Celestial navigation has been guiding the mariner for hundreds of years; more recently, the GPS (GNSS, in general) has made 3-D positioning and precise time a commodity for everyday use. With an increased awareness of the vulnerabilities associated with any satellite-based system, there is a renewed interest in non-satellite PNT systems. We have been reexamining celestial navigation from the vantage point of those conversant in GNSS language. Two aspects of these parallel approaches have attracted our attention: how to choose which stars to process (similar to the GNSS satellite selection problem) and if (and at what accuracy) precise time can be estimated from celestial measurements.

This paper clarifies two common misconceptions about celestial navigation: (1) that the stars selected for celestial navigation do not need to be “evenly” distributed across 360 degrees of azimuth, that multiple stars from nearly similar directions (i.e. Polaris and stars in the Big Dipper) can be very useful together in forming a good celestial fix; and (2) that while “star clocks” are advertised, their result is not precise time, and that celestial techniques can only provide a low quality source of time information.

Index Terms—celestial navigation, DOP

I. INTRODUCTION

Celestial navigation, the art of determining one’s location on the surface of the earth based upon measurements of a collection of stars, has been guiding the mariner for hundreds of years. Since 1802, Bowditch’s classic text “The New American Practical Navigator” (and its multiple revisions) has provided details of the methods of celestial navigation including the use of the sextant and examples of the sight reduction process [1]. Since the 1950’s electronic navigation systems have overtaken celestial navigation, providing accurate measurements of position to the user. Early terrestrial systems, such as Omega and Loran, provided 2-D positioning to maritime users; more recently, the GPS (and GNSS, in general) has made 3-D positioning and time a commodity for everyday use. Given increased awareness of the vulnerabilities associated with any satellite-based system, there is currently a renewed interest in non-satellite PNT (position, navigation, and time) systems.

In the world of GNSS, measurements are taken for each satellite and determining the user’s position (and time) involves solving the set of pseudorange equations; when overdetermined, this is in a least squares sense. While these equations are nonlinear, they can be solved via an iterative, linearized approach. The resulting 3-D position accuracy is often described by the geometric dilution of precision, or GDOP [2]. This measure, a function of GNSS constellation geometry (specifically the azimuths and elevation angles to the satellites employed in the solution), is a condensed version of the covariance matrix of the errors in the position and time estimates. Combining the GDOP value and an estimate of the user range error allows one to establish the 95% confidence ellipsoid.

For 2-D positioning on the surface of the earth, celestial navigation is based upon measuring the altitudes (elevation angles, relative to an artificial or visible horizon, historically measured with a sextant) of celestial bodies to locate the user [1], [3]–[5]. Assuming multiple measurements of star elevations, finding ones’ position similarly involves solving an overdetermined set of nonlinear equations; again, an iterative, linearized approach is known [6], [7].

In recent work these authors began to reexamine celestial navigation from the vantage point of engineers conversant in GNSS language. Specifically, having developed lower bounds to GDOP for GNSS performance assessment in [8], we similarly considered the limits of performance of celestial navigation by developing in [9] a lower bound to HDOP (horizontal dilution of precision), the natural metric to describe the impact of the geometry of the celestial bodies employed in the position solution. As the development of this bound was constructive (meaning that we developed “balance” conditions on the azimuth of the stars employed in the position solution), we were also able to characterize what a good positioning constellation might look like.

In GNSS a problem of interest is how to select a subset of the available satellites for best performance (e.g. for a 4 satellite fix it is common to choose those satellites that form the tetrahedron with greatest volume [10]); a sizable amount of work has been done on this problem (see, for example, [11]–[13]). Paralleling concepts from GNSS and celestial navigation leads to the similar question of how to select the stars for the best celestial fix. Over 200 years ago Bowditch presented a basic solution for the star selection problem (which is still taught today!): “Choose the stars and planets that give the best bearing spread …” [1, p.276]. This guidance is commonly interpreted as choosing stars that provide as uniform an azimuth distribution as possible (and never, never select stars with nearly equal azimuths).

In our 2019 paper [9] we proved that this regular practice is a sufficient, but not necessary, condition for best HDOP and is, in fact, overly restrictive. Specifically, we showed that while azimuths forming the vertices of a regular polygon (i.e. a “uniform distribution”) meets the balance conditions, many other selections of azimuths can achieve the balance conditions.
A. Review of the Basics

The classical method used to solve for position at sea is called the “altitude-intercept” method, also known as the Marcq St. Hilaire method (after the French navigator who developed the technique). In this method one uses the known celestial bodies’ locations plus an assumed position (AP), presumably reasonably close to the actual position, to calculate the altitudes one would expect at known times at that AP. This can be accomplished using the Nautical Almanac and Publication 229 or https://aa.usno.navy.mil/data/docs/celeavtable.php to determine both the computed altitudes and azimuth angles. The difference between the computed and observed altitudes (measured in minutes in arc), combined with the azimuth, is used to plot each celestial line of position. More specifically, the line of position produced is oriented perpendicular to the azimuth toward each celestial body, and is initially drawn through the AP. The final LOP is shifted towards or away from the celestial body by the amount equal to the arc difference between the computed and observed altitudes (recognizing that one minute of arc is approximately one nautical mile). If the computed altitude is greater than the observed altitude, we shift the LOP away from the celestial body along the azimuth line; if the computed altitude is less than the observed altitude we shift the LOP towards the celestial body, along the azimuth line. This is done for each celestial body, and the final fix is computed from the intersection of all LOPs. If the lines do not intersect in a single point, the “estimation of the ship’s position from the somewhat chaotic image of a number of position lines is left to the ‘keen judgement’ of the observer” [6].

