Phys 326 Discussion 12 – GR, SR, & The Global Positioning System

In lecture we introduced **General Relativity** (GR), Einstein's geometric theory of gravity. GR has been extremely successful in describing 100 years worth of experimental tests and reveals that Newtonian gravity, $F = -GM/r^2$, is an approximation for the case of weak gravitational fields = situations where the dimensionless combination $GM/(c^2 r) \ll 1$. GR also incorporates **Special Relativity** (SR) in its very fabric. This overcomes the other limitation of Newtonian mechanics: that $F = ma$, $KE = \frac{1}{2} mv^2$, and the other formulae of traditional mechanics are approximations for slow speeds = situations where the dimensionless combination $v/c \ll 1$.

At the heart of GR are **proper time**, τ, and its close cousin, **proper distance**, σ. The significance of these quantities as "watch time" and "ruler distance" comes from SR, and was explored in PHYS 225. If you would like a refresher, see the Appendix. In lecture, we traced the development of GR via a series of thought experiments, starting with the **Equivalence Principle**. The result is a set of axioms that summarize the principal content of GR; we haven't talked about them all yet, but here they are:

(1) The **effect of mass** on the universe is to create a **curvature in the spacetime surrounding the mass**. This curvature is summarized by the spacetime **metric** for $d\tau$.

(2) The metric in the region of spacetime outside a non-rotating spherical mass M is the **Schwarzschild metric**:

$$d\tau = \sqrt{dt^2\left(1-2M/r\right) - \frac{dr^2}{\left(1-2M/r\right)} - r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)}$$

This expression is written in **natural units** where both M and t are expressed in meters, just like r, using the physical constants c and G: $M \equiv M_{kg}\, G/c^2$ and $t \equiv t_{\text{sec}}\, c$. NOTE: The Schwarzschild metric is not an axiom of GR, it is a result of the **Einstein field equations**, which are addressed in a graduate GR class. The metric and the field equations are analogous to Coulomb's Law and Maxwell's equations in E&M: Coulomb's Law for the E-field of a static charge is a *solution* of Maxwell's equations, but you *learn* Coulomb's Law *first* to understand what an electric field E actually is (!) and how it behaves.

(3) The **dynamics** of GR are summarized by this elegant axiom:

The paths of particles due to gravity are **geodesics** of the metric, i.e. the "straight lines" that extremize the proper time interval $\int d\tau$ – the *distance* – between two fixed spacetime endpoints.

Geodesics in 3D space extremize the spatial distance $\int dl$ between two fixed endpoints. To get unique geodesics in 3D space we must find the *minimum* distance, hence: "a straight line is the path of shortest distance between two points". In 4D spacetime, it turns out that unique geodesics are obtained by *maximizing* total proper time. This is sometimes called the **Principle of Maximal Aging**: the path taken by a particle to get from one fixed event to another is the one that *maximizes* the total proper time along the path. Thus:

☞ Gravity isn't a force at all; mass warps spacetime, and everything just travels in straight lines.

Problem 1 : The GPS System and the Weak-Field, Slow-Speed Schwarzschild Metric

Today we will explore a famous example where both GR and SR are required in an *engineering* situation: the Global Positioning System (GPS). The GPS consists of a network of 32 satellites that orbit the earth in 12-hour orbits (see wikipedia: GPS for a lovely animated picture of this satellite network.) Each satellite regularly sends out radio signals that record the satellite's current time and position. A GPS receiver on earth collects signals from 3-4 satellites. Knowing the time and position at which each signal was sent, the local time on earth, and the fact that the signals travelled at the speed of light, the receiver can figure out roughly where it is: it is somewhere on a spherical surface of radius $c(t_\oplus - t_S)$ around the satellite's location. Here $t_\oplus$ is time on earth (when the signals were received) and t_S is time at each satellite (when the signal was sent). The receiver finally uses some overlapping-sphere geometry to determine its exact location.

The Problem: time on earth and time at the satellite cannot be directly compared! *Checkpoints*[1]

The satellite is moving relative to the earth's surface *and* it is at a different gravitational potential than the earth's surface. Any comparison between $t_\oplus$ and t_S thus requires *corrections* for both **Lorentz time dilation** and **gravitational time dilation**. We must calculate the size of these corrections! We will need some earth data: as usual,

- the radius of the earth is $R_\oplus = 6.4 \times 10^6$ m
- all appearances of the earth's mass $M_\oplus$ will be in the combination $GM_\oplus$, which is equal to $gR_\oplus^2$.

