THE UNIVERSITY OF RHODE ISLAND

University of Rhode Island
DigitalCommons@URI

Department of Electrical, Computer, and Biomedical Engineering Faculty Publications

Department of Electrical, Computer, and Biomedical Engineering

2019

Rethinking Star Selection in Celestial Navigation

Peter F. Swaszek
University of Rhode Island, swaszek@uri.edu

Richard J. Hartnett

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/ele_facpubs

The University of Rhode Island Faculty have made this article openly available.
Please let us know how Open Access to this research benefits you.

This is a pre-publication author manuscript of the final, published article.

Authors

Peter F. Swaszek, Richard J. Hartnett, and Kelly C. Seals

Rethinking Star Selection for Celestial Navigation

Peter F. Swaszek, *University of Rhode Island*
Richard J. Hartnett, *U.S. Coast Guard Academy*
Kelly C. Seals, *U.S. Coast Guard Academy*

ABSTRACT

In celestial navigation the altitude (elevation) angles to multiple celestial bodies are measured; these measurements are then used to compute the position of the user on the surface of the Earth. Methods described in the literature include the classical "altitude-intercept" algorithm as well as direct and iterative least-squares solutions for over determined situations. While it seems rather obvious that the user should select bright stars scattered across the sky, there appears to be no established results on the level of performance that is achievable based upon the number of stars sighted nor how the "best" set of stars might be selected from those visible. This paper addresses both of these issues by examining the performance of celestial navigation noting its similarity to the performance of GNSS systems; specifically, modern results on GDOP for GNSS are adapted to this classical celestial navigation problem.

INTRODUCTION

In the world of GNSS, position accuracy is often described by the geometric dilution of precision, or GDOP [1]. 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 of position.

Many papers in the navigation literature have considered GDOP: early investigations computed the GDOP as a function of time and location (on the surface of the Earth) to show the quality of the position performance achievable (e.g. [1, p.211]); the mathematics of the GDOP measure have been considered (e.g. [2]); and lower bounds on GDOP have been established as a function of number of satellites (e.g. [3]). A related problem is satellite selection for best GDOP. Specifically, in these days of multiple, full constellations, there could be times when there are many more satellites available than the receiver needs or is able to handle (e.g. [4], where GBAS channel bandwidth limitation was an issue); hence, some sort of satellite selection process is used, selecting those satellites that provide good (or best) performance. For example, early GPS receivers employed only 4 satellites to provide a solution while 6 or more satellites were typically available; multiple authors considered the best way to choose those 4 (e.g. [5]). Since then, many papers in the navigation literature have considered methods to find the best, or nearly best, subset of satellites both from single and multiple constellations (e.g. [6–14]). It is also possible to extend these approaches to horizontal DOP (HDOP) for navigating on the surface of the Earth.

For 2-D positioning on the surface of the earth, one can also consider measuring altitudes of celestial bodies (elevation angles, relative to an artificial or visible horizon, measured with a sextant), to locate the user [15–18]. Assuming multiple measurements of star elevations, solving for position can be cast as an overdetermined set of equations; hence, it is natural to employ a least squares solution. Whatever the position units, latitude and longitude or East and North in a local coordinate frame, these equations are nonlinear so can be solved via an iterative, linearized approach (just like in GNSS) [19, 20]. Presumably a user is interested in the "most accurate" position estimate, which would suggest that the geometry of the celestial bodies chosen for measurement is important. In fact, guidance provided in Bowditch [15, p.276] is to *". . . Predict expected altitudes and azimuths for up to eight bodies when preparing to take celestial sights. Choose the stars and planets that give the best bearing spread. Try to select bodies with a predicted altitude between 30 degrees and 70 degrees."* (Underline added for emphasis.) While this guidance seems reasonable, it is a bit ambiguous as to what constitutes a "best bearing spread," and the general interpretation has been to

choose visible celestial bodies that are fairly evenly spread around the horizon in azimuth. As we show below, this is a sufficient condition for the best geometry, but it is not a necessary condition, so can be overly restrictive.