B. HDOP

The lack of a common intersection of the LOPs in the overdetermined case is familiar to those in the GNSS community; the usual response is to reformulate the problem in a least squares sense, minimizing the residual error for each LOP. Since the equations are nonlinear, this involves choosing an initial assumed position, linearizing the LOP equations, solving for the least squares position, using this result as the next assumed position, and iterating. While this can be done in latitude/longitude coordinates [7], it is more convenient when describing positioning accuracy to employ East/North coordinates. Specifically, adapting the notation in [6], let $\delta_e$ and $\delta_n$ be the East and North least squares offsets to the solution from the AP, respectively, and $\delta_k$ be the difference between the measured and computed altitudes to the celestial object at azimuth $\theta_k$. Assuming $m$ measurements, the simultaneous equations are

$$
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\vdots \\
\delta_m
\end{bmatrix} =
\begin{bmatrix}
sin \theta_1 & cos \theta_1 \\
sin \theta_2 & cos \theta_2 \\
\vdots & \vdots \\
sin \theta_m & cos \theta_m
\end{bmatrix}
\begin{bmatrix}
\delta_e \\
\delta_n
\end{bmatrix}
$$

We note that these are precisely the equations used in the altitude-intercept method to specify the location of the LOPs. In other words the altitude-intercept method combined with the ‘keen eye’ of the observer are equivalent to the first iteration in a typical least squares position solution.

While the least squares solution of this set of simultaneous equations is found in [6], it is more recognizable to the GNSS community to write these equations in vector form

$$\delta = G \delta$$

where $\delta$ being the vector of altitude differences and $G$ the directions cosine matrix

$$G =
\begin{bmatrix}
sin \theta_1 & cos \theta_1 \\
sin \theta_2 & cos \theta_2 \\
\vdots & \vdots \\
sin \theta_m & cos \theta_m
\end{bmatrix}
\begin{bmatrix}
e_1 & n_1 \\
e_2 & n_2 \\
\vdots & \vdots \\
e_m & n_m
\end{bmatrix}$$
where we further simplify the notation using $e_k$ and $n_k$ as the East and North components of the unit vector pointing toward azimuth $\theta_k$. Assuming that the measurement errors are zero mean and uncorrelated with common variances $\sigma^2$, the least squares solution to this overdetermined set of equations is

$$
\begin{bmatrix}
\delta_e \\
\delta_n
\end{bmatrix} = (G^T G)^{-1} G^T \delta
$$

and the covariance matrix of the resulting position error is

$$
\text{Cov}\left(\begin{bmatrix}
\delta_e \\
\delta_n
\end{bmatrix}\right) = \sigma^2 (G^T G)^{-1}
$$

Since the solution is a 2-D fix, on the surface of the Earth, it is common to focus on the Horizontal Dilution of Precision or HDOP

$$
\text{HDOP} = \sqrt{\text{trace}\left\{ (G^T G)^{-1} \right\}}
$$

which is the square root of the sum of the variances in the East and North directions without the measurement variance scaling. The fact that $G$ has only two columns allows us to simplify this performance measure. Specifically, since

$$
G^T G = \begin{bmatrix}
\sum_{k=1}^{m} e_k^2 & \sum_{k=1}^{m} e_k n_k \\
\sum_{k=1}^{m} e_k n_k & \sum_{k=1}^{m} n_k^2
\end{bmatrix}
$$

we have

$$
\text{HDOP} = \sqrt{\frac{m}{\sum_{k=1}^{m} e_k^2 \sum_{k=1}^{m} n_k^2 - \left(\sum_{k=1}^{m} e_k n_k\right)^2}}
$$

(2)

As an example consider the mariner observing the sky at the twilight hours of Wednesday Jan. 30, 2019 somewhere to the east of Reston, VA; say at 39°N, 74°W, about 30-40 miles off of the coast from Wildwood, NJ. Using information from Her Majesty’s Nautical Almanac Office and the US Naval Observatory, 29 catalogued stars are readily visible. Their azimuths and elevations (rounded to the nearest whole degree) are:

azimuths = \{ 1 16 25 31 64 76 \\
95 97 114 122 122 125 \\
125 135 172 178 205 209 \\
212 219 227 260 265 269 \\
309 319 322 332 360 \}

and

elevations = \{ 39 7 19 76 60 27 \\
54 15 36 40 10 56 \\
33 31 54 10 3 72 \\
26 53 3 36 56 18 \\
26 4 63 9 23 \}

A sky plot showing these 29 locations appears in Figure 1.

Imagine that our goal is to select $m = 8$ stars for determining our position. A common question is which 8 to choose for best performance. With 29 available and $m = 8$, the problem is not too large; a full search requires testing a total of 4.3 million possible choices. Figure 2 shows the resulting histogram of the HDOP over all of the choices; the best HDOP of 0.70711 through the worst case value of 2.8020 (the constellation yielding the minimum HDOP is shown in blue in Figure 1). We note that this histogram is quite skewed toward the left; there are very many 8-star constellations with excellent performance!! Fully 50% have HDOP below 0.7195 and 90% with HDOP below 0.752; hence, even choosing stars at random is likely to produce a pretty good result!

III. WHAT DO THE BEST CONSTELLATIONS LOOK LIKE?

Some relevant questions for celestial navigation include “What is the best possible performance (minimum HDOP) achievable?” and “How should we select the celestial objects employed to achieve performance close to this best value?” In [9] we started answering these and related questions by first developing the lower bound

$$
\text{HDOP} \geq \sqrt{\frac{4}{m}}
$$

(3)
which is achieved by $m$ celestial objects whose azimuths satisfy two balance conditions

$$
\sum_{k=1}^{m} e_k n_k = 0 \quad \text{and} \quad \sum_{k=1}^{m} e_k^2 = \sum_{k=1}^{m} n_k^2 = \frac{m}{2}
$$

Next, we started to characterize star constellations that met these conditions. Specifically, we developed a series of results (proofs of these appear in [9]):

**Theorem 1.** Imagine that a set of azimuths $\{\theta_1, \theta_2, \ldots, \theta_m\}$ is balanced. Then rotating the entire constellation by common angle $\phi$ results in an equivalent balanced constellation. Similarly, negating all of the azimuths to $\{-\theta_1, -\theta_2, \ldots, -\theta_m\}$ yields another balanced constellation.

