

A Closed-Form Solution for Earthquake Location in a Homogeneous Half-Space Based on the Bancroft GPS Location Algorithm

by Demián Gómez, Charles Langston, and Robert Smalley Jr.

Abstract The traditional approach to both earthquake and Global Positioning System (GPS) location problems in a homogeneous half-space produces a nonlinear relationship between a set of known positions, seismic stations or GPS satellites, and an unknown point, an earthquake hypocenter or GPS receiver. Linearization, followed by an iterative inversion, is typically used to solve both problems. Although sources and receivers are swapped in the earthquake and GPS location problems, the observation equation is the same for both, due to the principle of reciprocity. Consequently, the mechanical part of the solution of the equations is the same and single-step closed-form solutions for the GPS location problem, such as the Bancroft algorithm, can also be used to solve for earthquake hypocenters in a homogeneous half-space. This article applies the Bancroft algorithm to synthetic and real data for the Charlevoix seismic zone and compares the location of ~ 1200 events estimated with both the Bancroft algorithm and HYPOINVERSE. The Bancroft algorithm shows quantifiable improvements in accuracy compared with traditional methods. We also show how tools commonly used by the GPS community, such as the geometric dilution of precision, can be used to better estimate the precision of the results obtained by a seismic network.

Introduction

The standard introduction to earthquake and GPS location sets the problem up in a homogeneous half-space, in which ray paths are straight lines. Although the earthquake location problem usually requires a more complex velocity model, and bending ray paths, this does not affect the basic development. In addition, there are cases in which the homogeneous half-space velocity model is the best practical model to locate earthquakes. In the case of GPS, unlike seismology, ray-path bending is a second-order effect and a constant velocity model is appropriate. Bancroft (1985) found a closed-form rather than an iterative solution for the GPS location problem with a uniform velocity model. We will show that both the GPS and earthquake location problems in a uniform velocity half-space are analogous, and the Bancroft algorithm can be used to locate hypocenters. In theory, this method has a few shortcomings, such as larger sensitivity to data outliers, because it minimizes the sum of the fourth power of the errors rather than the square (as in least squares). However, we will use real data to show the method is capable of finding solutions with smaller root mean square (rms) errors than such traditional earthquake location methods as HYPOINVERSE (Klein, 1978). We will begin by examining the development of the earthquake location problem.

Given a uniform half-space and P arrival times from an earthquake at a number of seismic stations (N) with known locations (X_n, Y_n, Z_n) , the hypocenter and origin time

(X, Y, Z, t) of the earthquake can be found from the following analysis. The travel time T_n to the n th seismic station is

$$T_n = \frac{d_n}{v} = \frac{\sqrt{(X - X_n)^2 + (Y - Y_n)^2 + (Z - Z_n)^2}}{v}, \quad (1)$$

in which d_n is the distance between the n th station and the earthquake, and v is the P or S velocity for the half-space. The left side of equation (1) is travel time, but the seismic observables are arrival times. As the origin time of an event is unknown, we need some way to estimate it. Subtracting the earliest arrival time from each observation yields $N - 1$ nonzero arrival-time differences that are travel times offset by a constant value equal to the travel time from the hypocenter to the closest seismic station. We can, therefore, rewrite equation (1) in terms of relative arrival times by subtracting the travel time t to the nearest station:

$$\tilde{T}_n = \frac{\sqrt{(X - X_n)^2 + (Y - Y_n)^2 + (Z - Z_n)^2}}{v} - t. \quad (2)$$

With P arrival times from four or more stations, the hypocenter and origin time can be found by simultaneously solving the set of equations. Although this analysis also works with S waves, earthquakes are not generally located using only S waves. If both P and S waves are available, d_n can

be estimated directly from the T_S-T_P arrival times, and data from only three stations are needed (two arrival times per station, with one for P and another for S ; i.e., six observations).

Equation (2) is nonlinear and cannot be solved analytically as it is written. Such problems are typically solved using a truncated Taylor series to linearize the problem and least squares to find the solution that minimizes the observation errors. A complete discussion on the history of least squares and some applications to geodesy can be found in [Nievergelt \(2000\)](#). Following [Geiger \(1910, 1912\)](#) for an initial estimate for the earthquake location of (X_0, Y_0, Z_0, t_0) , we can write the travel time to the n th station as

$$\tilde{T}_n \approx \frac{\partial T_n}{\partial X} \Big|_{X_0} \delta X_0 + \frac{\partial T_n}{\partial Y} \Big|_{Y_0} \delta Y_0 + \frac{\partial T_n}{\partial Z} \Big|_{Z_0} \delta Z_0 + \frac{\partial T_n}{\partial t} \Big|_{t_0} \delta t_0. \quad (3)$$

A common value for the initial estimate is the location of the station having the first arrival with fixed depth and time offsets. Equation (3) gives the exact solution of the four unknowns using the following four equations:

$$\begin{aligned} X &= X_0 + \delta X_0, & Y &= Y_0 + \delta Y_0, \\ Z &= Z_0 + \delta Z_0, & t &= t_0 + \delta t_0. \end{aligned} \quad (4)$$

Each seismic station provides one equation (observation) of the type shown in equation (3). If four observations are available, the unknown parameters may be found using standard matrix inversion methods. In the case of an overdetermined system (more than four measurements), least squares can be used to minimize the errors for the set of equations for the corrections in equation (3) to generate a new trial solution (X_0, Y_0, Z_0, t_0) to be used in the next estimate of this iterative procedure, until convergence (or failure) is obtained. Because the relation between the time corrections and the coordinates of the hypocenter is nonlinear, dropping the higher degrees of the Taylor series in equation (3) requires assuming that the correction terms in equation (4) are small enough to correctly approximate equation (2). If the correction terms are large, a single iteration might not be sufficient for convergence, and additional iterations may be necessary. This introduces a limitation to the method, because the initial estimate (X_0, Y_0, Z_0, t_0) has to be close enough to the real solution to keep the error terms small. If this condition is not satisfied, the iteration may either converge very slowly or not converge at all ([Lee and Stewart, 1981](#); [Thurber, 1985](#)).

We now examine the Global Positioning System (GPS) absolute positioning case and its solution and show it is analogous to the earthquake location problem. This comparison will validate the use of a series of tools developed in GPS analysis for seismology.

GPS Model for Absolute Positioning

GPS consists of a constellation of 24–32 satellites, orbiting the Earth at an altitude of 20,200 km, that provide navigation and time transfer signals. In the navigation appli-

cation, GPS equipment provides the location of its antenna with an error that varies from meters to a few millimeters, depending on the positioning technique used.