The contents of this paper are:

- We very briefly review the fundamental methods of celestial navigation, highlight a mathematically precise definition of HDOP for the celestial navigation problem similar to that used in GNSS applications, and note the similarity of the classical altitude-intercept method to current GNSS positioning algorithms.

- We minimize this HDOP expression over the choice of the stars' locations; the results are a set of necessary conditions on the stars' azimuths and an achievable lower bound to HDOP as a function of the number of stars employed (parameterized by m). These conditions demonstrate the limitations of the conventional wisdom on star selection (and, hence, motivate the title of this paper). Specifically, these conditions do not, in fact, require that the m celestial observations be separated by $360/m$ degrees of azimuth in order to minimize HDOP. No longer requiring an even spread around the azimuth circle provides added flexibility to a navigator, especially if there is lack of a visible horizon in some direction (due to land, fog, or lighting conditions).

- We present a simple star selection algorithm for HDOP minimization. Of the roughly 9000 stars visible to the naked eye, 58 stars (including Polaris) plus 4 planets (Venus, Mars, Jupiter, Saturn), plus the Sun and Moon, have been selected for tabulation in each year's production of the Nautical Almanac [21]. These were chosen based upon visibility and ease of identification. Imagine if approximately one-half of these were visible (say 30); an exhaustive search for those 8 celestial objects that minimize HDOP is a "30 choose 8" problem, involving an evaluation of 5.8 million possible subsets of objects for minimum HDOP. Exploiting our knowledge of the HDOP formulation, we present a real time algorithm with significantly reduced complexity for selecting a near-optimum subset of stars for position determination.

- Finally, we include a real-world example, taken from tabulated visible celestial bodies in the 2019 Nautical Almanac, and show the utility of our results. We highlight that the celestial bodies selected in the optimal solution are, in fact, not necessarily equally spread in azimuth.

REVIEW OF CELESTIAL NAVIGATION

Celestial Navigation to a mariner is considered to be the art of calculating a 2-D position on the surface of the earth with the aid of the sun, moon, planets, and major stars. For aviators it typically means calculating that 2-D position relative to a known (barometric) altitude above the earth. In either case it usually means the navigator is obtaining lines of position (LOPs) by measuring the altitudes of celestial bodies relative to an actual or artificial horizon, typically by using a sextant. The general strategy in celestial navigation is to (1) choose a set of reasonable celestial bodies for measurement, based on visibility, azimuth, and altitude, (2) measure observed altitudes of that set of celestial bodies at known times, (3) correct altitude measurements for effects such as refraction, height of eye, or parallax, and (4) determine each celestial body's location on the celestial sphere at the observation time (e.g. declination and Greenwich Hour Angle). Each celestial body's altitude measurement creates a "circle of equal altitude" on the surface of the earth. The intersection of these circular lines of position (LOPs) results in a unique celestial fix. For a more local "small-scale" position solution, and with observed altitudes below 75 degrees, one can draw these "circular arcs" as actual lines of position, and this is what is typically done in maritime navigation.

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 we use known celestial body locations plus an assumed position (AP), presumably reasonably close to the actual position, to calculate the computed 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/celnavtable.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 (LOP). 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 in the amount of 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 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 did 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" [20].

The lack of a common intersection 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 [19], is more convenient when describing positioning accuracy to employ East/North coordinates. Specifically, adapting the notation in [20], let δ_e and δ_n be the East and North least squares offsets to the solution from the AP, respectively, and δ_k be the difference between the measured and computed altitudes to the celestial object at azimuth θ_k. The simultaneous equations are

$$\delta_k = \delta_e \sin\theta_k + \delta_n \cos\theta_k \qquad k = 1, 2, \ldots m$$