Since reflecting a constellation across any line through the origin is equivalent to a rotation, negation, and rotation, we can also form a balanced constellation via reflection. Further, since there are infinitely many choices for the rotation angle we can form equivalence classes of constellations based upon a rotation and reflection, with the need to describe only one version in each class.

**Theorem 2.** For $m \geq 3$ the constellation forming a regular $m$-gon (azimuths evenly distributed over $360^\circ$) is balanced. Without loss of generality the base constellation describing this class has azimuths

$$
\left\{0^\circ, \frac{360}{m}, \frac{2 \cdot 360}{m}, \frac{3 \cdot 360}{m}, \ldots, \frac{(m-1) \cdot 360}{m}\right\}
$$

We note that this solution certainly fits Bowditch’s notion of bearing spread.

**Theorem 3.** The union of two balanced constellations yields another balanced constellation.

Combining these three theorems we can build larger constellations from several smaller ones, each with its own arbitrary rotation; Figure 3 compares three such constellations:

- Left subfigure – 8 celestial objects each separated by $45^\circ$ (drawn at common elevation for visual simplicity since HDOP is not a function of the elevations).
- Central subfigure – two sub-constellations of 4 celestial objects. Each group can have an arbitrary rotation in azimuth; in fact, we could allow overlap meaning that a constellation of pairs of stars with $90^\circ$ azimuth separation would meet the HDOP bound.
- Right subfigure – one set of 5 objects each separated by $72^\circ$ and the remaining 3 each separated by $120^\circ$, each with arbitrary rotation.

For larger $m$ there are more ways to construct union constellations (e.g. for $m = 10$ we could have $10 \times 3$ or $6 + 4$ or $5 + 5$ or $4 + 3 + 3$), so there are many classes of balanced constellations, all meeting the HDOP bound. The theory of partitions [14] can help in enumerating the number of possibilities for larger $m$.

These prior results on balanced constellations can be extended by three new theorems:

**Theorem 4.** Imagine that a set of azimuths $\{\theta_1, \theta_2, \ldots, \theta_m\}$ is balanced. Then reflecting a single one of these azimuths though the origin, say $\theta_j$ to $\theta_j + 180^\circ$, results in a balanced constellation.

Proof. We check both conditions for the constellation with a single reflection; the method is primarily trigonometric manipulations. Without loss of generality let $j = 1$. Using hats to indicate the direction components of this modified constellation, the first condition is

$$
\sum_{k=1}^{m} \hat{e}_k \hat{n}_k = \sin(\theta_1 + 180^\circ) \cos(\theta_1 + 180^\circ) + \sum_{k=2}^{m} \sin \theta_k \cos \theta_k
$$

$$
= [- \sin \theta_1] [- \cos \theta_1] + \sum_{k=2}^{m} \sin \theta_k \cos \theta_k
$$

$$
= \sum_{k=1}^{m} \sin \theta_k \cos \theta_k = \sum_{k=1}^{m} e_k n_k
$$

The last expression is the first balance condition for the
original constellation, equal to zero by assumption, so this condition is also zero for the constellation with a single reflection. The second balance condition looks first at the sum of squares of the east components

\[ \sum_{k=1}^{m} e_k^2 = \sin^2 (\theta_1 + 180^\circ) + \sum_{k=2}^{m} \sin^2 \theta_k \]

\[ = [-\sin \theta_1]^2 + \sum_{k=2}^{m} \sin^2 \theta_k \]

\[ = \sin^2 \theta_1 + \sum_{k=2}^{m} \sin^2 \theta_k \]

\[ = \sum_{k=1}^{m} \sin^2 \theta_k = \sum_{k=1}^{m} e_k^2 \]

which is \( \frac{m}{2} \) by assumption. In the obvious, trivial way the sum of the squares of the north components for the constellation with one reflection is also \( \frac{m}{2} \) and the result is a balanced constellation.

This theorem naturally leads to several interesting classes of balanced constellations as follows:

**Theorem 5.** For \( m \) odd, the constellation class consisting of azimuths \( \left\{ 0^\circ, \frac{180}{m}, \frac{180}{m} - \frac{360}{m}, \ldots, \frac{180 - (m-1)}{m} \right\} \) is balanced.

**Proof.** The proof is by construction. Start with the regular \( m \)-gon constellation with azimuths \( \left\{ 0^\circ, \frac{360}{m}, \frac{360}{m} - \frac{360}{m}, \ldots, \frac{360 - (m-1)}{m} \right\} \) of Theorem 2 and reflect the \( \frac{m-1}{2} \) largest of these (those greater than \( 180^\circ \)) back into the range \( [0, 360^\circ] \); the result is the desired uniform spacing of \( \frac{180}{m} \) in azimuth.

Employing this theorem with \( m = 3 \) yields that a three point constellation with inter-azimuth spacing of \( 60^\circ \) is balanced; with \( m = 5 \), five stars with \( 36^\circ \) inter-azimuth spacing is balanced.

In these authors’ opinion the most useful result of the single reflection theorem is as follows:

**Theorem 6.** A pair of stars separated by \( 90^\circ \) in azimuth is balanced.

**Proof.** Following Theorem 4 this could be proved by construction using reflection (start with the balanced square constellation with azimuths \( \left\{ 0^\circ, 90^\circ, 180^\circ, 270^\circ \right\} \), reflect the last two yielding the balanced constellation with four azimuths \( \left\{ 0^\circ, 0^\circ, 90^\circ, 90^\circ \right\} \), argue that this can be split this into two two-star sub constellations each of which must be balanced); however, it is more direct to just compute the balance conditions. Assuming azimuths of \( 0^\circ \) and \( 90^\circ \)

\[ \sum_{k=1}^{2} e_k n_k = \cos^2 0^\circ + \sin 90^\circ \cos 90^\circ = 0 \]

\[ \sum_{k=1}^{2} e_k^2 = \sin^2 0^\circ + \sin^2 90^\circ = 1 \]

\[ \sum_{k=1}^{2} e_k^2 = \cos^2 0^\circ + \cos^2 90^\circ = 1 \]

and the two point constellation is balanced.