The most common use of GPS is to obtain the absolute position of a single receiver on or near the surface of the Earth. This is done by multiplying the travel time of the GPS radio signal from a satellite whose position is known, by the speed of light to produce a pseudorange (Fig. 1) that is used to calculate locations with precisions of 5 m or better. The receiver estimates the travel times using the arrival times of coded time-synchronized signals transmitted by the GPS satellites. In earthquake location terms, this is similar to using $S-P$ times and V_P and V_S to determine the distance from a station to the source (Fig. 1). When applied to earthquake location, the pseudoranges ($S-P$ distances) are used to solve for the XYZ coordinates of a hypocenter.

Because of the large value of the speed of light, very accurate time measurements are required to correctly determine the pseudoranges. One solution to achieve sufficient precision is to use atomic clocks. This is the solution used in the GPS satellites; however, because the size and cost of such devices are prohibitive for most GPS receivers, consumer-quality quartz clocks are typically used. Such clocks are not precise enough to perform pseudorange measurements (as a reference, a 1 ms timing error for a signal traveling at the speed of light becomes a 300 km pseudorange error), but this problem can be overcome by using a fourth pseudorange measurement to estimate a correction for the receiver clock. This solution allows the receiver to determine time almost as accurately as the satellite clocks.

By taking into account the clock correction term, the equation for the pseudorange can be written as

$$\rho_i = t_i c = \sqrt{(X_r - X_i)^2 + (Y_r - Y_i)^2 + (Z_r - Z_i)^2} - \delta t c \quad (5)$$

([Leick, 2004](#)), in which c is the appropriate speed of light, t_i is the travel time determined by the receiver, (X_i, Y_i, Z_i) are the coordinates of the i th satellite, (X_r, Y_r, Z_r) are the unknown coordinates of the GPS receiver, and δt is the receiver clock error. With four or more of these equations, we can solve for the location of the receiver and the receiver clock error.

Because the signals transmitted by the satellites are precisely synchronized and their departure times can be calculated by the receiver, the δt term in equation (5) represents the correction to apply to the receiver's clock to synchronize it to the GPS satellite time frame. A full discussion on GPS pseudorange measurements can be found in [Leick \(2004\)](#). To compare this error term with its seismological counterpart, we can think of the time of an event with respect to the synchronized time frame of a seismic station network. When an event occurs, the t term on the right side of equation (2) is estimated to place the event in the network time frame. This is equivalent to placing the receiver in the GPS satellite time frame.

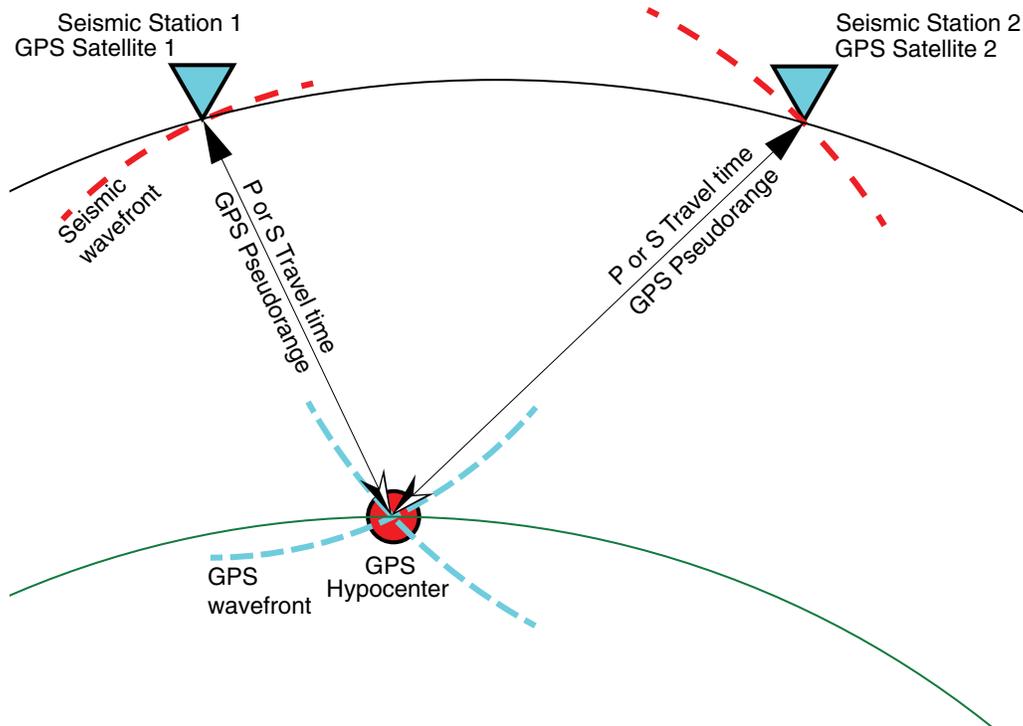


Figure 1. Relationship between the sources (earthquakes or Global Positioning System [GPS] satellites) and receivers (seismic stations or GPS receivers). The circle indicates an unknown position (GPS receiver or hypocenter), the triangles indicate known position (GPS satellites or seismic stations), the solid labeled lines show ray paths and the dashed lines show the GPS or seismic wavefronts. The color version of this figure is available only in the electronic edition.

Dividing equation (5) by the speed of light, we obtain

$$t_i = \frac{\sqrt{(X_r - X_i)^2 + (Y_r - Y_i)^2 + (Z_r - Z_i)^2}}{c} - \delta t, \quad (6)$$

in which t_i is the measured travel time. This expression is identical to equation (2) that we found for the earthquake location problem. This shows that for a uniform half-space, the earthquake location and GPS positioning problems are equivalent. This is the key observation supporting the application of the Bancroft algorithm to earthquake location (it was originally developed for GPS location). The principal difference is in the location of receivers and sources (Fig. 1). In GPS, the known locations are the sources (GPS satellites), and the unknown is the GPS receiver location. In contrast, in earthquake location, the receivers (the seismic stations) have the known locations, and the hypocenter has an unknown location. Because of the principle of reciprocity, interchanging sources and receivers between the two problems yields the same results.

In GPS, two signals (L1 and L2) with different frequencies are employed. They provide a first-order correction for a time delay produced by a dispersive change of the speed of light in the ionosphere. A similar but nondispersive change in the speed of light occurs in the troposphere, with a time delay estimated using a theoretical model. The length difference between the geometrical and the optical ray paths due to refraction in the ionosphere and troposphere is very small, thus

the bending of the ray paths can be neglected when computing GPS solutions. Only the corrections for the ionospheric and tropospheric delays are required to properly compute the pseudoranges, allowing us to consider the ray paths as straight, at least to first order. In seismology, straight rays are associated with a constant velocity, which will allow us to apply the GPS closed-form solutions to earthquake location. More complex velocity models having nonconstant velocities, in which one cannot neglect bending of the seismic ray path, are outside the scope of this article.