We learned in class how gravitational time dilation works qualitatively: "lower is slower". **Local time** elapses more **slowly** when you are **lower** in a gravitational field than when you are higher up. If you look up to the top of a very tall mountain, you see the people / plants / birds above you moving and aging at a faster rate than is normal; if you look down, everything at the bottom of a valley seems to be moving and aging at a slower rate than is normal. As our first exercise in using the Schwarzschild metric,

$$d\tau = \sqrt{dt^2\left(1-2M/r\right) - \frac{dr^2}{\left(1-2M/r\right)} - r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)} \qquad \text{in natural units:} \quad M \equiv M_{\text{kg}} G/c^2, \quad t \equiv t_{\text{sec}}\, c$$

we will use it to obtain the gravitational time correction. Here is how to read the metric to reproduce the effect:

- t is faraway time. We will also call it **Bookkeeper** time by inventing a mythical observer, "The Bookkeeper", who is located infinitely far from all gravitational fields and records the coordinates of events without taking any space-time distortions into account.
- τ is local time. It is proper time, of course, and we know from SR that proper time is the **wristwatch** time recorded by an observer who is *at* the events being measured (thus, *local* time).

(a) Take the Schwarzschild metric and factor out dt; the result will be an expression for $d\tau$ in terms of dt and the time-derivatives $\dot{r}, \dot{\theta}, \dot{\phi}$. Further, this entire calculation will take place in a single plane (we will work with one satellite at a time), so take $\theta = 90°$ for simplicity. That will kill off one term from your metric.

(b) Next switch from natural units to SI units: inject factors of G and c into your metric so that t and τ are in seconds and M is in kg. Shuffle your constants so that they are all on the right-hand side, giving $d\tau = \ldots$ Your metric now looks like this: $d\tau = dt\sqrt{1 - A - \dot{r}^2 C - r^2\dot{\phi}^2 B}$. We will be comparing local time intervals $d\tau_\oplus$ measured by the receiver's clock at the earth's surface to the intervals $d\tau_S$ measured by the satellite's clocks. Both clocks are sitting at fixed radii, so $\dot{r} = 0$; the metric we need for this problem therefore simplifies to $d\tau = dt\sqrt{1 - A - r^2\dot{\phi}^2 B}$. What are A and B when all quantities are expressed in SI units?

(c) As advertised, GR and SR are both going to make an appearance, but the effects are not *huge* ... we must estimate how big these effects are for the GPS system so that we can make some reasonable approximations. First, evaluate the dimensionless term A numerically for the two relevant radii: at $r = R_\oplus$ (earth's surface) and at $r = R_S$ (satellite's orbit). Start by finding R_S in terms of $R_\oplus$ using the fact that the GPS satellites are in circular orbits with 12-hour periods.

(d) Before we get to the next term, let's decide to neglect the rotation of the earth. (We must if we are going to use the Schwarzschild metric, because it only applies in the vicinity of *non-rotating* spherical masses M.) Make a quick check that this is a reasonable approximation: calculate the orbital velocity v_S of the satellite and the rotational speed $v_\oplus$ of the earth's surface. Hopefully you will find that $v_S \gg v_\oplus$ is a decent approximation.

---

[1] (b) $A = 2GM_\oplus/rc^2 = 2gR_\oplus^2/rc^2$ & $B = 1/c^2$ (c) $R_S = 4.2\,R_\oplus \longrightarrow A = 1.4 \times 10^{-9}$ (earth) & 3.3×10^{-10} (sat) (d) $v_S \approx 3{,}900$ m/s, $v_\oplus \approx 500$ m/s (e) 1.7×10^{-10} (f) ☺ (g) -0.84×10^{-10} (h) $+5.3 \times 10^{-10}$ (i) look at upper (gravity) & lower (speed) curves at $R_S = 27{,}000$ km (j) $\delta t \approx \delta d / c = 6.7$ nsec (k) 15 sec! (l,m,n) next question answers previous one