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 [20], it is more recognizable to the GNSS community to write these equations in vector form

$$\begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{bmatrix} = \begin{bmatrix} \sin\theta_1 & \cos\theta_1 \\ \sin\theta_2 & \cos\theta_2 \\ \vdots & \vdots \\ \sin\theta_m & \cos\theta_m \end{bmatrix} \begin{bmatrix} \delta_e \\ \delta_n \end{bmatrix} \qquad \text{or} \qquad \boldsymbol{\delta} = \mathbf{G} \begin{bmatrix} \delta_e \\ \delta_n \end{bmatrix}$$

$\boldsymbol{\delta}$ being the vector of altitude differences and $\mathbf{G}$ the directions cosine matrix

$$\mathbf{G} = \begin{bmatrix} \sin\theta_1 & \cos\theta_1 \\ \sin\theta_2 & \cos\theta_2 \\ \vdots & \vdots \\ \sin\theta_m & \cos\theta_m \end{bmatrix} = \begin{bmatrix} e_1 & n_1 \\ e_2 & n_2 \\ \vdots & \vdots \\ e_m & n_m \end{bmatrix}$$

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 θ_k. Assuming that the measurement errors are zero mean and uncorrelated with common variances σ^2, the least squares solution to this overdetermined set of equations is

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

The covariance matrix of the resulting position error is

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

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

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

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 $\mathbf{G}$ has only two columns allows us to simplify this performance measure. Specifically, since

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

we have

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

We note that these expressions could be modified to account for different measurement accuracies (e.g. replace σ^2 with individual σ_k^2's, perhaps functions of the altitudes themselves to accommodate atmospheric effects), essentially yielding a weighted least squares solution. It is also possible to include a common bias in the altitude measurements (e.g. perhaps due to a systematic error in the sextant) [22]; the result is to add a column of ones to $\mathbf{G}$, similar to the ones in the direction cosines matrix in GNSS, with the result that the least squares solution also estimates this bias. Neither of these modifications is pursued here, but is listed in our future work section below.

A LOWER BOUND TO HDOP

Imagine that we could place celestial objects wherever we wanted in the sky. One question that comes to mind is *"What is the best possible performance (minimum HDOP)?."* A second is *"Where should the celestial objects be placed to achieve this best value?."* The form for HDOP in Eq. (1) allows us to answer both questions.

Theorem 1. *A lower bound to the HDOP is*

$$HDOP \geq \sqrt{\frac{4}{m}} \tag{2}$$

and this value is achieved by m celestial objects whose azimuths satisfy two balance conditions

$$\sum_{k=1}^{m} e_k n_k = 0 \qquad and \qquad \sum_{k=1}^{m} e_k^2 = \sum_{k=1}^{m} n_k^2 = \frac{m}{2} \tag{3}$$

Proof. First, note that the HDOP in Eq. (1) is dependent upon the locations of the celestial objects only in its denominator; hence, minimizing HDOP is equivalent to maximizing the denominator, or maximizing its square

$$\max \left\{ \left(\sum_{k=1}^{m} e_k^2\right)\left(\sum_{k=1}^{m} n_k^2\right) - \left(\sum_{k=1}^{m} e_k n_k\right)^2 \right\}$$

Introducing the notation

$$a = \frac{1}{m}\sum_{k=1}^{m} e_k^2 \qquad b = \frac{1}{m}\sum_{k=1}^{m} n_k^2 \qquad d = \frac{1}{m}\sum_{k=1}^{m} e_k n_k \tag{4}$$

we have $b = 1 - a$ (since $e_k^2 + n_k^2 = 1$) and the maximization becomes

$$\max_{a,d} \left\{ m^2 \left[a(1-a) - d^2 \right] \right\}$$

a function of two unknowns, a and d. The derivatives of this expression with respect to the unknowns are $\frac{\partial}{\partial a} = m^2(1-2a)$ and $\frac{\partial}{\partial d} = -2m^2 d$ which equal zero at the single solution $d = 0$ and $a = \frac{1}{2}$ (so $b = \frac{1}{2}$ as well). The matrix of second partial derivatives

$$J = \begin{bmatrix} -2m^2 & 0 \\ 0 & -2m^2 \end{bmatrix}$$

is trivially negative definite; hence, a unique minimum of HDOP occurs when the stated balance conditions are satisfied. Substituting in the best choices for a and d yields the lower bound. $\qquad\square$

Answering the two posed question, the minimum HDOP is inversely proportional to the square root of the number of stars, Eq. (2), and that the best set(s) of star locations satisfy the azimuth balance conditions in Eq. (3).

