be $\hat{\mathbf{r}}$, and the distance between the dipoles as r.

$$H = -\mathbf{m_2}\bullet\mathbf{B_1}\quad = -\ (\mu_0/4\pi r^3\ )\ [\ 3\ (\mathbf{m_1}\bullet\hat{\mathbf{r}})\ (\mathbf{m_2}\bullet\hat{\mathbf{r}})\ \ -\ (\mathbf{m_1}\bullet\mathbf{m_2})\ ]$$

Since the result is a scalar, it is true for any location (and orientation) of the dipoles. The direction of $\hat{\mathbf{r}}$can be taken in either way since it appears quadratically. The final step is to replace $\mathbf{m_i} = \boldsymbol{\mu}_\mathbf{i} = \gamma_i\ \mathbf{I_i}$ where $\mathbf{I}_i$ is the spin of each dipole. We then get

$$H = -\mathbf{m_2}\bullet\mathbf{B_1}\quad = \{\ -\ (\mu_0/4\pi)\ \gamma_1\gamma_2\ /\ r^3\ \}[\ 3\ (\mathbf{I_1}\bullet\hat{\mathbf{r}})\ (\mathbf{I_2}\bullet\hat{\mathbf{r}})\ \ -\ (\mathbf{I_1}\bullet\mathbf{I_2})\ ]$$

where the each I spin still has the units of $\hbar$. If we take these out, we add $\hbar^2$ to the {} constant. This is the result Levitt gives on page 204, but he only puts one power of $\hbar$ instead of 2, a mistake I think. ******

Notice this would be the same result for H = $-\mathbf{m_1 \bullet B_2}$. Each spin sees this same interaction Hamiltonian.

****

Question: What is the angular average of a 3x3 matrix?

Answer: The only way I can think of to do this is to compute $\langle RAR^{-1} \rangle$ where R is a general Euler angle rotation, and then somehow average this over angles. I have not solved this problem, but based on Levitt page 212 I think the answer is that the average is a constant times the identity matrix where the constant is 1/3 the sum of the diagonal elements of the original matrix. I imagine this problem shows up in lots of situations where you want an isotropic average of something. Maybe I will see a solution of this problem some day when I am reading on another subject. I did not see a web solution with a quick search.

Well, I just found a derivation in tensors.pdf which I have saved. An "isotropic tensor" is a 3x3 matrix which is the same no matter how axes are rotated. The *only* such tensor is a constant times the identity matrix. Thus, we know that $\langle A \rangle = \alpha I$ so the only problem is to find α. In doing our average we are going to look at matrices of the form RAR^{-1} . But $tr(RAR^{-1}) = tr(A)$ since cyclic, so all the matrices we are going to average have the same trace which is of course $A_{xx} + A_{yy} + A_{zz}$. Thus, this must be the value of the trace of the averaged matrix $\langle A \rangle$, But this is then equal to $tr(\alpha I) = \alpha tr(I) = 3\alpha$. Therefore, the average matrix is $\langle A \rangle = (1/3) [A_{xx} + A_{yy} + A_{zz}]$ as guessed above.

A lot easier than messing with Maple!

****

Comparison of Schiff and Levitt on the Density Operator

Schiff always seems to have everything. He does the density operator on page 378. He comments on the idea in classical physics of a region of phase space and the idea of how a quantum state might be blurred out into this phase space region and that is somehow the role of the quantum theory density operator.

Equation (42.1) tells you about the expectation value of some variable Ω in state α, but expresses it in an odd form where you project out of each i,j state pair the part belonging to α and you add these i,j contributions up with an implied i,j summation. The right side of (42.1) shows that (42.2) is true, and you have your original expectation value as $trace(P_\alpha \Omega)$ where $P_\alpha = |\alpha\rangle\langle\alpha|$ is a **projection operator**. We are at this point dispatched back to page 166 of Schiff where projection operators were introduced. Notice that $tr(P_\alpha) = \Sigma_i \langle i|\alpha\rangle\langle\alpha|i\rangle = \langle\alpha|\alpha\rangle = 1$.And also $P_\alpha^2 = |\alpha\rangle\langle\alpha|\alpha\rangle\langle\alpha| = |\alpha\rangle 1 \langle\alpha| = P_\alpha$. Schiff goes on to write $i\hbar\partial_t P_\alpha = [H, P_\alpha]$ as the equation of motion. So far, this discussion has nothing to do with statistics.