The Bancroft Algorithm

The application of the closed-form solution of Bancroft to earthquake location is limited to homogeneous half-spaces, a nonrealistic Earth velocity structure model in many cases. This work is not intended to present a new, general method to replace current earthquake location techniques. The homogeneous half-space model, however, is sometimes the most appropriate. The Geological Survey of Canada, for example, uses a uniform half-space velocity model to locate events in the Charlevoix seismic zone (CSZ). In the following sections, we will use data from Charlevoix to test the performance of the Bancroft algorithm against a standard earthquake location program (HYPOINVERSE, Klein, 1978).

We will now present development of a solution to equation (2) and its equivalents, equations (5) and (6), developed by Bancroft (1985), and discuss the algorithm's contribu-

tions to the earthquake location problem. Equation (5) is nonlinear due to the square root term. By simply isolating the square root on the right side and squaring both sides of the equation, we obtain

$$(\rho_i + \delta tc)^2 = (X_r - X_i)^2 + (Y_r - Y_i)^2 + (Z_r - Z_i)^2. \quad (7)$$

Expanding and rearranging this expression yields

$$\begin{aligned} -\{X_r^2 + Y_r^2 + Z_r^2 - (\delta tc)^2\} &= (X_i^2 + Y_i^2 + Z_i^2 - \rho_i^2) \\ -2(X_i X_r + Y_i Y_r + Z_i Z_r - \rho_i \delta tc). \end{aligned} \quad (8)$$

Solving equation (8) for (X_r, Y_r, Z_r) and δt is no easier than solving equation (2). To recast equation (8) into an equation with a closed-form solution, Bancroft introduces the Minkowski functional (Minkowski, 1907/1908), also known as the Lorentz inner product, from relativity theory. Although the justification for the use of this inner product was not discussed in Bancroft (1985), the application of the Minkowski functional implies the geometry of a system of pseudorange equations is hyperbolic (Chaffee and Abel, 1994). Other authors such as Pujol and Smalley (1990) have taken advantage of this geometry to solve for the coordinates of earthquake hypocenters. Regardless of the hyperbolic geometry, Sirolo (2010) showed the Bancroft algorithm can also be derived without the use of the Minkowski functional. In this article, however, we will follow Bancroft's notation and derivation.

The Minkowski functional for four space is defined as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{M} \mathbf{b}, \quad (9)$$

in which \mathbf{M} is the 4×4 matrix:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \quad (10)$$

The Minkowski functional is also the standard dot product between two relativistic four-space position vectors with

elements $\begin{bmatrix} X \\ Y \\ Z \\ \sqrt{-1}t \end{bmatrix}$. Defining the vectors

$$\bar{\mathbf{X}}_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{X}}_r = \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}, \quad (11)$$

the terms of equation (8) can be rewritten using the Minkowski functional:

$$\begin{aligned} X_i^2 + Y_i^2 + Z_i^2 - \rho_i^2 &= \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix} \right\rangle, \\ X_i^2 + Y_i^2 + Z_i^2 - (\delta tc)^2 &= \left\langle \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle, \\ X_i X_r + Y_i Y_r + Z_i Z_r - \rho_i \delta tc &= \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle, \end{aligned} \quad (12)$$

and equation (8) becomes

$$\begin{aligned} \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix} \right\rangle + \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle \\ - \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle &= 0. \end{aligned} \quad (13)$$

To further simplify equation (13), Bancroft introduces the following two terms:

$$E = \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle \quad \text{and} \quad a_i = \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix} \right\rangle. \quad (14)$$

For each observation (to a satellite or seismic station), a term a_i can be calculated. These terms can be arranged into a column vector:

$$\mathbf{a} = \begin{bmatrix} \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_1 \\ \rho_1 \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_1 \\ \rho_1 \end{bmatrix} \right\rangle \\ \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_2 \\ \rho_2 \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_2 \\ \rho_2 \end{bmatrix} \right\rangle \\ \vdots \\ \frac{1}{2} \left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix} \right\rangle \end{bmatrix}. \quad (15)$$

Because the last equation of equation set (12) has both known and unknown terms and there are i observation equations, the Minkowski functional form for the matrix equation to be solved can be rewritten in a noncompact fashion that isolates the unknown terms:

$$\left\langle \begin{bmatrix} \bar{\mathbf{X}}_i \\ \rho_i \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix} \right\rangle = \mathbf{B} \mathbf{M} \begin{bmatrix} \bar{\mathbf{X}}_r \\ \delta tc \end{bmatrix}, \quad (16)$$

in which \mathbf{B} is an $i \times 4$ matrix containing all the known terms from equation (8),

$$\mathbf{B} = \begin{bmatrix} X_1 & Y_1 & Z_1 & \rho_1 \\ X_2 & Y_2 & Z_2 & \rho_2 \\ \vdots & \vdots & \vdots & \vdots \\ X_i & Y_i & Z_i & \rho_i \end{bmatrix}. \quad (17)$$

Substituting equations (16) and (14) into equation (13) and

defining $\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, we obtain

$$\mathbf{a} - \mathbf{B}\mathbf{M} \begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix} + E\mathbf{e} = 0. \quad (18)$$

Solving equation (18) for the vector containing the unknowns $\begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix}$ yields the so-called GPS navigation equation,

$$\begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix} = (\mathbf{B}\mathbf{M})^{-1}(E\mathbf{e} + \mathbf{a}). \quad (19)$$

Because the E term defined in equation (14) also includes the unknowns $\begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix}$, equation (19) is nonlinear and cannot be solved as given. Applying the Minkowski functional again to both sides of equation (19) yields

$$\left\langle \begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix}, \begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix} \right\rangle = \langle (\mathbf{B}\mathbf{M})^{-1}(E\mathbf{e} + \mathbf{a}), (\mathbf{B}\mathbf{M})^{-1}(E\mathbf{e} + \mathbf{a}) \rangle. \quad (20)$$

This can be rewritten as

$$2E = \{(\mathbf{B}\mathbf{M})^{-1}(E\mathbf{e} + \mathbf{a})\}^T \mathbf{M}(\mathbf{B}\mathbf{M})^{-1}(E\mathbf{e} + \mathbf{a}). \quad (21)$$

Rearranging equation (21) and using $\mathbf{M}^{-1} = \mathbf{M}$ and $\langle \mathbf{M}\mathbf{a}, \mathbf{M}\mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle$, we obtain a quadratic equation in E :

$$\langle \mathbf{B}^{-1}\mathbf{e}, \mathbf{B}^{-1}\mathbf{e} \rangle E^2 + 2(\langle \mathbf{B}^{-1}\mathbf{e}, \mathbf{B}^{-1}\mathbf{a} \rangle - 1)E + \langle \mathbf{B}^{-1}\mathbf{a}, \mathbf{B}^{-1}\mathbf{a} \rangle = 0. \quad (22)$$

In this expression, all the components of the quadratic coefficients \mathbf{B} , \mathbf{e} , and \mathbf{a} are known, thus the value of E can be found. This solution will yield two possible values for E , from which the physically correct answer is chosen, that is, a positive pseudorange value. If both values for E are positive, the correct answer is the one that yields a solution that has a coordinate below the seismic network. If a solution for the problem cannot be found, both values for E might be negative or imaginary. Once E is obtained, equation (19) can be used to solve for $\begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix}$. Abel and Chafee (1991) and Chafee and Abel (1994) more fully discuss details about the uniqueness and existence of the solution.