Sub-constellations of two stars with only a \( 90^\circ \) separation allows for the inclusion of meteorological limitations to real star sitting; i.e. allowing for portions of the sky and/or horizon to be obscured (as occurs in a fog bank, when near land, or with partial cloud cover). This is a significant improvement from Bowditch’s original recommendation of a uniform, 360 degree spread in azimuth, allowing the mariner to optimize his/her positioning accuracy even in limited sky conditions.

At this point we can attempt to characterize all classes of small sub-constellations (the statements below can all be shown to be true by simple, but tedious, algebraic operations):

- \( m = 2 \) – there is only one class, the pair with \( 90^\circ \) spacing.
- \( m = 3 \) – there are two classes, those with \( 120^\circ \) spacing and those with \( 60^\circ \) spacing.
- \( m = 4 \) – there are no new unique classes; all balanced \( m = 4 \) sub-constellations are formed from two \( m = 2 \) selections each with \( 90^\circ \) separation (but the two groups can have any arbitrary relative rotation; hence, this is a very rich class).
- \( m = 5 \) – there are at least four classes: five azimuths with \( 72^\circ \) spacing, five with \( 36^\circ \) spacing, and the two sub-sub constructions of \( 5 = 3 + 2 \). While it is possible that additional classes could exist, we have neither found them nor proved that they do not exist.
- \( m > 5 \) – we conjecture that the only interesting new classes are when \( m \) is prime, providing two new classes with \( \frac{360}{m} \) and \( \frac{180}{m} \) spacing, respectively.

In summary, the union-based construction of a balanced constellation of Theorem 3 has a much, much larger set of building blocks then imagined in [9]; for example, optimum constellations for the \( m = 8 \) example above could be constructed as \( 5 + 3, 3 + 3 + 2, \) or \( 2 + 2 + 2 + 2 \). Further, we note that the ability to have arbitrary rotations of each of the sub-constellations means that the optimum choices based upon the decomposition of \( 8 = 2 + 2 + 2 + 2 \) by far dominate the totality of optimum constellations (and we use this fact in the algorithm development below). Combining this richness of optimum constellations with the relative insensitivity of HDOP to a small mismatch (also noticed in [9]) helps to explain the empirical fact (seen in Figure 2) that the histogram of HDOP performance over all possible star selections is strongly biased toward good performance.

**IV. A SIMPLE STAR SELECTION ALGORITHM**

Recall that the goal is to select a subset of the visible stars to yield small HDOP. In [9] we developed a simple search algorithm which looked for multiple small, but nearly balanced sub-constellations with the intent of building a larger constellation as the union of smaller ones. For the night sky example shown in Figure 1 our best result using Theorems 1-3 consisted of a pair of 4-tuples; the resulting HDOP was
Fig. 4. Star selection from two 4-tuples; the resulting HDOP is 0.7075.

Fig. 5. Sky plot for the example. The red and blue circles show the azimuths and elevations of the full set of 29 star locations; the blue identifies the best set of \( m = 8 \) stars found using the pairwise search algorithm (with HDOP (0.7071). The azimuths for this set are shown in the title.

TABLE I

<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \Delta \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>121.9(^{\circ})</td>
<td>211.9(^{\circ})</td>
<td>0.0(^{\circ})</td>
</tr>
<tr>
<td>30.9(^{\circ})</td>
<td>121.0(^{\circ})</td>
<td>0.1(^{\circ})</td>
</tr>
<tr>
<td>218.9(^{\circ})</td>
<td>308.3(^{\circ})</td>
<td>0.6(^{\circ})</td>
</tr>
<tr>
<td>24.3(^{\circ})</td>
<td>113.2(^{\circ})</td>
<td>1.1(^{\circ})</td>
</tr>
<tr>
<td>177.0(^{\circ})</td>
<td>208.2(^{\circ})</td>
<td>1.2(^{\circ})</td>
</tr>
<tr>
<td>63.3(^{\circ})</td>
<td>331.8(^{\circ})</td>
<td>1.5(^{\circ})</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

0.7075 (very close to the bound of 0.70711). Theorems 4-6 provide for an alternative search method; specifically, search for pairs of stars separated by approximately 90\(^{\circ}\) keeping the \( \frac{m}{2} \) closest pairs. Specifically, the algorithm is:

1) For a total of \( N \) stars, compute the \( N(N-1) \) azimuth differences (remember to take into account the modulo 360\(^{\circ}\) wraparound).
2) Sort the list by those closest to 90\(^{\circ}\).
3) Select the \( \frac{m}{2} \) pairs at the top of the list.

For the \( N = 29 \) azimuths listed above this sorted list is shown in Table 1 in which \( \Delta \theta \) denotes the offset from a 90\(^{\circ}\) difference. The resulting constellation has HDOP = 0.70712; its sky plot is shown in Figure 5. While different then the solution shown in Figure 1, it has nearly equal performance.

Of, perhaps, greatest utility of this pairwise algorithm is its ability to find excellent constellations even if a large section of the sky (as measured in azimuth) is unavailable as happens when a portion of the horizon is occluded either by a land mass or a fog bank (or other meteorological event). As an example, consider the sky of the example above, but in which a 100\(^{\circ}\) swath of the NorthWest sky is obscured as suggested by Figure 6 (azimuths between 240\(^{\circ}\) and 340\(^{\circ}\)). The top subfigure shows the resulting sky with the full search best constellation (HDOP = 0.70711); the bottom subfigure shows the best pairwise constellation (HDOP = 0.70716). Figure 7 shows an even more extreme example; both the brute force and pairwise approaches find constellations with HDOP = 0.70711 (interestingly, slightly better than the pairwise result with more sky availability, but this just speaks to the slight suboptimality of the pairwise algorithm). It is pretty clear from the top subfigure in this case that the full search is also finding patterns with pairwise separation of 90\(^{\circ}\). While they are not identical solutions, the performance of the fast algorithm is equivalent.