CHARACTERIZING GOOD STAR LOCATIONS

The best HDOP is provided by star selections that satisfy the above balance conditions; to start to understand this better we consider various facts about balance through several theorems. Their proofs, generally trigonometric arguments, are relegated to an appendix.

Theorem 2. *Imagine that a set of azimuths $\{\theta_1, \theta_2, \ldots, \theta_m\}$ is balanced. Then rotating the entire constellation by angle ϕ 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 3. *For $m \geq 3$ the constellation forming a regular m-gon (azimuths evenly distributed over $360°$) is balanced.*

Without loss of generality the constellation has azimuths

$$\left\{0°, \frac{360}{m}, 2\frac{360}{m}, 3\frac{360}{m}, ..., (m-1)\frac{360}{m}\right\}$$

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

Theorem 4. *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 1 compares three such constellations:

- Left subfigure – 8 celestial objects each separated by $45°$ as stipulated by Theorem 3 (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 (Theorem 3 again). Each group can have an arbitrary rotation in azimuth.

- Right subfigure – one set of 5 objects each separated by $72°$ and the remaining 3 each separated by $120°$, again both with arbitrary rotations.

For larger m more and more complicated constructions are possible (e.g. for $m = 10$ we could have 10 or $7 + 3$ or $6 + 4$ or $5 + 5$ or $4 + 3 + 3$). In other words, for larger m there are many classes of balanced constellations; hence, many meeting the HDOP bound.
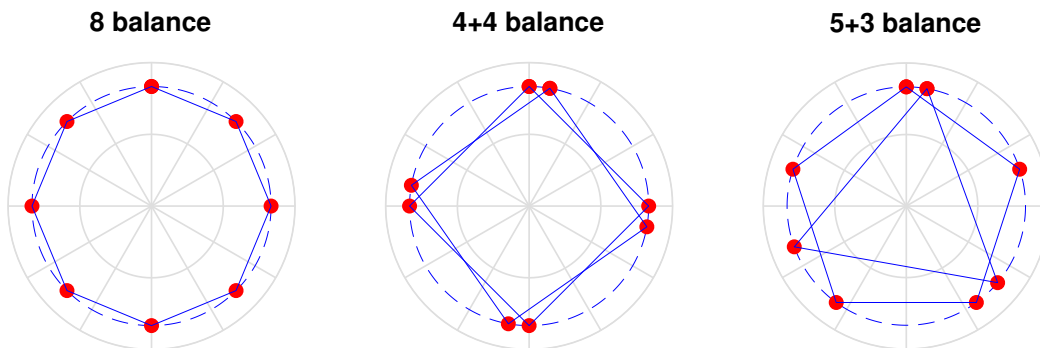