For the case of a system with more than four observations, the matrix \mathbf{B} will have more than four rows, and thus it will not be invertible. This is addressed in the standard method by applying the Moore–Penrose pseudoinverse to \mathbf{B} :

$$\mathbf{B}^+ = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T. \quad (23)$$

Using the Moore–Penrose pseudoinverse for \mathbf{B}^{-1} in (19) one obtains the regular least-squares form with an extra term \mathbf{M}^{-1} :

$$\begin{bmatrix} \bar{\mathbf{x}}_r \\ \delta tc \end{bmatrix} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{M}^{-1}(E\mathbf{e} + \mathbf{a}). \quad (24)$$

Taking a closer look at the term $(E\mathbf{e} + \mathbf{a})$, however, the reader will notice the observations in the vector \mathbf{a} are squared due to the Minkowski functional. This squaring of observations will have a few consequences that will be discussed later.

The development of the Bancroft method was primarily motivated by the need to solve equation (5) using minimum computer and power resources. This algorithm has been widely used in GPS receivers, because it does not require any *a priori* information about the initial location, reducing the receivers' cold start up time. The Bancroft algorithm is only one of many other closed-form solutions developed for GPS positioning that could also be implemented for earthquake location. For other examples, see Sirola (2004).

Application in Seismology and Tests using Synthetic Data

We have reviewed the GPS inverse problem and the earthquake location problem and observed that they are identical when using a uniform half-space velocity model. We will now test the Bancroft algorithm and compare the results with a standard earthquake location program, HYPOINVERSE (Klein, 1978), using both synthetic and real data for the Charlevoix seismic network. We will relocate ~ 1200 events from the Charlevoix catalog that occurred during the December 1989–August 1999 time period. Finally, we will introduce a method of quantifying the quality of the solution, known as dilution of precision (DOP) that is based on the hypocenter network geometry, and show that DOP is more informative than azimuthal gap.

Using the CSZ seismicity catalog (see Data and Resources), we obtained synthetic data by calculating travel times to the Charlevoix seismic network stations in a uniform velocity half-space for 1330 events during the selected time period. Random Gaussian noise, with a standard deviation of 0.02 s, was added to the synthetic data to simulate typical picking time errors (Powell *et al.*, 2010). At this time, we are not assuming long tails caused by erroneous picks or phase association misidentifications, because the goal of the synthetic analysis is to establish the capability of each method (Bancroft and HYPOINVERSE) to obtain the original hypocenter.

Histograms of the difference between the true and estimated solutions are shown in Figure 2a. Both methods successfully located 100% of the hypocenters. For latitude and longitude, Bancroft and HYPOINVERSE solutions appear to be equally accurate, with Bancroft showing a slightly more dispersed distribution. To quantify these results, we fit normal distributions to the histograms in Figure 2a. The results of these fits are shown in Table 1 (“Gaussian Noise” [GN] columns). For depth and origin time estimations, Bancroft seems to be slightly more accurate but more dispersed than HYPOINVERSE, confirming that HYPOINVERSE solutions are slightly biased.

To better understand the behavior observed in Figure 2a, we performed a second synthetic test without adding Gaussian

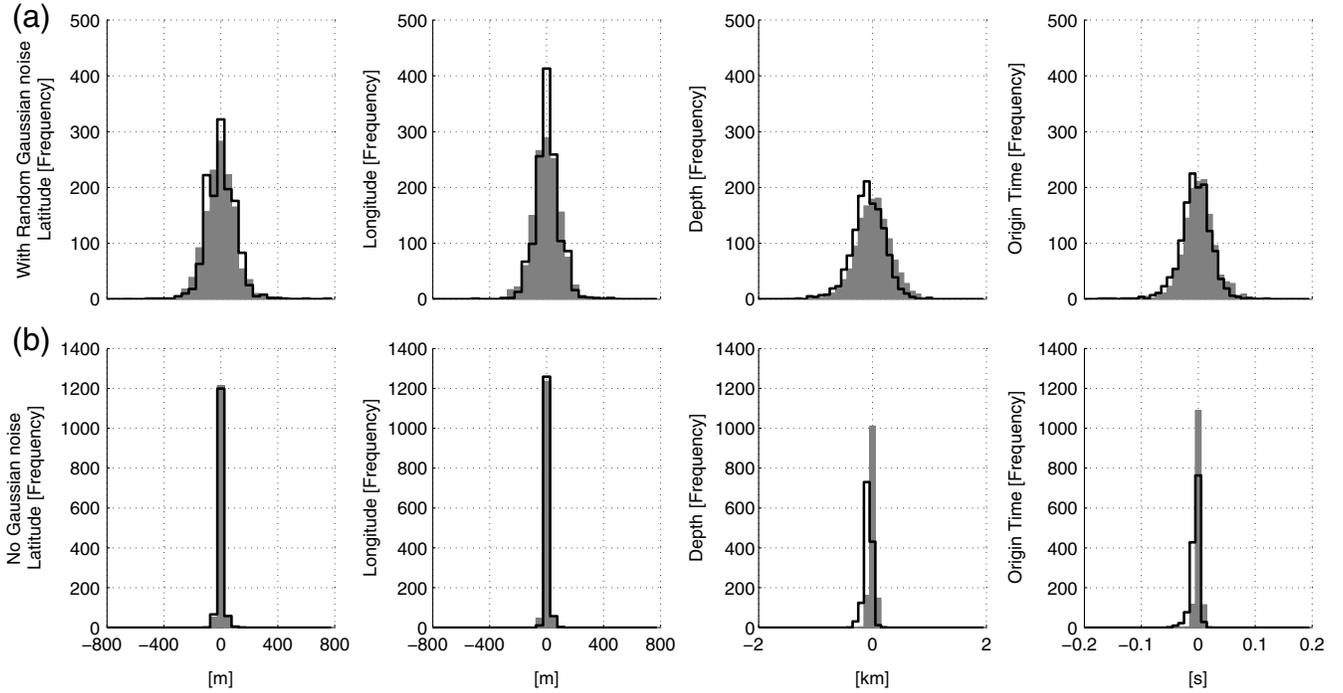


Figure 2. (a) Comparison of Bancroft algorithm and HYPOINVERSE location residuals using synthetic data. Dark gray indicates Bancroft, and the black lines indicate HYPOINVERSE. Histograms are shown for latitude, longitude, depth, and origin time residuals calculated from the true and estimated solutions using HYPOINVERSE and the Bancroft algorithms. The x axis shows residual values and y axis shows frequency. (b) Histograms of the same variables but without adding Gaussian noise to synthetics.