The next step is to assume that you have some kind of *mixture of states* with random phase differences between the states (whatever that means). If, in your mixed state, the probability of being in state α is p_α, then you can define a **density matrix operator** $\rho = \Sigma_\alpha |\alpha\rangle p_\alpha \langle\alpha| = \Sigma_\alpha p_\alpha P_\alpha$. This is the first time a sum over α has appeared in this section. He says the α states are orthonormal but they don't need to be complete, but I will think of this α as labeling the states in some "mixture of states".

Now consider:

$$tr(\Omega\rho) = \Sigma_i < i \mid \Omega\rho \mid i > = \Sigma_i < i \mid \Omega \left[\Sigma_\alpha \mid\alpha\rangle p_\alpha \langle\alpha \right] \mid i > = \Sigma_\alpha\, p_\alpha \Sigma_i < i \mid \Omega \mid\alpha\rangle\langle\alpha \mid i >$$

$$= \Sigma_\alpha\, p_\alpha \Sigma_i \langle\alpha \mid i \rangle\langle i \mid \Omega \mid\alpha\rangle = \Sigma_\alpha\, p_\alpha \langle\alpha \mid \Omega \mid \alpha\rangle = \langle\Omega\rangle$$

So this shows that in a "mixed state" described by ρ, you have $\langle\Omega\rangle = tr(\Omega\rho)$. Note that this is not the same as expectation in state α, it is in the mix of states that are labeled by index α.

So, we started with the idea of projection operator P_α for each state α, and we ended up with the idea of a density operator ρ describing a mixture of states. And $i\hbar\partial_t\rho = [H,\rho]$. Everything follows over from P_α to ρ.

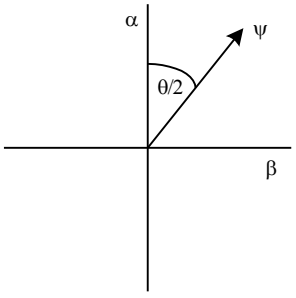
Now let's review what Leavitt does on page 274. We are restricted here to the spin-1/2 world. In equation (10.3), the state $|\psi\rangle$ is a specific system state characterized by c_α and c_β. But then he says to construct a mixed state that is the average of states 1,2,3....N where the probability of being in any of these states is $p_\alpha = 1/N$. Then top page 275 really says $\rho = \Sigma_\alpha\, |\alpha\rangle p_\alpha \langle\alpha|$ and (10.5) is then just a fine way to say this. We know from Schiff above that $\langle\Omega\rangle = tr(\Omega\rho)$ and hence we get Malcolm's (10.6). This is the average of the expectation value of some operator in your many states. It is the ensemble average.

Now let's think of our ρ operator as a 2x2 matrix. Then $\langle i|\rho|j\rangle = \Sigma_\alpha \left[\langle i| \alpha\rangle\, p_\alpha \langle\alpha| j\rangle = \Sigma_\alpha\, p_\alpha\, c_i^\alpha\, c_j^{\alpha\,\bullet} \right.$ which you would write as $\langle c_i\, c_j^{\,\bullet} \rangle$ and this is Malcolm's (10.7), just fine. Now he calls the states α and β instead of 1 and 2, so the Schiff α now has a different meaning in Leavitt.

So I am completely happy with Leavitt's and Schiff's density matrix equations, there is no difference as I once thought there was.

Question: Why do the rotations for j = 1/2 have half angles in them?

The space is of course 2 dimensional. Can think of (01) and (10) as your basis vectors. Lets call these $|\alpha\rangle$ and $|\beta\rangle$. I think this would be a reasonable picture, but only in the case of real mixtures of the base states:



We know that $\psi = R(\theta)\, |\alpha\rangle = C_{\theta/2}\, |\alpha\rangle + S_{\theta/2}\, |\beta\rangle$ The angle that appears in your $R(\theta)$ is not the same as the angle that appears in the picture! If $\theta = 180$, then $C_{\theta/2} = \cos(90) = 0$ and $R(180)\, |\alpha\rangle = |\beta\rangle$. The interesting fact is that $R(360)\, |\alpha\rangle = -|\alpha\rangle$ and that is just the way this "spinor" representation works.