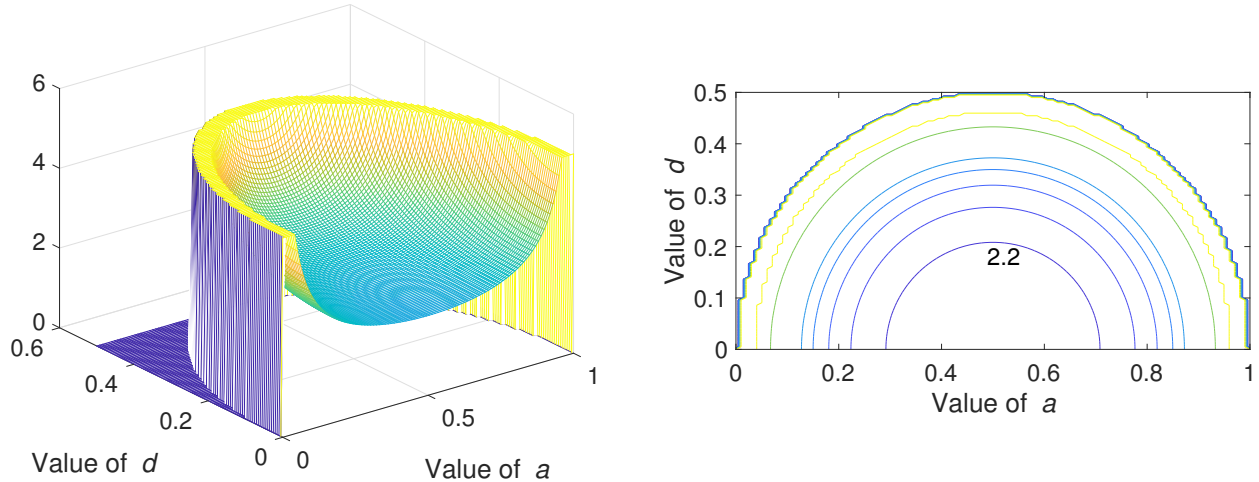Figure 1: Skyviews of optimum 8 star configurations.

Figure 2: Mesh and contour plots showing the impact of lack of balance.

THE LACK OF BALANCE

What happens if we don't balance? Can we be close enough to still get good HDOP performance? Using the definitions in Eq. (4) the HDOP is

$$\text{HDOP} = \sqrt{\frac{1}{m\left(ab - d^2\right)}} = \sqrt{\frac{1}{m\left(a(1-a) - d^2\right)}}$$

The constraints on these variables are: $0 \leq a \leq 1$ and $d^2 < a(1-a)$; clearly we can also restrict attention to $|d| \leq \frac{1}{2}$. Figure 2 shows surface and contour plots of the square root of m times the HDOP for the full range of a and the positive values of d (the full images are even functions of d and this reduction makes the characteristics of the results near $d = 0$ much easier to see). In the mesh plot (left) the HDOP coefficient is truncated at 5 to aid visualization as it becomes unbounded as a and d go toward their extreme values. In the contour plot (right) the first contour line, at value 2.2, shows when the performance has worsened by a factor of 10% (the minimum value, 2.0, occurs at $a = \frac{1}{2}$ and $d = 0$). This contour line is quite far from the $(\frac{1}{2}, 0)$ point; hence, the HDOP performance is not too sensitive to the optimum values of a and d (this is also visible in the mesh plot of the HDOP).

A SIMPLE STAR SELECTION ALGORITHM

It is unlikely that the sky will present m bright stars forming a perfect m-gon, i.e. equally distributed in azimuth; hence, it is worthwhile developing some sort of algorithmic approach to deciding which set of stars yields a near balanced configuration. A brute force search can find the best set, but suffers from combinational growth in the search space. For example choosing m stars from a collection of M possible allows for $\binom{M}{m}$ possible constellations; for $m = 8$ and $M = 30$, this is over 5.8 million possible situations.

A similar problem occurs in minimizing GDOP for GNSS system and many authors have presented sub-optimum satellite selection procedures. Below we propose and test one simple algorithm based upon trying to satisfy balance described above as nearly as possible. It is based upon the decomposition of m into smaller sets of stars, finding a set of each count by a simple overlap and add approach, and selecting the best combination. We describe the approach by means of an example.

1. We start with the full set of possible azimuths, θ_k, $k = 1, 2, \ldots M$. For 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 39N, 74W, 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, $M = 29$ stars are readily visible. Their azimuths and elevations

