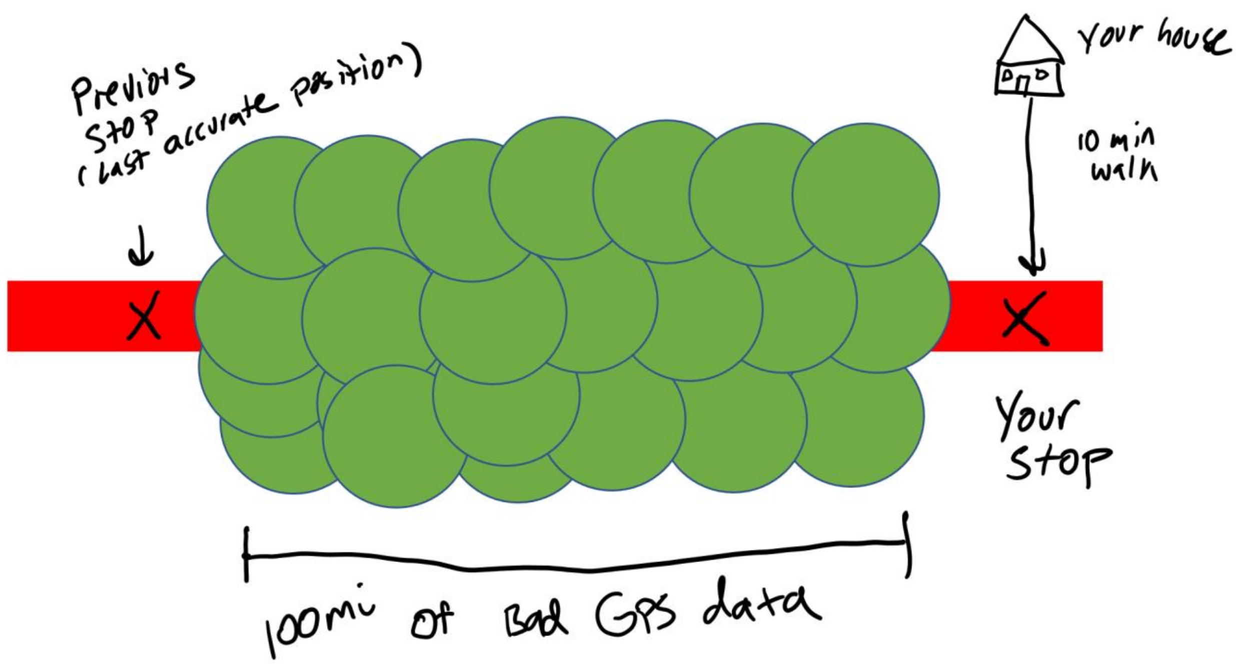
Train Stop

You just got a new job at a fancy tech start up 80 miles away from where you live. You would move, but you still have 6 months left on your lease. Luckily, there is a new high-speed train which connects you to your new place of work. You have found that it takes you 10 minutes to walk from your house to the train station. Wanting to leave your house at the latest possible time, you decide to make a train tracker you can reference in the morning to decide when to leave. You managed to find the real time GPS data for the train online, but the 100 miles between the previous stop and the stop where you catch the train runs mostly under dense tree cover so that the GPS measurements can be off by as much as 10 miles. You also find online that the train typically runs through the dense, sometimes treacherous, woodland with an average speed coming from a normal distribution with mean 100mph and standard deviation of 15 mph. You decide that you want to know where the train is 50 minuets after it enters the forest, typically 10 min from then until your station, that way you can decide to rush to the station or have another cup of coffee. Your goal will be to use what you know about the trains average speed through the woodland and its poor GPS location to estimate where the train actually is 50 min after it enters the woodland.

We are going to search for what we will call the analysis position, x^a. The analysis position will be our best guess for where the train is based on what we know about its average speed with a model like $d = vt$ in conjunction the poor resolution GPS data.



1) Clearly, we will want to use the uncertainties in the positions given by our model and the GPS measurements to find our optimal estimate. We have the uncertainty in GPS observation +/- 10 mi. Let's take that possible error as the standard deviation of a normal distribution with mean zero. What about the uncertainty in the model predicted distance? The distance predicted by

the model at time 50 min is, $x^f = \frac{100}{60}(50)$=83.33 miles (here f stands for forecast). Given what we know about the typical speeds the train goes through the forest, what can we do to estimate the error in the forecast at the 50 min mark? Discuss some of your ideas. There are a few ways one could estimate this!

2) Heading back on line you find that the train company actually gives some data on where the train is 50 min after it enters the forest since there is a rail switch at the 83.3 mile mark. You find that the position of the train at 50 min comes from a normal distribution with mean 83.333 and standard deviation $\sigma^0 = 12$, we can use this for an uncertainty in x^f. Let's call the position given by the GPS tracker x^o (o for observation). One way we might combine the two pieces of information might be as a linear combination,

$$x^a = \alpha x^f + \beta x^0 \quad with\ \ \alpha + \beta = 1$$

Discuss how we might choose the α, β weights using what we know about their uncertainties.

3) Now let's find the optimal α, β. Consider the cost functional,

$$J(x) = \frac{1}{(\sigma^f)^2}\left(x - x^f\right)^2 + \frac{1}{(\sigma^0)^2}(x - x^0)^2.$$

We might consider our optimal guess distance x^a to be the one which minimizes the weighted sum of squares J. Here we have choses the weights as the reciprocal of the variances in the in model and observation. This way, the smaller the variance the more the term contributes to the cost or the larger the error the less it contributes.

Find x^a such that it minimizes $J(x)$, then rewrite your answer in the form $x^a = \alpha x^f + \beta x^0$ as in (2)

4) Finally, since $\alpha = 1 - \beta$ using your result from 3 and rewrite x^a as

$$x^a = x^f + K\left(x^0 - x^f\right).$$

K is known as the Kalman Gain! Discuss what happens to the value of K when $\sigma^0 \gg \sigma^f$ and when $\sigma^f \gg \sigma^0$, what does that say about our how much our analysis relies on the model or the data?

5) Finally, let's see if we have time for another coffee. Suppose we check the real time GPS at the 50 min mark and get a value of $x^0 = 75mi$, which would suggest that the train is going only 90mph and has 25 more miles to go, so we 16 min leaving us with 4 min for another cup of coffee, leaving a 2 min buffer for the 10 min walk. However, what does x^a suggest?