noise to the arrival times. Figure 2b shows that after removing the Gaussian noise, the HYPOINVERSE depth and origin time estimations show asymmetric histograms, whereas the Bancroft solution is always centered at zero. The “No Gaussian Noise” (NN) column in Table 1 shows the results of the normal distribution fits. The persistent bias observed in depth and origin time is probably a result of small differences between our forward calculation and HYPOINVERSE’s seismic stations coordinate transformations used to solve the inverse problem. We arrive at this conclusion after always observing the same bias running HYPOINVERSE with different trial depths and iteration criteria. We also conclude that these biases are too small to modify the result of the tests using the real CSZ arrival-time data, presented in the next section.

We should mention that HYPOINVERSE’s input format for arrival-time data is FORTRAN F5.2, which only allows precision to a hundredth of a second, and this round off contributes

to some of the dispersion observed in HYPOINVERSE result with noiseless data. To compare against the Bancroft algorithm, we used the same input precision. Using full double-precision arrival times (in MATLAB; see [Data and Resources](#)), the Bancroft algorithm finds a solution that differs from the true solution by $\sim 10^{-3}$ m in latitude and longitude, $\sim 10^{-3}$ m in depth, and $\sim 10^{-6}$ s in origin time in the region examined.

Relocation of Events using Real Arrival-Time Data

We proceeded to relocate the 1330 events in the CSZ using both methods, from which 1232 were relocated (92%). The remaining 98 events were not relocated, either due to quality of the data or because of missing arrival times. We only relocated events with more than four observations that were marked as “good” on the arrival-time quality column. This avoids exact solutions of unknown quality when the number of observations and parameters are equal.

Table 1
Normal Fit Distributions to Histograms

Solution		Latitude (m)		Longitude (m)		Depth (m)		Origin Time (s)	
		GN	NN	GN	NN	GN	NN	GN	NN
Bancroft	Mean	3.8	-0.35	-5.8	0.14	15	0.04	0.001	3×10^{-5}
	Standard deviation	111.0	16.00	94.0	14.00	315	47.0	0.029	0.004
HYPOINVERSE	Mean	-4.3	-0.68	8.9	5.50	-61	-76.0	0.004	-0.005
	Standard deviation	97.0	20.00	86.0	17.00	279	63.0	0.028	0.008

Distributions with Gaussian noise (GN) values and without Gaussian noise (NN) values. Bancroft fits show that its solutions are slightly more accurate but more dispersed than those from HYPOINVERSE.

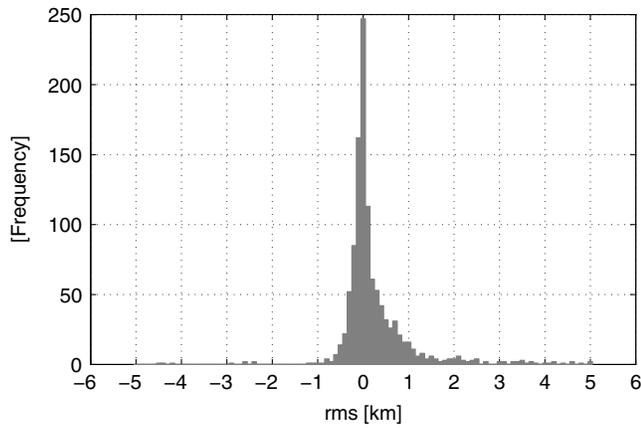


Figure 3. Histogram of the root mean square (rms) difference between the HYPOINVERSE and Bancroft solutions using Charlevoix seismic-zone real data. An $\text{rms}_D > 0$ represents a solution in which Bancroft produced a better rms than HYPOINVERSE, and $\text{rms}_D < 0$ represents a solution in which HYPOINVERSE produced a better rms than Bancroft.

The rms residual difference for each earthquake, $\text{rms}_D = \text{rms}_H - \text{rms}_B$ (in which the subscripts D, H, and B indicate difference, HYPOINVERSE, and Bancroft) was calculated to quantify the comparison of the results from both methods (Fig. 3). An $\text{rms}_D > 0$ represents a solution in which Bancroft had a smaller rms than HYPOINVERSE, and vice versa. Figure 3 shows a clear tendency toward positive values, showing that Bancroft's solution has smaller residuals than HYPOINVERSE. Out of 1232 events, 548 solutions had $\text{rms}_D < 0$, whereas 684 had $\text{rms}_D > 0$. The asymmetry of the histogram shows that in the cases in which HYPOINVERSE wins ($\text{rms}_D < 0$), the difference between the two solutions is generally small, as shown by the steep decline of the histogram. In the cases in which Bancroft wins ($\text{rms}_D > 0$), the difference between the two solutions is generally larger, as shown by the wider histogram tail.

Quantifying Solution Quality

The most common GPS solution quality estimator is a DOP formulation. There are three DOP estimators (time, position, and geometric), and they are computed using the diagonal elements of the covariance matrix associated with the design matrix. DOP represents the effect of the satellite-GPS receiver geometry on the accuracy of the solution (Leick, 2004) and is independent of measurement errors. Because the equation for earthquake location and GPS is the same, the DOP concept can provide a useful solution quality metric in the earthquake location problem. The precision with which a seismic network will be capable of locating an event is, therefore, also largely determined only by the event-station geometry.

The DOP factor may be conceptually interpreted using the ratio

$$\text{DOP} = \frac{\Delta(\text{Output location})}{\Delta(\text{Measured data})}. \quad (25)$$

Ideally, small changes in the measured data should produce small variations in the result, which will yield small DOP values. When small changes in the input result in large changes in the result, this indicates the solution is very sensitive to errors in the observations.