V. SOLVING FOR TIME

In GNSS we get time (i.e. an estimate of UTC) from the solution algorithm; can we do the same with celestial? Web sites such as that shown in Figure 8 suggest that you can. However, as we demonstrate below, these “clocks” yield local time only (not traceable back to UTC, much like a sundial). In contrast, there is also the “Lunar Distance” method which estimates precise time from the movement of the moon relative to a star [16]. Below we examine estimating precise time from celestial measurements, demonstrating both of these facts and, further, presenting a way to predict the accuracy of any such estimate.

Referring to [6], [7], [17]–[20] we start with the single measurement equation for a celestial body

\[
\sin H = \sin \phi \sin d + \cos \phi \cos d \cos (G - \theta)
\]

in which \( \phi \) and \( \theta \) are, respectively, the user’s latitude and longitude (for which we are trying to solve), \( G \) is the celestial
Fig. 6. Sky plots under the assumption of 100° of sky occlusion. The top subfigure is the brute force best constellation while the bottom subfigure is the best pairwise constellation.

Fig. 7. Sky plots under the assumption of 180° of sky occlusion. The top subfigure is the brute force best constellation while the bottom subfigure is the best pairwise constellation.

Fig. 8. A typical star clock website [15].

The Star Clock is a simple, easy-to-use aid that lets you tell time using the stars.

Kelly Beatty

Object’s Greenwich hour angle, $d$ is the object’s declination (the latitude and longitude of the Ground Point under the object, respectively, both of which are tabulated if the time is known), and $H$ is the altitude measurement. Solving for the altitude

$$H = \sin^{-1} \{ \sin \phi \sin d + \cos \phi \cos d \cos (G(t) - \theta) \}$$

$$\equiv g(\phi, \theta)$$

we can linearize the measurement equation about some nominal point (with values $H_0, \phi_0, \theta_0$, the $\delta$’s denote the perturbations)

$$H_0 + \delta h \approx g(\phi_0, \theta_0) + \frac{\partial g}{\partial \phi}(\phi - \phi_0) + \frac{\partial g}{\partial \theta}(\theta - \theta_0)$$

to yield a differential measurement equation

$$\delta h \approx \left. \frac{\partial g}{\partial \phi} \right|_{\phi_0, \theta_0} \delta \phi + \left. \frac{\partial g}{\partial \theta} \right|_{\phi_0, \theta_0} \delta \theta$$

with

$$\frac{\partial g}{\partial \phi} = \frac{\cos \phi \sin d - \sin \phi \cos d \cos (G - \theta)}{\sqrt{1 - (\sin \phi \sin d + \cos \phi \cos d \cos (G - \theta))^2}}$$

and

$$\frac{\partial g}{\partial \theta} = \frac{\cos \phi \cos d \sin (G - \theta)}{\sqrt{1 - (\sin \phi \sin d + \cos \phi \cos d \cos (G - \theta))^2}}$$

If time, $t$, is unknown then the location of the celestial object’s ground point, $(d, G)$, should also vary with time, say $(d(t), G(t))$. For example, the Ground Point for stars on the celestial sphere rotate daily in longitude at constant latitude, so a simple model for the stars is

$$d(t) = d_0 \quad \text{and} \quad G(t) = G_0 + \kappa (t - t_0)$$

with $\kappa$ approximately equal to 15 degrees per hour (actually slightly more at 360° per sidereal day). Objects within our
solar system also move relative to the Earth; hence, have additional motion beyond the rotation of the celestial sphere. Figure 9 shows the declination and GHA over a 24 hour period for the standard visible celestial objects (sun, moon, visible planets, and 58 tabulated stars). On this scale the declinations appear to be constant and the GHAs to vary linearly. Figure 10 shows the results of fitting straight lines to this data. From the top subfigure we observe that the stars all have constant declination (the ‘blue’ slopes are all equal to zero), the Sun and planets have almost zero slopes (η ≈ 0), and only the moon has a significant declination change. From the bottom subfigure we observe that the stars all have slope of κ = 15.04 degrees per hour, the Sun and three inner planets have somewhat different slopes, and the moon again shows the largest difference. The quality of these fits suggest models of the form

\[ d(t) = d_0 + \eta (t - t_0) \quad \text{and} \quad G(t) = G_0 + \kappa (t - t_0) \]

Further, we will add subscripts to index the different celestial objects as needed. Adding in this time dependence we have the linearized perturbation equation

\[ \delta h \approx \frac{\partial g}{\partial \phi} \bigg|_{\phi_0, \theta_0, t_0} \delta \phi + \frac{\partial g}{\partial \theta} \bigg|_{\phi_0, \theta_0, t_0} \delta \theta + \frac{\partial g}{\partial t} \bigg|_{\phi_0, \theta_0, t_0} \delta t \]

with

\[ \frac{\partial g}{\partial t} = \eta (\sin \phi \cos d - \cos \phi \sin d \cos (G - \theta)) \frac{\Delta}{\Delta} - \kappa \cos \phi \cos d \sin (G - \theta) \frac{\Delta}{\Delta} \]

in which

\[ \Delta = \sqrt{1 - (\sin \phi \sin d + \cos \phi \cos d \cos (G - \theta))^2} \]

While this linearized model is interesting, it is more insightful to recast it in a local coordinate frame (East, North, Up versus latitude and longitude). We can accomplish this by a change of variables.