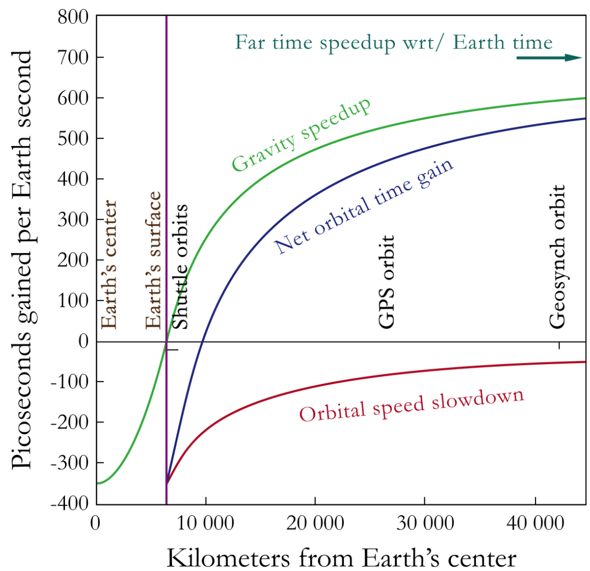
(e) Now evaluate the term $r^2\dot{\phi}^2 B$ numerically for the satellite using the speed v_S you found in (d).

(f) I hope you found that both A and $r^2\dot{\phi}^2 B$ are very small numbers at the two radii where we need them! A Taylor approximation is most certainly in order: apply a 1st-order Taylor approximation to the metric to get rid of that annoying square root in $d\tau = dt\sqrt{1 - A - r^2\dot{\phi}^2 B}$.

(g) Our metric is now in the form of two nice additive corrections that we must apply to the "Bookkeeper time" dt to obtain earth-surface time $d\tau_\oplus$ and satellite time $d\tau_S$. The "A" term is due to gravity (gravitational time dilation) while the "B" term is due to speed (the familiar Lorentz time dilation from special relativity). Since they are additive corrections, we can treat them one at a time. Let's do the **special relativity correction** first: calculate the fractional time-difference $(d\tau_S - d\tau_\oplus)/d\tau_\oplus$ caused by "B" term. Also, make sure you can recover the familiar *form* of Lorentz time-dilation from your expressions: that a moving clock ticks slower than a stationary one by a factor of $\gamma = 1/\sqrt{1-\beta^2}$.

(h) Now we come to the **gravitational correction**. Calculate the fractional time-difference $(d\tau_S - d\tau_\oplus)/d\tau_\oplus$ caused by "A" term. How does it compare to the Lorentz correction? Everyone expects that the exotic theory of General Relativity couldn't possibly have significant consequences for any engineering applications … you may be surprised by what you find! Bonus: for practice, show that $d\tau_S/d\tau_\oplus = 1 + (\Phi_S - \Phi_\oplus)/c^2$ where Φ is gravitational potential. This is the weak-gravity result $\Delta\tau_{TOP}/\Delta\tau_{BOT} \approx 1 + \Delta\Phi/c^2$ that we obtained from our Alice / Bob / photon-emitter thought experiment using the Doppler shift; you can now get it straight from the Schwarzschild metric. ☺

(i) The plot at right comes from Wikipedia and shows "picoseconds gained [on the satellite clocks] per earth second" as a function of the satellite's orbital radius. Check your work against this plot: use your two $(d\tau_S - d\tau_\oplus)/d\tau_\oplus$ values and the orbital radius of the GPS satellites to compare your findings to the values on this graph. Pay close attention to the signs of the two corrections: they are of *opposite* sign! If you didn't find that, please go back and debug!



(j) These are tiny corrections … are they *really* needed to make the GPS system work? Let's find out. GPS specifications quote a position accuracy of 2 m for military applications. (It is about 15 m for civilian GPS receivers.) To achieve a position error of at most $\delta d = 2$ m, what is the maximum error δt that we can make in our time measurements? Hints: Remember how the GPS receiver calculates positions: each satellite i reports its current location (x_i, y_i, z_i) and time t_i ... the receiver knows its own current time t ... so the receiver can deduce its own position (x, y, z) by figuring out how far it is from each satellite. The satellite signals travel at the speed of light, so the approximate relation between δt and δd is … <something really simple> (don't over-think it!). So: $\delta d = 2$ m requires a time accuracy of $\delta t = what$? If you're really stuck, jump to parts (l,m,n), then come back.

(k) That is some *serious* time accuracy! The combined $(d\tau_S - d\tau_\oplus)/d\tau_\oplus$ value you got from SR & GR is the fractional correction the earth receiver must make to the satellite times it receives to translate them into its own frame. Suppose the GPS system did not make these tiny corrections. Pick some moment when we reset all clocks in the system to zero. How much time must pass on earth before the time-error exceeds δt and the GPS system drifts out of spec?

➔ Message: the GR & SR corrections are ESSENTIAL for this system to work at all!