Using the classic least-squares method and equation (2), each seismic station will provide an observation. The solution of the event hypocenter and origin time (X, Y, Z, t) can be found from the linearized version of equation (2). Assuming an overdetermined system of equations, the matrix notation for a least-squares solution has the form

$$\bar{\mathbf{X}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{L}, \quad (26)$$

in which \mathbf{L} contains the observations (measured pseudoranges between event and seismic stations), $\bar{\mathbf{X}}$ contains the unknown parameters in vector form, and \mathbf{A} is known as the design matrix:

$$\mathbf{A} = \begin{bmatrix} \frac{(X-X_1)}{R_1} & \frac{(Y-Y_1)}{R_1} & \frac{(Z-Z_1)}{R_1} & -v \\ \frac{(X-X_2)}{R_2} & \frac{(Y-Y_2)}{R_2} & \frac{(Z-Z_2)}{R_2} & -v \\ \vdots & \vdots & \vdots & \vdots \\ \frac{(X-X_n)}{R_n} & \frac{(Y-Y_n)}{R_n} & \frac{(Z-Z_n)}{R_n} & -v \end{bmatrix} \quad \text{and}$$

$$R_n = \sqrt{(X - X_n)^2 + (Y - Y_n)^2 + (Z - Z_n)^2}. \quad (27)$$

The term $(\mathbf{A}^T \mathbf{A})^{-1}$ is usually called the inverse of the normal matrix, or simply the covariance matrix \mathbf{Q} . From this matrix, several DOPs can be calculated as

$$\begin{aligned} \text{GDOP} &= \sqrt{Q_{11} + Q_{22} + Q_{33} + Q_{44}}, \\ \text{TDOP} &= \sqrt{Q_{44}}, \\ \text{VDOP} &= \sqrt{Q_{33}}, \\ \text{HDOP} &= \sqrt{Q_{11} + Q_{22}}, \quad \text{and} \\ \text{PDOP} &= \sqrt{Q_{11} + Q_{22} + Q_{33}}. \end{aligned} \quad (28)$$

The geometric dilution of precision (GDOP) estimates the total effect on the solution of the combined network-event geometry and origin time determination. The time dilution of precision (TDOP), vertical dilution of precision (VDOP), horizontal dilution of precision (HDOP), and position dilution of precision (PDOP) reflect the precision of the network with respect to determination of the origin time, depth, epicenter, and hypocenter of an event, respectively.

Because the covariance matrix does not depend on the observed data, the net effect of the geometry of the network on the solution can be estimated by calculating the covariance matrix as a function of hypocenter location. PDOP can be compared with the traditional method of using the azimuthal gap. The gap method estimates the quality of the epicenter by examining the gaps between azimuthally adjacent stations. PDOP and GDOP are more powerful, as PDOP provides quality estimation for the hypocenter and GDOP provides quality

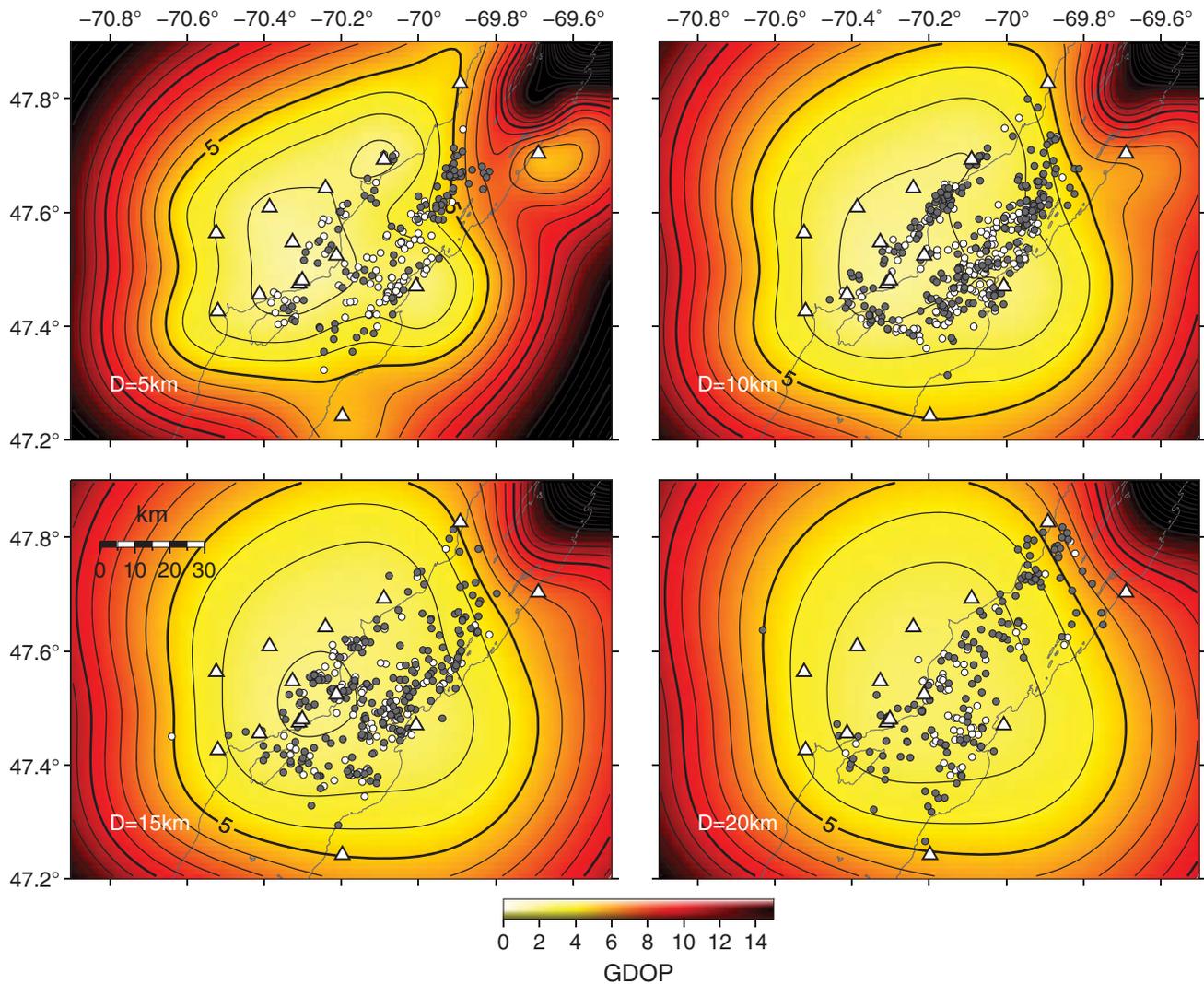


Figure 4. Example geometric dilution of precision (GDOP) for the Charlevoix Seismic Network at 5, 10, 15, and 20 km depths. Seismic stations are shown (triangles). For catalog epicenters, dark circles indicate hypocenters with $\text{rms}_D > 0$, and light circles are those with $\text{rms}_D < 0$. Depth ranges on each depth slice are 0–7.5 km, 7.5–12.5 km, 12.5–17.5 km, 17.5 km, and deeper. The gray thin lines show Saint Lawrence River shore. The color version of this figure is available only in the electronic edition.

estimation for the hypocenter and origin time, based on the earthquake-station geometry. This procedure can be applied to the covariance matrix associated with the least-squares design matrix of standard earthquake location programs, such as the various versions of the HYPO family, double difference, and 3D inversion.