The spin vector **J** for j=1/2 is still a vector in 3D space, however. We can apply the usual $R_{3Dz}(\theta)$ rotation to **J** and then we have rotation matrices which do NOT have half angles. In particular, we know that $R_{3Dz}(180)\, \mathbf{J} = -\mathbf{J}$. So in Levitt's picture on page 241, the dark arrows shown in the two pictures are

really these **J** vectors, now some quantum state ψ. The vector **J** is quantized in QM and can only have the two positions shown if you select the z axis say by a magnetic field. Each position is labeled, so really the up arrow is $\langle\alpha|\mathbf{J}|\alpha\rangle = (1/2)\,\hat{\mathbf{z}}$ and the lower arrow is $\langle\beta|\mathbf{J}|\beta\rangle = -(1/2)\,\hat{\mathbf{z}}$. The diagonal elements of the other J components vanish.

****

Question: Why are the rotation matrices orthogonal?

This is true because a rotation does not change the length of a vector. Here is the fast proof:

$$v'^{T}v' = (Rv)^{T}(Rv) = v^{T}R^{T}Rv = v^{T}v \qquad \text{so} \qquad R^{T}R = 1$$

And here is the slower proof. Suppose we have $x'_i = R_{ij}x_j$. Then write

$$x'_i\,x'_i = (R_{ij}x_j)\,(R_{ik}x_k) = (R_{ij}R_{ik})x_jx_k \quad \text{must be} \quad (\delta_{jk})\,x_jx_k = x_jx_j$$

Thus we must have $R_{ij}R_{ik} = \delta_{jk}$. But this says $R^{T}_{ji}R_{ik} = \delta_{jk}$ or $R^{T}R = 1$. But this last result says that $R^{T} = R^{-1}$ which defines an "orthogonal matrix".

Why is this word "orthogonal" used ? Consider

$$R_{ij} = (\mathbf{c}_j)_i \qquad \text{where } \mathbf{c}_j \text{ is the } j^{th} \text{ column vector of the matrix } R_{ij}$$

We examine the elements of this column vector by stepping down the row index i. So we have

$$\delta_{jk} = R_{ij}R_{ik} = (\mathbf{c}_j)_i\,(\mathbf{c}_k)_i = \mathbf{c}_j \bullet \mathbf{c}_k = \delta_{jk}$$

Thus, the column vectors formed from the matrix are in fact **orthonormal**, which includes orthogonal. Some writers have claimed such matrix should be called "orthonormal".

The rows are NOT orthonormal, to wit:

$$R_{ij} = (\mathbf{r}_i)_j \qquad \text{where } \mathbf{r}_i \text{ is the } i^{th} \text{ row vector of the matrix } R_{ij}$$

We examine the elements of this row vector by stepping across the column index of R_{ij}. Then

$$\delta_{jk} = R_{ij}R_{ik} = (\mathbf{r}_i)_j(\mathbf{r}_i)_k$$

but this does not say the rows are orthonormal. The rows of R^{T} would be orthonormal.

The orthogonal matrices make up the group O(n), that is what the O stands for.

Question: What is the corresponding statement for a Lorentz Transformation Matrix?

We have $g_{\mu\nu}x^{\mu}x^{\nu}$ as the "length" of a vector, and this is what gets preserved. So consider

$x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu}$ => $g_{\mu\nu} x'^{\mu} x'^{\nu} = g_{\mu\nu}(a^{\mu}{}_{\alpha} x^{\alpha})(a^{\nu}{}_{\beta} x^{\beta}) = g_{\alpha\beta} x^{\alpha} x^{\beta}$

Therefore, we must have that

$g_{\mu\nu}\, a^{\mu}{}_{\alpha}\, a^{\nu}{}_{\beta} = g_{\alpha\beta}$

We can tilt the indices on a the other way like so:

$a^{\nu}{}_{\beta} = g^{\nu\nu'} g_{\beta\beta'}\, a_{\nu',}{}^{\beta'}$

Insert this on the left side to get

$g_{\mu\nu}\, a^{\mu}{}_{\alpha}\, a^{\nu}{}_{\beta} = g_{\mu\nu}\, a^{\mu}{}_{\alpha}\, (g^{\nu\nu'} g_{\beta\beta'}\, a_{\nu',}{}^{\beta'}) = \delta_{\mu}{}^{\nu'}\, a^{\mu}{}_{\alpha}\, g_{\beta\beta'}\, a_{\nu',}{}^{\beta'} = a^{\mu}{}_{\alpha}\, g_{\beta\beta'}\, a_{\mu}{}^{\beta'}$