First, starting with latitude \( \phi \), longitude \( \theta \), and height \( h \) (= 0) and, assuming a spherical Earth of radius \( r_e \) (= 3959 miles), we can convert to ECEF (Earth Centered Earth Fixed) coordinates by

\[
\begin{align*}
x &= r_e \cos \phi \cos \theta \\
y &= r_e \cos \phi \sin \theta \\
z &= r_e \sin \phi
\end{align*}
\]

A small perturbation in latitude and longitude, \( \phi \to \phi + \delta \phi \) and \( \theta \to \theta + \delta \theta \), yields the corresponding perturbations in ECEF

\[
\begin{align*}
\delta x &\approx -r_e \cos \theta \sin \phi \delta \phi - r_e \sin \theta \cos \phi \delta \theta \\
\delta y &\approx -r_e \sin \theta \sin \phi \delta \phi + r_e \cos \theta \cos \phi \delta \theta \\
\delta z &\approx r_e \cos \phi \delta \phi
\end{align*}
\]

Next, the conversion from ECEF to ENU (East, North, Up) at the point \((\phi, \theta, h)\) is a linear transformation

\[
\begin{bmatrix}
e \\
n \\
u
\end{bmatrix}
= \begin{bmatrix}
-\sin \theta & \cos \theta & 0 \\
-\sin \phi \cos \theta & -\sin \phi \sin \theta & \cos \phi \\
\cos \phi \cos \theta & \cos \phi \sin \theta & \sin \phi
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z
\end{bmatrix}
\]
The differentials in this ENU frame are
\[
\begin{bmatrix}
\delta u \\
\delta n \\
\delta e
\end{bmatrix} = \begin{bmatrix}
-\sin \theta & \cos \theta & 0 \\
-\sin \phi \cos \theta & -\sin \phi \sin \theta & \cos \phi \\
\cos \phi \cos \theta & \cos \phi \sin \theta & \sin \phi
\end{bmatrix} \begin{bmatrix}
-r_e \cos \theta \sin \phi \delta \phi - r_e \sin \phi \cos \phi \delta \theta \\
r_e \sin \phi \sin \phi \delta \phi + r_e \cos \phi \cos \phi \delta \theta \\
r_e \cos \phi \delta \phi
\end{bmatrix}
\]
\[
\times \begin{bmatrix}
r_e \cos \phi \delta \phi \\
r_e \sin \phi \delta \phi \\
0
\end{bmatrix}
\]
so
\[
\delta e \approx r_e \cos \phi \delta \phi \quad \text{and} \quad \delta n \approx r_e \delta \phi
\]
or
\[
\delta \phi \approx \frac{\delta n}{r_e} \quad \text{and} \quad \delta \theta \approx \frac{\delta e}{r_e \cos \phi}
\]

Returning to the perturbation model
\[
\delta h \approx \frac{\partial h}{\partial \phi} \delta \phi + \frac{\partial h}{\partial \theta} \delta \theta + \frac{\partial h}{\partial t} \delta t
\]
\[
\approx \frac{\cos \phi_0 \sin d_0 - \sin \phi_0 \cos d_0 \cos (G_0 - \theta_0)}{r_e \Delta} \delta n
\]
\[
+ \frac{\cos \phi_0 \sin d_0 - \sin \phi_0 \cos d_0 \cos (G_0 - \theta_0)}{r_e \Delta} \delta e
\]
\[
+ \frac{\eta (\sin \phi_0 \cos d_0 - \cos \phi_0 \sin d_0 \cos (G_0 - \theta_0))}{\Delta} \delta t
\]
\[
- \frac{\kappa \cos \phi_0 \cos d_0 \sin (G_0 - \theta_0)}{\Delta} \delta t
\]

As noted above in Section II.B, it is usual in celestial navigation to use the azimuth from the user’s position toward the celestial object’s Ground Point; this is the full 360° arc tangent with components \( \cos d_0 \sin (G - \theta_0) \) and \( \cos \phi_0 \sin d_0 - \sin \phi_0 \cos d_0 \cos (G - \theta_0) \). Equivalently, the East and North components are proportional to the individual terms
\[
east = \sin \az = \frac{\cos \phi_0 \sin d_0 - \sin \phi_0 \cos d_0 \cos (G_0 - \theta_0)}{\beta}
\]
\[
north = \cos \az = \frac{\cos d_0 \sin (G_0 - \theta_0)}{\beta}
\]
in which \( \beta \) is the constant
\[
\beta = \sqrt{\left(\frac{\cos \phi_0 \sin d_0 - \sin \phi_0 \cos d_0 \cos (G_0 - \theta_0)}{\beta}\right)^2 + \left(\frac{\cos d_0 \sin (G_0 - \theta_0)}{\beta}\right)^2}
\]

A tedious trigonometric argument shows that this is the same as the square root term in the denominator of the East and North derivative expressions, \( \beta = \Delta \), so the linearized equation becomes
\[
\delta h \approx \frac{\sin \az_0}{r_e} \delta n + \frac{\cos \az_0}{r_e} \delta e
\]
\[
+ \eta (\sin \phi_0 \cos d_0 - \cos \phi_0 \sin d_0 \cos (G_0 - \theta_0)) \delta t
\]
\[
- \kappa \cos \phi_0 \cos \az_0 \delta t
\]

Similarly, the first \( \delta t \) term’s coefficient can be related to what we’ll call the complementary azimuth, \( \az^c \), the azimuth from the celestial object’s Ground Point back to the observer’s position (note that in spherical trigonometry this is not, in general, just a 180° rotation)
\[
\frac{\sin \phi_0 \cos d_0 - \cos \phi_0 \sin d_0 \cos (G_0 - \theta_0)}{\Delta} = \sin \az^c
\]
so
\[
\delta h \approx \frac{\sin \az_0}{r_e} \delta n + \frac{\cos \az_0}{r_e} \delta e
\]
\[
+ (\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t
\]

Finally, it is convenient to write these equations in units of distance. If \( \delta h \) is measured in arc minutes (as is typical in celestial navigation), then multiplying by \( r_e \) yields units of nautical miles \((r_e \cdot \frac{2\pi}{360} = 1 \text{ nm})\), so
\[
\delta h_{\text{nm}} \approx \sin \az_0 \delta n + \cos \az_0 \delta e
\]
\[
+ r_e (\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t
\]