Figure 4 shows a map of the Charlevoix seismic network with the resulting GDOP surface as a function of position for source depths of 5, 10, 15, and 20 km, and the Bancroft relocated catalog events. In GPS, a GDOP lower than five is considered acceptable or good, whereas anything greater is considered poor quality. This is an arbitrary limit that has been proven to provide a useful threshold for GPS solution quality estimation (Leick, 2004). We will later show that this limit can also be used in seismology to describe the boundary between good and poor quality locations. As expected, events occurring within clusters of seismic stations will be

well located. Note that isolated stations near the edge of the network locally reduce GDOP, showing the reasonable result that a nearby station improves the quality of the location even though the azimuthal gap may be quite large.

It should also be noted that one can weight the inverse of the normal matrix $(\mathbf{A}^T \mathbf{A})^{-1}$ in the usual way by using $(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}$, in which \mathbf{P} is a matrix containing the weights. In this way, we can take into account *a priori* information such as a station being located in a region where the data is noisier (e.g., station located in unconsolidated sediments, cultural noise, etc.) or any other information that can affect the confidence in the data.

We will now analyze the GDOP of the CSZ relocations by classifying them by rms_D . Figure 5 shows the histograms of the GDOP for the events that had $\text{rms}_D > 0$ and $\text{rms}_D < 0$. As expected, a large number of events occurred in the region of $\text{GDOP} < \sim 4$, which is a reasonable value within the seismic

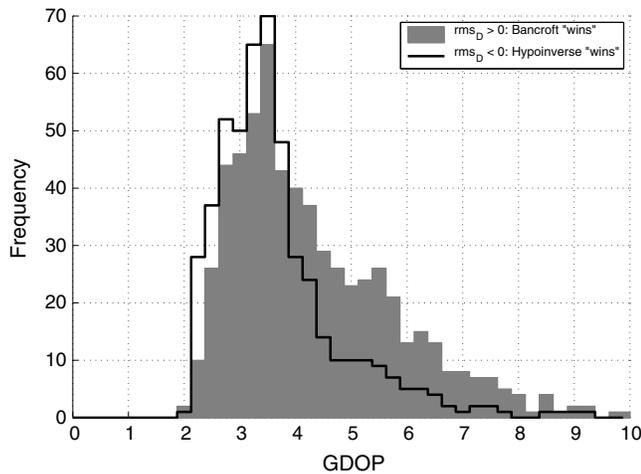


Figure 5. Histograms of the GDOP classified by rms value. For hypocenters with $\text{GDOP} < \sim 4$, Bancroft had fewer cases with smaller rms than did HYPOINVERSE. For hypocenters with $\text{GDOP} > \sim 4$, Bancroft had more cases with smaller rms than did HYPOINVERSE.

network limits, as seen in Figure 4. It should be noted, however, the GDOP map in Figure 4 was constructed assuming that all the seismic stations provide arrival-time data, which for some events is not true, resulting in a higher GDOP value for that particular hypocenter. In the region where $\text{GDOP} < \sim 4$, there are fewer events where Bancroft had a smaller rms than HYPOINVERSE. For the locations where $\text{GDOP} > \sim 4$, however, we observe the opposite, that is, there are more cases in which Bancroft had a smaller rms than HYPOINVERSE. This result confirms the findings of Bancroft (1985), that the algorithm improves the accuracy in situations with large GDOP and validates the use of $\text{GDOP} = 5$ as a threshold between good and poor quality locations. In the particular case examined here, the threshold could be tightened to $\text{GDOP} = 4$.

Discussion

In this article, we presented a comparison between the GPS and earthquake location inverse problems for a uniform half-space velocity model and found that they are analogous and expressed by the same equation. We then showed how to apply a noniterative algorithm developed to determine position using GPS data to the earthquake location problem in seismology. Advantages of the closed-form algorithm include no need for an initial estimate, improved efficiency as there is no need to iterate, and a numerically stable solution that improves accuracy in situations with large GDOP (Bancroft, 1985), as shown by Figure 5.

Despite this improvement in precision, some questions have arisen regarding the statistics and norm optimization criteria of the results obtained by the Bancroft algorithm when used in overdetermined cases (Chaffee and Abel, 1994). A complete discussion of noniterative algorithms for solving the GPS location problems can be found in Sirola (2010). Although some drawbacks about these methods are discussed by Sirola

(2010), such as numerical instability in certain cases of negative pseudorange measurements), these drawbacks do not present problems when applying the Bancroft algorithm for earthquake location.

There are two points from Sirola (2010), however, that deserve special attention. In overdetermined systems, the Bancroft algorithm minimizes the sum of the fourth power of the errors, not the sum of the square of the errors (as in least squares). This is easily observed in equation (24) in which one finds the observations are squared. As a result, the solution is more influenced by outliers in the data than in least squares. Sirola calls these types of solutions “least quartic”. However, in cases in which errors are of similar magnitude, this effect should not be critical for data without large outliers. Cases of events with stations providing large outlier data can be easily filtered by running the algorithm twice. A first run is used to calculate a first hypocenter, from which outliers can be detected from the residuals; and a second run is used to obtain the final solution. A similar argument regarding the influence of the norm optimization criteria exists between the use of the L1 norm and the L2 norm (least squares) (Aster *et al.*, 2005). As demonstrated by the test using real data, 55% of the solutions showed a reduction in rms residuals. In the cases in which HYPOINVERSE wins ($\text{rms}_D < 0$), the difference between the two solutions is generally small. In the cases in which Bancroft wins ($\text{rms}_D > 0$), the difference between the two solutions is generally larger.

Furthermore, the consequences of the application of this least-quartic solution can be observed in both synthetic and real data tests. During the synthetic tests (using random Gaussian noise), the normal distribution fits performed on the histograms revealed a larger dispersion on the Bancroft algorithm results. Because the Bancroft algorithm finds the exact solution but using a different norm optimization, the solutions have a tendency to be more dispersed than those using least squares. For the real data test, Bancroft had fewer results with lower rms than HYPOINVERSE in areas with low GDOP, most likely because HYPOINVERSE did not have convergence problems in these regions, making the norm optimization difference more evident (although the rms difference rarely exceeded ~ 1 km). In regions with large GDOP (in which standard-linearized location algorithms have higher chances of failing), the Bancroft algorithm provided more solutions with lower rms than did HYPOINVERSE because, although the norm optimization used in Bancroft provides a larger rms when compared with least squares, the linearized algorithm was unable to converge to a solution with lower rms. We should point out that there are many parameters to stabilize HYPOINVERSE that were not investigated and could potentially provide a lower rms. However, the application of Bancroft without any *a priori* information of the problem provided a better solution than that of HYPOINVERSE for 55% of the events.

