Nathan Sanford
Earthquake Inference Project

This short project infers the hypocenter of an earthquake from seismic data using least-squares multilateration. The seismic data consists of arrival times (relative to a common time) of an earthquake's seismic P-waves[1] at a portion of the stations that make up Earthscope's USArray.

Let the velocity of the P-waves associated with the earthquake travel with a velocity of αkm/s. We can then use the arrival times at each seismic observation station and the locations of each station to set up a system of equations which the hypocenter and origination time of the earthquake obey. Let $t_i, x_i, y_i,$ and z_i be the arrival time and location of the i$^{\text{th}}$ station and $t_e, x_e, y_e,$ and z_e be the origination time and hypocenter location of the earthquake. Then we have

$$t_i - t_e = \frac{1}{\alpha} D_i(x_e, y_e, z_e) \tag{1}$$

where $D_i(x_e, y_e, z_e) = \sqrt{(x_i - x_e)^2 + (y_i - y_e)^2 + (z_i - z_e)^2}$. This equation simply comes from distance equals rate multiplied by time.

We can see that retrieving the hypocenter from the arrival times is a nonlinear inverse problem as (1) is nonlinear in $x_e, y_e,$ and z_e. We then make a number of simplifying assumptions in order to make this problem more tractable. First, we will assume that the depth of the hypocenter is $z_e = 10$ km below sea level and that the P-wave velocity is a constant $\alpha = 5.7$km/s[2]. Then we will linearize the problem by making an initial guess for the origin time and epicenter coordinates of the earthquake t_e^o, x_e^o, y_e^o and expand around this guess so that $t_e = t_e^o + dt, x_e = x_e^o + dx$, and $y_e = y_e^o + dy$. Then, the distance between the earthquake and the i$^{\text{th}}$ station becomes approximately

$$D_i(x_e, y_e) \approx D_i(x_e^o, y_e^o) + \frac{1}{D_i(x_e^o, y_e^o)} [(x_e^o - x_i)dx + (y_e^o - y_i)dy] \tag{2}$$

where $D_i(x_e^o, y_e^o) = \sqrt{(x_i - x_e^o)^2 + (y_i - y_e^o)^2 + (z_i + 10)^2}$. Then, the full nonlinear problem (1) can be approximated by

$$t_i - t_e^o = dt + \frac{1}{\alpha}\{D_i(x_e^o, y_e^o) + \frac{1}{D_i(x_e^o, y_e^o)}[(x_e^o - x_i)dx + (y_e^o - y_i)dy]\}. \tag{3}$$

This simplified linear problem with three unknowns (dx, dy, dt) is the problem we will solve.

[1] More on P-waves can be found here.

[2] These simplifying assumptions were made solely so that the problem could be done quickly in the context of the class.

As this problem is linear, it can be written as a matrix vector equation. To do this, we isolate the data that is known *a priori* (upon a choice for the initial guess) and define

$$d_i = t_i - t_e^o - \frac{1}{\alpha} D_i(x_e^o, y_e^o). \tag{4}$$

Now, substituting (4) into (3) we get a problem for the unknown model vector $\underline{m} = (dt, dx, dy)^T$ that looks like

$$\underline{d} = \underline{\underline{G}}\,\underline{m} \tag{5}$$

where the i$^{\text{th}}$ row of the design matrix $\underline{\underline{G}}$ is given by

$$\Big(1, \frac{x_e^o - x_i}{\alpha D_i(x_e^o, y_e^o)}, \frac{y_e^o - y_i}{\alpha D_i(x_e^o, y_e^o)}\Big).$$

The locations of the observation stations were naturally given in terms of their latitudes and longitudes. However, this is not optimal since the distances in (5) are assumed to be in kilometers. Therefore, we roughly convert between degrees latitude and longitude using approximate conversion factors of 1° latitude being 111.19 km and 1° longitude being 85.18 km. These conversion factors are only appropriate for the general region encompassing the southeastern United States, so the latitudes and longitudes were first centered on a reference point near the Nevada-Utah border. After making a random initial guess for the epicenter, the initial guess for the origin time was computed by calculating the distance to the nearest station and using its P-wave arrival time. The generalized inverse

$$\underline{\underline{G}}^{-g} = (\underline{\underline{G}}^T \underline{\underline{G}})^{-1} \underline{\underline{G}}^T$$

was used to compute the least-squares solution to (5) which is

$$\underline{m} = \underline{\underline{G}}^{-g}\,\underline{d}. \tag{6}$$

A solution model vector $\underline{m} = (dt, dx, dy)$ from (6) then gives an updated guess for the epicenter $(x_e^o + dx, y_e^o + dy)$ and origin time $t_e^o + dt$ of the earthquake. This then provides the basis for an iterative method where the hypocenter is updated by least-squares-computed corrections.