Matching units, the right hand side must also be in units of nautical miles. Examining the first and second terms on the right, the resulting \( \delta n \) and \( \delta e \) will each have units of nautical miles since the sine and cosine terms are unitless. The remaining terms, \( r_e \eta \sin \az^c \delta t \) and \( -r_e \kappa \cos \phi_0 \cos \az_0 \delta t \), need some explanation. In Figure 10, \( \eta \) and \( \kappa \) are defined with units of degree per hour, so the products have units of miles * degrees/hour * unit of \( \delta t \). Multiplying by \( \frac{2\pi}{360} \) converts these terms to
\[
\frac{2\pi}{360} r_e (\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t
\]
with units of miles * radians/hour * unit of \( \delta t \). Multiplying now by \( \frac{1}{60} \) converts these last two terms to
\[
\frac{2\pi}{360 \cdot 60} r_e (\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t
\]
with units of miles * radians/minute of time * unit of \( \delta t \). However, as noted above, \( \frac{2\pi}{360} r_e \) reduces to 1 nm, so the terms are
\[
(\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t
\]
with units of nautical miles per minute of time * unit of \( \delta t \). In other words, the perturbation equation is of the form
\[
\delta h_{\text{nm}} \approx \sin \az_0 \delta n + \cos \az_0 \delta e
\]
\[
+ (\eta \sin \az^c - \kappa \cos \phi_0 \cos \az_0) \delta t_{\text{min}}
\]
with \( \delta e \) and \( \delta n \) measured in nautical miles and \( \delta t \) measured in minutes (hence, the subscripts).

In vector-matrix form, for \( m \) measurements we have
\[
\begin{bmatrix}
\delta h_1 \\
\delta h_2 \\
\vdots \\
\delta h_m
\end{bmatrix} = \begin{bmatrix}
\sin \az_1 & \cos \az_1 & \text{term}_1 \\
\sin \az_2 & \cos \az_2 & \text{term}_2 \\
\vdots & \vdots & \vdots \\
\sin \az_m & \cos \az_m & \text{term}_m
\end{bmatrix} \begin{bmatrix}
\delta n \\
\delta e \\
\delta t
\end{bmatrix}
\]
TABLE II
DOP VALUES UNDER VARIOUS CELESTIAL OBJECT CHOICES.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Time known</th>
<th>HDOP bound</th>
<th>Time unknown</th>
<th>Position known</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EDOP</td>
<td>NDOP</td>
<td>H DOP</td>
<td>EDOP</td>
</tr>
<tr>
<td>stars (16) only</td>
<td>0.35511</td>
<td>0.36025</td>
<td>0.50585</td>
<td>0.5000</td>
</tr>
<tr>
<td>stars + Mercury</td>
<td>0.35272</td>
<td>0.34630</td>
<td>0.49431</td>
<td>0.48507</td>
</tr>
<tr>
<td>stars + Mercury + Venus</td>
<td>0.35270</td>
<td>0.32890</td>
<td>0.48507</td>
<td>0.47140</td>
</tr>
<tr>
<td>stars + Mercury + Venus + moon</td>
<td>0.34792</td>
<td>0.31215</td>
<td>0.46742</td>
<td>0.44721</td>
</tr>
<tr>
<td>stars + moon (4 at 5 minute intervals)</td>
<td>0.34130</td>
<td>0.29600</td>
<td>0.45178</td>
<td>0.44721</td>
</tr>
<tr>
<td>stars + moon (4 at 15 minute intervals)</td>
<td>0.33833</td>
<td>0.29809</td>
<td>0.45092</td>
<td>0.44721</td>
</tr>
<tr>
<td>sun + moon (4 each at 15 minute intervals)</td>
<td>1.4707</td>
<td>0.43848</td>
<td>1.5347</td>
<td>0.70711</td>
</tr>
<tr>
<td>moon only (4 at 5 minute intervals)</td>
<td>24.018</td>
<td>8.7772</td>
<td>25.631</td>
<td>1.0000</td>
</tr>
<tr>
<td>moon only (4 at 15 minute intervals)</td>
<td>7.6123</td>
<td>3.2783</td>
<td>8.2882</td>
<td>1.0000</td>
</tr>
<tr>
<td>stars + moon (9 at 15 minute intervals)</td>
<td>0.31325</td>
<td>0.27028</td>
<td>0.41373</td>
<td>0.40000</td>
</tr>
<tr>
<td>stars + moon (13 at 15 minute intervals)</td>
<td>0.28644</td>
<td>0.26635</td>
<td>0.39114</td>
<td>0.37139</td>
</tr>
</tbody>
</table>

in which we employ the notation

\[ t \text{term}_k = \eta_k \sin \alpha - \kappa_k \cos \phi_0 \cos \alpha_0 \]

for the \( k \)-th celestial object. This is the extension of (1) adding in the appropriate column for the unknown time (different, of course, from the column of ones for the unknown clock bias in GNSS). An alternative, compact form is

\[ \delta \mathbf{x} = \mathbf{G} \delta \mathbf{x} \]

in which \( \delta \mathbf{x} \) represents the desired position/time update vector. Solving for this update in a least squares sense

\[ \delta \mathbf{x} = (\mathbf{G}^{T} \mathbf{G})^{-1} \mathbf{G}^{T} \delta \mathbf{h} \]

The covariance matrix of the resulting position/time error is

\[ \text{Cov}(\delta \mathbf{x}) = \sigma^2 (\mathbf{G}^{T} \mathbf{G})^{-1} \]

in which \( \sigma^2 \) is the common variance of the assumed independent star altitudes; the diagonal elements of \( (\mathbf{G}^{T} \mathbf{G})^{-1} \) are the squares of the East DOP (EDOP), North DOP (NDOP), and time DOP (TDOP), respectively. The full DOP is the square root of the trace of this covariance matrix

\[ \text{DOP} = \sqrt{\text{EDOP}^2 + \text{NDOP}^2 + \text{TDOP}^2} \]