Now we have $a^{\mu}{}_{\alpha}\, g_{\beta\beta'}\, a_{\mu}{}^{\beta'} = g_{\alpha\beta}$, so apply $g^{\beta\kappa}$ to both sides so RHS becomes $\delta_{\alpha}{}^{\kappa}$. Then we have

$\delta_{\alpha}{}^{\kappa} = a^{\mu}{}_{\alpha}\, g_{\beta\beta'}\, a_{\mu}{}^{\beta'}\, g^{\beta\kappa} = a^{\mu}{}_{\alpha}\, a_{\mu}{}^{\beta'}\, \delta_{\beta',}{}^{\kappa} = a^{\mu}{}_{\alpha}\, a_{\mu}{}^{\kappa}$

So we then end up with the way this result is presented without any g's, namely

$a^{\mu}{}_{\alpha}\, a_{\mu}{}^{\kappa} = \delta_{\alpha}{}^{\kappa}$ or $a_{\mu}{}^{\kappa}\, a^{\mu}{}_{\alpha} = \delta^{\kappa}{}_{\alpha}$ as in BD vol 1 page 16 equation 2.3.

Here is another fact which we know

$(a^{-1})^{\kappa}{}_{\mu}\, a^{\mu}{}_{\alpha} = \delta^{\kappa}{}_{\alpha}$

Compare to the right result above to conclude that

$(a^{-1})^{\kappa}{}_{\mu} = a_{\mu}{}^{\kappa} = (a^{T})_{\kappa}{}^{\mu} = g_{\kappa\kappa'} g^{\mu\mu'} (a^{T})^{\kappa'}{}_{\mu'}$ $\neq>$ $(a^{-1})^{\kappa}{}_{\mu} = (a^{T})^{\kappa}{}_{\mu}$

So in general a LT is not orthogonal and is not unitary. If you fiddle or see TK sheets, you see that pure rotations are unitary, while pure boosts are hermitian. Notice above the "meaning" of applying the T superscript. You have some kind of matrix $a_{\mu}{}^{\kappa}$ with some up/down index positions, and you write a picture of the matrix. Transpose means swapping the rows and columns, and that means swapping the indices without changing their positions up/down, hence $(a^{T})_{\kappa}{}^{\mu} = a_{\mu}{}^{\kappa}$. This does "mess up" the covariance sense of the indices, however, and probably people don't use a^{T} for this reason.

Question: Why can you reverse the tilt of a pair of summed indices without changing the result?

Suppose we have some general thing like $R_{\alpha\beta\delta\ldots} R^{\alpha\beta\delta\gamma\ldots}$. We know that

$R_{\alpha}{}^{\beta}{}_{\delta\gamma\ldots} = g^{\beta\beta'} R_{\alpha\beta'\delta\ldots}$ and $R^{\alpha}{}_{\beta}{}^{\delta\gamma\ldots} = g_{\beta\beta'}\, R^{\alpha\beta'\delta\gamma\ldots}$

Therefore

$$R_{\alpha}{}^{\beta}{}_{\delta\gamma\ldots}\, R^{\alpha}{}_{\beta}{}^{\delta\gamma\ldots} = g^{\beta\beta'} R_{\alpha\beta'\delta\ldots}\, g_{\beta\beta''} R^{\alpha\beta''\delta\gamma\ldots} = \delta^{\beta'}{}_{\beta''} R_{\alpha\beta'\delta\ldots} R^{\alpha\beta''\delta\gamma\ldots} = R_{\alpha\beta\delta\ldots} R^{\alpha\beta\delta\gamma\ldots}$$

So the answer is this: if you reverse the "tilt" of a pair of summed indices, you incur a summed pair of g matrices which reduce to a Kronecker delta and give the claimed result.

Question: Why does d/dx^{μ} transform covariantly?

We know that $x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu}$ is the usual contravariant transformation. We can then write by the chain rule

$$df/dx'\, dx'/dx = df/dx \qquad \text{or} \qquad df/dx'^{\mu}\, dx'^{\mu}/dx^{\nu} = df/dx^{\nu}$$

But this says $df/dx'^{\mu}\, a^{\mu}{}_{\nu} = df/dx^{\nu}$ which just then be written $\partial'_{\mu} f\ a^{\mu}{}_{\nu} = \partial_{\nu} f$ or

$$\partial'_{\mu}\, a^{\mu}{}_{\nu} = \partial_{\nu}$$

Question Revisited: What is the corresponding statement for a Lorentz Transformation Matrix?