(rounded to the nearest whole degree) are:

$$
\text{azimuths} = \left\{
\begin{array}{cccccccccc}
1 & 16 & 25 & 31 & 64 & 76 & 95 & 97 & 114 & 122 \\
122 & 125 & 125 & 135 & 172 & 178 & 205 & 209 & 212 & 219 \\
227 & 260 & 265 & 269 & 309 & 319 & 322 & 332 & 360 &
\end{array}
\right\}
$$

and

$$
\text{elevations} = \left\{
\begin{array}{cccccccccc}
39 & 7 & 19 & 76 & 60 & 27 & 54 & 15 & 36 & 40 \\
10 & 56 & 33 & 31 & 54 & 10 & 3 & 72 & 26 & 53 \\
3 & 36 & 56 & 18 & 26 & 4 & 63 & 9 & 23 &
\end{array}
\right\}
$$

Our goal is to select $m = 8$ stars for positioning; we decide that the decomposition of interest is $8 = 3 + 5$ so that we will initially look for a 3-tuple of stars that are nearly balanced.

2. Finding the set of 3 stars involves the implementation and iteration of several steps:

   (a) Define the length 360 indicator vector $\mathbf{x}$ nominally equal to zeros with ones at the M ($= 29$) azimuth values (map $0°$ to the 360th element, if needed).

   (b) Separate the 360 elements of $\mathbf{x}$ into 3 subvectors of 120 entries each; overlap these subvectors and add them term by term creating the vector $\mathbf{y}$. $\mathbf{y}$ will have 120 entries each of which are integer valued between 0 and 3. Tables of $\mathbf{x}$ and $\mathbf{y}$ after this step appear as the top section in Table 1 labeled "Iteration 1", $\mathbf{x}$ consisting of the three binary rows above the line and $\mathbf{y}$ being the single row below the horizontal line.

   (c) Check the maximum of $\mathbf{y}$. If it equals 3 than there is a set of 3 stars forming a balanced equilateral triangle so we stop this step and go to step 3 below; if not, continue to sub-step (d). (Clearly we do not stop here for the example as the maximum value in $\mathbf{y}$ is equal to 2 which occurs in two places – marked by arrows in Table 1.)

   (d) Widen $\mathbf{x}$ by logically OR-ing it with a circularly-shifted version of itself (one degree to the right). Note that the OR function means that the elements of $\mathbf{x}$ are still equal to 0 or 1. Effectively this creates a widened indicator vector of the available azimuths.

   (e) Subdivide this new version of $\mathbf{x}$ into 3 subvectors and again add term-by-term to produce a new version of $\mathbf{y}$. If this $\mathbf{y}$ contains one (or more) element equal to 3 then a nearly balanced triangle is found and we can continue to step 3. If not, return to sub-step (d) and widen $\mathbf{x}$ further by OR-ing it with a copy of itself circularly shifted one degree to the left. Note that successive implementations of sub-step (d) should alternate expansions to the right and left. For the example, successive applications of sub-step (d) appear in Table 1. Widening $\mathbf{x}$ to groups of 4 per star finally produced a single score of 3 near an offset of $94°$.

3. Once the 3-tuple is discovered remove those 3 stars from contention and repeat step 2 looking for a 5-tuple (i.e. start with the indicator with $M - 3$ stars; separate it into 5 subvectors each of length 72; look for a maximum equal to 5; if found, stop, if not, further widen $\mathbf{x}$ and repeat). For the example, Table 2 shows the initial and a few of the intermediate computations of $\mathbf{x}$ and $\mathbf{y}$; finally after widening $\mathbf{x}$ to 14 ones per star a 5-tuple is identified at an offset of $38°$.

4. The resulting nearly balanced $3 + 5$ constellation is shown in Figure 3. The left subfigure shows the stars as red dots and the 3-star triplet marked by blue lines (the title shows the azimuths of the 3 stars selected); the central subfigure shows the 5-tuple of stars; the right subfigure shows the total collection of 8 stars as the blue dots. A quick computation yields that the resulting HDOP is 0.7087.