The second point raised by Sirola (2010) is also related to the fact that the observations in equation (24) are squared as a consequence of applying the Minkowski functional. Dealing with the covariance of these squared measurements is more complicated. Moreover, applying traditional L1 or L2 norm

statistics to least-quartic methods will not provide useful information about the solution quality. For seismologic applications, this point should not represent a problem, because it is fairly common to work with quantities for which neither the error distribution nor the confidence level are known. Furthermore, quality estimators such as GDOP, taken together with the residuals, can provide information about the solution quality. Despite these drawbacks, the results of our tests show that Bancroft's synthetic solutions are comparable with those of HYPOINVERSE, with a higher dispersion due to the norm optimization difference. We have also shown that Bancroft improved 55% of the real locations. Additionally, we showed how to implement a set of simple and straightforward solution quality estimates (TDOP, HDOP, VDOP, PDOP, and GDOP) based on the source and seismic network geometry.

A homogeneous half-space is oftentimes too simple a model for determining earthquake locations. In certain cases, however, such as the CSZ, a uniform half-space is the most reasonable model. The Bancroft algorithm relies on a constant velocity medium (ray paths are straight lines), to make use of the Minkowski functional. At a minimum, the Bancroft algorithm can be used to obtain good initial earthquake locations (rather than using the location of the station with the first arrival) for use with more realistic nonhomogeneous velocity models and corresponding iterative methods. The approach of using a closed-form solution as a starting position for a conventional iterative least-squares method was also suggested by Sirola (2004). Although we have not tested this option, HYPOINVERSE requires setting a trial depth and origin time before running the location routine. If using Bancroft as a first rough approximation of the location, there would be no need to set these trial parameter values. The Bancroft method should also be directly applicable to relative location methods based on using a master event with a known location. In these methods, events near the master event are located with respect to the master event by using relative arrival times, assuming a constant velocity in the vicinity of the reference event. These are topics for future work.

Data and Resources

We would like to thank Maurice Lamontagne from the Geological Survey of Canada for providing the Charlevoix seismic-zone earthquake catalog data and metadata. Figure 4 was made using the Generic Mapping Tools version 5.1.1 (Wessel and Smith, 1998). MATLAB is available from www.mathworks.com/products/matlab.

Acknowledgments

D. Gómez is supported by the National Science Foundation (NSF) Office of Polar Programs grant, Collaborative Research: IPY: POLENET-Antarctica: Investigating Links between Geodynamics and Ice Sheets (NSF-ANT-062339 and supplement -0948103), and by the Center for Earthquake

Research and Information, University of Memphis. We would like to thank Bill Barnhart, an anonymous reviewer, and Associate Editor Roland Bürgmann for critical reviews that improved the manuscript, and we thank Christine Powell for her contributions to this work.

References

- Abel, J. S., and J. W. Chaffee (1991). Existence and uniqueness of GPS solutions, *IEEE Trans. Aero. Electron. Syst.* **27**, no. 6, 952–956, doi: [10.1109/7.104271](https://doi.org/10.1109/7.104271).
- Aster, R., B. Borchers, and C. Thurber (2005). *Parameter Estimation and Inverse Problems*, Academic Press/Elsevier, Burlington, Vermont.
- Bancroft, S. (1985). An algebraic solution of the GPS equations, *IEEE Trans. Aero. Electron. Syst.* **AES-21**, 56–59.
- Chaffee, J., and J. Abel (1994). On the exact solutions of pseudorange equations, *IEEE Trans. Aero. Electron. Syst.* **30**, no. 4, 1021–1030, doi: [10.1109/7.328767](https://doi.org/10.1109/7.328767).
- Geiger, L. (1910). Herbstimmung bei Erdbeben aus den Ankunftszeiten, *K. Gessell. Wiss. Goett.* **4**, 331–349 (in German).
- Geiger, L. (1912). Probability method for the determination of earthquake epicenters from the arrival time only, *Bull. St. Louis Univ.* **8**, 60–71.
- Klein, F. W. (1978). Hypocenter location program HYPOINVERSE, *U.S. Geol. Surv. Open-File Rept.* 78-694, 113 pp.
- Lee, W. H. K., and S. W. Stewart (1981). Principles and applications of microearthquake networks, *Adv. Geophys. Suppl.* 2, 293 pp.
- Leick, A. (2004). *GPS Satellite Surveying*, Third Ed., John Wiley & Sons, New Jersey.
- Minkowski, H. (1907/1908). Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 53–111 (in German).
- Nievergelt, Y. (2000). A tutorial history of least squares with applications to astronomy and geodesy, *J. Comput. Appl. Math.* **121**, no. 1/2, 37–72, ISSN 0377-0427, doi: [10.1016/S0377-0427\(00\)00343-5](https://doi.org/10.1016/S0377-0427(00)00343-5).
- Powell, C. A., M. M. Withers, H. R. DeShon, and M. M. Dunn (2010). Intrusions and anomalous V_p/V_s ratios associated with the New Madrid seismic zone, *J. Geophys. Res.* **115**, no. B08311, doi: [10.1029/2009JB007107](https://doi.org/10.1029/2009JB007107).
- Pujol, J., and R. Smalley Jr. (1990). A preliminary earthquake location method based on a hyperbolic approximation to travel times, *Bull. Seismol. Soc. Am.* **80**, 1629–1642.
- Sirola, N. (2004). A versatile algorithm for local positioning in closed form, *Proceedings of the 8th European Conference GNSS 2004*, Rotterdam, The Netherlands, 16–19 May 2004.
- Sirola, N. (2010). Closed-form algorithms in mobile positioning: Myths and misconceptions, *7th Workshop Positioning Navigation and Communication (WPNC)*, Dresden, Germany, 11–12 March 2010, 38–44, doi: [10.1109/WPNC.2010.5653789](https://doi.org/10.1109/WPNC.2010.5653789).
- Thurber, C. (1985). Nonlinear earthquake location: Theory and examples, *Bull. Seismol. Soc. Am.* **75**, 779–790.
- Wessel, P., and W. H. F. Smith (1998). New, improved version of generic mapping tools released, *Eos Trans. AGU* **79**, no. 47, 579–579, doi: [10.1029/98EO00426](https://doi.org/10.1029/98EO00426).

Center for Earthquake Research and Information
University of Memphis
3876 Central Avenue, Suite 1
Memphis, Tennessee 38152-3050

Manuscript received 16 March 2014;
Published Online 27 January 2015