Now that we know the "reverse tilt rule" let's start again on this one:

We have $g_{\mu\nu} x^{\mu} x^{\nu}$ as the "length" of a vector, and this is what gets preserved. So consider

$$x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu} \qquad \Rightarrow \qquad g_{\mu\nu} x'^{\mu} x'^{\nu} = g_{\mu\nu} (a^{\mu}{}_{\alpha} x^{\alpha})\, (a^{\nu}{}_{\beta} x^{\beta}) = g_{\alpha\beta} x^{\alpha} x^{\beta}$$

Therefore, we must have that

$$g_{\mu\nu}\, a^{\mu}{}_{\alpha}\, a^{\nu}{}_{\beta} = g_{\alpha\beta}$$

Write this as

$$g_{\alpha\beta} = a_{\nu\alpha}\, a^{\nu}{}_{\beta} = a^{\nu}{}_{\alpha}\, a_{\nu\beta}$$

Apply $g^{\beta\beta'}$ to goth sides

$$\delta_{\alpha}{}^{\beta'} = a^{\nu}{}_{\alpha}\, a_{\nu}{}^{\beta'}$$

which is then a "faster route" to the result of interest.

Let's now do this little proof as fast as we can:

$$x'.x' = x'^{\mu}\, x'_{\mu} = a^{\mu}{}_{\nu} x^{\nu} * a_{\mu}{}^{\sigma} x_{\sigma} = a^{\mu}{}_{\nu} a_{\mu}{}^{\sigma} x^{\nu} x_{\sigma} \ =?= x.x$$

so must have $a^{\mu}{}_{\nu} a_{\mu}{}^{\sigma} = \delta_{\nu}{}^{\sigma}$ and then RHS becomes x.x. In this way, we get the "sum on first index" form of this relation.

Now let's invert our $x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu}$ using the rule we just found. Thus

$x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu} \Rightarrow x'^{\mu} a_{\mu}{}^{\sigma} = a^{\mu}{}_{\nu} x^{\nu} a_{\mu}{}^{\sigma} = \delta_{\nu}{}^{\sigma} x^{\nu} = x^{\sigma} \quad \Rightarrow \quad x^{\sigma} = a_{\mu}{}^{\sigma} x'^{\mu}$

Thus we have shown that

$x'^{\mu} = a^{\mu}{}_{\nu} x^{\nu}$ and inverse is $x^{\mu} = a_{\nu}{}^{\mu} x'^{\nu}$

sum on 2nd sum on 1st

Now repeat the above process to get

$x.x = x^{\mu} x_{\mu} = a_{\nu}{}^{\mu} x'^{\nu} * a^{\sigma}{}_{\mu} x'_{\sigma} = a_{\nu}{}^{\mu} a^{\sigma}{}_{\mu} x'^{\nu} x'_{\sigma}$

and this shows that we must have the rule

$a_{\nu}{}^{\mu} a^{\sigma}{}_{\mu} = \delta_{\nu}{}^{\sigma}$ // sum on second rule.

So here are the two rules side by side:

$a^{\mu}{}_{\nu} a_{\mu}{}^{\sigma} = \delta_{\nu}{}^{\sigma} = a_{\mu\nu} a^{\mu\sigma}$ // sum on first index

$a_{\nu}{}^{\mu} a^{\sigma}{}_{\mu} = \delta_{\nu}{}^{\sigma} = a_{\nu\mu} a^{\sigma\mu}$ // sum on second index

See page 881 Messiag Vol II for confirmation.

****

Question: Particle as Poincare Group Representation

We sometimes thing of a "particle" as having the Casimir eigenvalues of the Poincare Group. For example, one of the Hamiltonian symmetries is the p^{μ} operator so $[H,p^{\mu}] = 0$, so we feel we can "label" our solutions with $p.p = m^2$ and so this is the particle mass. What are the other casimirs of the PG? We suspect somehow "spin" is one, so a particle has a definite spin value such as ½ for an electron. Are we talking about a point particle, or a spread-out particle here? Clarify. TBC

****