While the set of stars found in the example above has a low HDOP we need to look at other possible balanced sets. For example, we should also test $8 = 5 + 3$ (looking for the 5-tuple first); for this example the result is slightly worse then the $3 + 5$ solution (HDOP $= 0.7096$). Finally we look for a $4 + 4$ configuration; the result is shown in Figure 4 and has the slightly better HDOP of 0.7075, remarkably close to the lower bound of 0.7071. (Note that there is no need to check for a fully balanced set of 8 stars. If that had existed, it would have been identified in the $4 + 4$ check with offsets differing by $45°$. Forcing the issue, the "best" near-balanced 8-tuple for this set of stars has HDOP $= 0.7079$ and appears in Figure 5.)

Table 1: The versions of $\mathbf{x}$ (above the line) and $\mathbf{y}$ (below the line) looking for a triplet.

Iteration 1:

```
1000000000000000010000000100000000000010000010100000000001000000
0100100000000010000000000000000000000000000010000001000000000000
0000000000000000010011000000000001000000000000000000000010000001
1100100000000110001000002000101000000000010010100000100000010000001
                              ⇑
```

Iteration 2:

```
1100000000000001100000001100011000000000000000000001110000000000
0110110000000011000000000000000000000000000000011011101100000000
0000000000000001100110000000000000000001101010100110011111000001
1210110000000012100100002200111100000111111100000111000001100001
                              ⇑
```

Iteration 3:

```
1100000000000011100000001110001110000000000000000001111000000000
1111111000000011100000000000000000000001110111111100001110000000
0000000000000001110011101101110000000001111100000000000011100011
2211100000000122101110002222012121100000000001110011100011100011
                              ⇑
```

Iteration 4:

```
1110000000000011110000011110111100000000000000001111100000000000
1111111100000011101111111110000000000000001111111100011111000000
0000000000000011110111101111110000000001111111100011121111122211
2221110000000122221111102222122211000000000001111101111110011110011
                              3
                              ⇑
```

Table 2: Some of the versions of **x** and **y** looking for a 5-tuple.

Iteration 1:
10000000000000001000000001000001000000000000000000000000000000000100000000
00010000000000000000000001000000000000000001000000010010000000001000000000
00000000000000000000000000001000010000000000000000000000000100010000000
00100000001000000000000000000000000000000010000100010000000000000000000
00000000000000000001000000000100100000000000000000000000000000000000001
10110000001000010000100020010020020000000101000011002000000101110000001

Iteration 2:
11000000000000001100000001100001100000000000000000000000000000000110000000
00011000000000000000000110000000000000011000001101100000001100000000
00000000000000000000000000011000011000000000000000000000000110011000000
00110000001100000000000000000000000000000011000110011000000000000000000
00000000000000000000011000000001101100000000000000000000000000000000001
11121000001100011000110022011022022000000011110001210220000001112210000001

                     ⋮

Iteration 13:
11111110011111111111111111111111111110000000000000000000001111111111111111
11111111110000000001111111111111100001111111111111111111111111111111111111000
00000000000000000000001111111111111111111100000000000001111111111111111111
11111111111111111000000000000000000000111111111111111111111110000000000000
00000000000001111111111111111111111111111000000000000000000000000001111111
33333332232222333233344444444433334444222222222222223334433333344444333

Iteration 14:
11111111011111111111111111111111111110000000000000000001111111111111111
11111111110000000011111111111111100011111111111111111111111111111111111100
00000000000000000000011111111111111111111110000000000001111111111111111111
11111111111111111000000000000000000000111111111111111111111110000000000000
00000000000001111111111111111111111111111000000000000000000000000001111111
333333332332223333334444444444443334454442222222222222233344433333344443
                     ⇑

3: 95 212 332　　　　**5: 31 97 172 219 309**　　　　**3+5: 31 95 97 172 212 219 309 332**

Figure 3: Star selections for $m = 8 = 3 + 5$; the HDOP equals 0.7086.

4: 1 95 178 269　　　　**4: 31 125 219 309**　　　　**4+4: 1 31 95 125 178 219 269 309**

Figure 4: Star selections for $m = 8 = 4 + 4$; the HDOP equals 0.7075.

8: 31 76 125 172 219 265 309 360　　　　**8: 31 76 125 172 219 265 309 360**

Figure 5: The set of $m = 8$ star best fitting Bowditch's sense of spread; the HDOP equals 0.7079.