(l) How the GPS system actually works is quite interesting. The equations the receiver must solve to determine its own position (x,y,z) based on its own time t and the satellite data (t_i, x_i, y_i, z_i) it receives are

$$c^2(t-\bar{t}_i)^2 = (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 \quad \text{for each satellite signal } i \text{ received.}$$

I put a bar over the satellite times $\bar{t}_i$ to clarify that the GR & SR corrections have been applied to translate them into earth-surface times. How many such equations — i.e. how many satellite signals — are needed to solve for (x,y,z)?

(m) I bet you said three: 3 equations for 3 unknowns. That makes perfect sense ... except that all those $(x-x_i)^2$ terms are *squared*, meaning we lose sign information. It turns out you can solve this problem exactly with four satellites. To understand this, sketch the problem: the info from each satellite restricts the receiver's location to a *spherical surface* of radius $c(t-\bar{t}_i)$ around the satellite's position (x_i, y_i, z_i). How many such spherical surfaces do you need for their intersection to be exactly one point?

(n) Your sketching hopefully showed that three satellite signals will restrict the receiver's position to *two* possible points. Here, the GPS system takes a clever approach: when it has figured out the position down to two possible points, it assumes that the receiver is somewhere near the earth's surface and *picks the point closest to the earth's surface*. (The other solution is usually way off.) That brings us back down to only three satellites needed … and now the final subtlety. As it happens, the needed time accuracy $\delta t \approx 7$ nsec is much better than can be achieved by the clock on a handheld receiver of reasonable price. To solve this issue, the system adds back in a fourth satellite reading and this time solves the equations

$$c^2(t-\bar{t}_i)^2 = (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 \quad \text{with } i = 1,2,3,4$$

for four unknowns: the desired position (x,y,z) of the receiver *and* the inaccurately-measured time t at the receiver. It again resolves the discrepancy due to the squares using the clever strategy of picking the result closest to the earth's surface. Pretty clever. ☺ And no, this is not actually a question. ☺

Appendix: Proper Time, Proper Distance, and the Metric

(1) Proper Time: We know from SR that proper time, τ, is the one and only Lorentz-invariant (Lorentz-scalar) measure of the spacetime-distance between two events. When gravity is absent,

$$d\tau = \sqrt{dt^2 - (dx^2 + dy^2 + dz^2)}\,.$$

Here we are using **natural units**, where time and position are both measured in meters using the physical constant c (i.e. natural t = t-in-meters = $c \times t$-in-seconds). As we know from our relativity studies – and can anyway see immediately from the definition of $d\tau$! – the proper time interval $\Delta\tau$ between two events corresponds to the "wristwatch time" Δt on the personal clock of the *one* special observer who is *at both events*, i.e. for whom $\Delta x = \Delta y = \Delta z = 0$.

(2) Proper Distance: If we have spacelike separated events, i.e. events that are so far apart in space compared with their time separation that even an observer traveling at c could never be at both of them, the proper time between them would be imaginary, so we usually switch to proper distance, σ, defined as

$$d\sigma \equiv \sqrt{-d\tau} = \sqrt{-dt^2 + (dx^2 + dy^2 + dz^2)}\,.$$

This quantity can be interpreted as "ruler distance": $\Delta\sigma$ is the spatial distance $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ that would be measured if you laid down a ruler between two events and read off the events' positions *at the same instant in time*, i.e in the frame where $\Delta t = 0$.

(3) The Metric: The metric of any space is the relationship that translates changes in coordinates into changes in physical distance. The "physical distance" between two points in a space depends on the space you are talking about. Some examples:

- In normal "Euclidean" 3D space, physical distance is spatial distance dl. In Cartesian coordinates, the metric is $dl = \sqrt{dx^2 + dy^2 + dz^2}$. The motivation for this definition of physical distance is that it is a **scalar**, meaning that is invariant under both rotations and translations.

- Same space, different coordinate systems: the metric of 3D space in spherical coordinates is $dl = \sqrt{dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\, d\phi^2}$, while in cylindrical coordinates it is $dl = \sqrt{ds^2 + s^2 d\phi^2 + dz^2}$.
- In 4D spacetime, physical distance is $d\tau$ or $d\sigma$. Again, the reasons for these definitions are that they are **4-scalars**, i.e. invariant under all ten of the fundamental symmetries of 4D spacetime: rotations, translations, and boosts. In an empty region of spacetime (one free of masses), the metric is the **Minkowski metric** $d\tau = \sqrt{dt^2 - dl^2}$ of SR, where the spatial component dl can be expressed in any spatial coordinate system.