For comparison to the results above, the HDOP is the L2-norm of the first two of these diagonal elements

\[ \text{HDOP} = \sqrt{\text{EDOP}^2 + \text{NDOP}^2} \]

Let’s consider this result:

- First, imagine that all \( \eta_k = 0 \), i.e. that we are examining only stars (i.e. not the moon or sun or any planets), so that \( \kappa_k \) is a constant, say \( \kappa \). The matrix \( \mathbf{G} \) simplifies to

\[
\mathbf{G} = \begin{bmatrix}
\sin \text{az}_1 & \cos \text{az}_1 & -\kappa \cos \phi_0 \cos \text{az}_1 \\
\sin \text{az}_2 & \cos \text{az}_2 & -\kappa \cos \phi_0 \cos \text{az}_2 \\
: & : & : \\
\sin \text{az}_m & \cos \text{az}_m & -\kappa \cos \phi_0 \cos \text{az}_m
\end{bmatrix}
\]

and immediately we note that the second and third columns are linearly dependent; hence, the matrix \( \mathbf{G} \) is rank deficient and the matrix equation is unsolvable. The result is that it is impossible to estimate both position and time from only measurements of the stars (i.e. the longitude and time variables are non-separable) and “star clocks” cannot provide precise time as was stated at the beginning of this section.

- Now consider \( \eta_k \neq 0 \) for at least one celestial object. In general, \( \mathbf{G} \) is now full rank, but the product \( \mathbf{G}^{T} \mathbf{G} \) might still be poorly conditioned yielding high DOP. Table II shows the results of computing DOP for a variety of celestial combinations as would be seen on June 1, 2020 at a location off of the coast of Oregon:

- Each row shows the performance for three problems: (1) time known, but position unknown (the traditional celestial navigation problem); (2) time and position unknown (i.e. celestial navigation without a clock); and (3) position known, time unknown (a celestial timing receiver).

- The first 4 rows assume that all of the celestial objects’ elevations are measured simultaneously; the remainder assume sightings over time for the moon or sun (the stars are still just one simultaneous measurement). We focus a number of these examples at moon sightings to show the value of these results to the evaluation of the performance of the Lunar Distance method.

- The entry “stars” means that the 16 brightest stars are employed (Alioth, Arcturus, Bellatrix, Betelgeuse, Capella, Deneb, Denebola, Dubhe, Elnath, Kochab, Mirfak, Polaris, Pollux, Procyon, Regulus, and Schedar).

- In the case of known time, the resulting HDOP pretty well tracks the lower bound, \( \sqrt{\frac{1}{m}} \), as long as the stars are involved; situations with only the sun and moon (or moon alone) have poor HDOP since the time separated measurements of these objects do not provide much azimuth spread unless the time separations become large.
- When time is unknown we observe relatively good NDOP whenever the 16 stars are involved (this is not too surprising in that Polaris by itself provides good information on latitude without any knowledge of time); the near inseparability of the estimates of time and East manifests itself in the much larger values for EDOP and TDOP.

- The very last row, stars and multiple moon measurements, shows that one can improve the accuracies of the East and time estimates, but that the performance without knowledge of time is still quite poor for this significant amount of measurement data; hence, the limited movement of the natural solar system's bodies makes for poor time estimation. It might be interesting to see how man-made bodies, such as a space station, might contribute to our ability to read time from the sky.

- The final column considers the accuracy of a time estimate assuming a precisely known position. Since the standard deviation of the altitude measurement is typically about 0.2 nm (0.2 minutes of arc), then the standard deviation of the time estimate for many of the presented situations is approximately 25 seconds (0.2 ± 2 ± 60), about 2 minutes for the moon alone (which is quite close to the empirical accuracy of the Lunar Method mentioned in the literature [16]). While not stratum 1, still not too bad for such a simple method.

VI. Conclusions

The primary significance of this paper is that it corrects the misconception, starting with Bowditch, that the stars selected in celestial navigation need to be “evenly” distributed across 360 degrees of azimuth (and this same misconception exists for GNSS with users thinking that the satellites should uniformly cover the sky); that multiple stars from nearly similar directions (i.e. Polaris and stars in the Big Dipper) can be very useful together in forming a good celestial fix. As a secondary contribution, we mathematically formulate what can and cannot be estimated about time from the observation of celestial objects.

It is obvious that the development could be extended to a more general noise model. Specifically, we assumed a common variance on the measurement noise for each equation; explicitly writing a noise term \( \epsilon_k \) for each altitude measurement

\[
\delta_k = \sin \theta_k \delta_e + \cos \theta_k \delta_n + t \text{time}_k \delta t + \epsilon_k
\]

we have been operating under the assumption that \( E \{ \epsilon_k \} = 0 \) and \( E \{ \epsilon_k^2 \} = \sigma^2 \). This could be extended as follows:

- Recognizing that each celestial object is of a different brightness, that lower elevation angles are subject to more atmospheric uncertainty, and that parts of the sky and/or horizon may be harder to see clearly, it would be more natural to allow for unequal variances, \( E \{ \epsilon_k^2 \} = \sigma_k^2 \).

Allowing for this variation, the natural solution is a weighted least squares

\[
\begin{bmatrix}
\delta_e \\
\delta_n
\end{bmatrix} = (G^T \Lambda^{-1} G)^{-1} G^T \Lambda^{-1} \delta
\]

with \( \Lambda = \text{diag} \{ \sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2 \} \).

- The assumption of a zero mean for each error term can be relaxed somewhat to allowing for a common bias in the altitude measurements (e.g. perhaps due to a systematic error in the sextant) [21]; the result is to add a column of ones to \( G \), similar to the ones in the direction cosines matrix in GNSS, with the result that the least squares solution also estimates this bias.

References