Figure 6: Histogram of the full search result.

Figure 7: The best and worst choices for stars, respectively.

COMPARISON TO FULL SEARCH

How do these results compare to the full search optimum? With $M = 29$ and $m = 8$ the problem is not too large, a total of 4.3 million subsets. Figure 6 shows the resulting histogram of the HDOP over all of the choices: the best case HDOP of 0.7071 through the worst case value of 2.802 (the constellations yielding both are shown in Figure 7). This histogram is quite skewed toward the left; there are very many 8-star constellation choices 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! While our simple algorithm found a constellation with quite good performance (the $4 + 4$ being the best), there are over 42,000 with better performance!

Examining Figure 7 we find it interesting that the azimuths of the best choice is not as spread in bearing as might be surmised from Bowditch's recommendations; nor does it appear to be the union of simple polygons. On the other hand, the worst is as we would expect; effectively collinear celestial objects.

CONCLUSIONS/FUTURE WORK

This paper makes several points:

- We make the observation that the classical altitude-intercept method of celestial navigation is really just the first iteration in the typical linearized least squares GNSS position solution.

- We develop a lower bound to HDOP for celestial navigation and define the balance conditions necessary for the selected star pattern to achieve it.

- We explore the characteristics of constellations that achieve balance, finding that there are many ways to do

so; hence, there are many actual constellations close to balance with near optimum performance so a simple selection algorithm should work pretty well.

- We develop an extremely simple, real-time algorithm for selecting stars to yield extremely good HDOP values, but with resulting constellations quite different than those envisioned by readers of Bowditch.

Our future work includes:

- Developing a greater understanding of balance, and near balance, and those impacts on HDOP.

- Characterizing all constellations that achieve balance for different values of m. The figure showing the best constellation suggests that limited view of the sky can still result in excellent HDOP. Identifying how to exploit this will aide celestial navigation in limited sky environments.

- Refining these results by including the varying measurement accuracy and the bias term as mentioned above in the review of celestial navigation. The varying accuracy will result in modified balance conditions, effectively turning them into weighted balance. The bias term relates to the ability to estimate the user's altitude and will result in additional east and north balance conditions for optimum performance.

REFERENCES

[1] P. Misra and P. Enge, *Global Positioning System: Signals, Measurement, and Performance*, Ganga-Jamuna Press, 2001.

[2] R. Yarlagadda, I. Ali, N. Al-Dhahir, and J. Hershey, "GPS GDOP metric," *IEE Proc. Radar, Sonar, Navigation*, Oct. 2000.

[3] P. F. Swaszek, R. J. Hartnett, and K. C. Seals, "Lower bounds to DOP," *Jour. Navigation*, June 2017.

[4] D. Gerbeth, M. Felux, M-S. Circiu, and M. Caamano, "Optimized selection of satellites subsets for a multi-constellation GBAS," *Proc. ION Int'l. Tech. Mtg.*, Jan. 2016.

[5] M. Kihara and T. Okada, "A satellite selection method and accuracy for the Global Positioning System," *Jour. Navigation*, vol. 31, no. 1, Spring 1984, pp. 8-20.

[6] M. S. Phatak, "Recursive method for optimum GPS satellite selection," *IEEE Trans. Aero. Elect. Sys.*, vol. 37, no. 2, Apr. 2001, pp. 751-754.

[7] C-W. Park and J. P. How, "Quasi-optimal satellite selection algorithm for real-time applications," *Proc. ION GPS 2001*, Salt Lake City UT, Sept. 2001.

[8] M. Liu, M-A. Fortin, and R. Landry, "A recursive quasi-optimal fast satellite selection method for GNSS receivers," *Proc. ION GNSS 2009*, Savannah GA, Sept. 2009.

[9] M. Zhang and J. Zhang, "A fast satellite selection algorithm: beyond four satellites," *IEEE Jour. Selected Topics in Signal Processing*, vol. 3, no. 5, Oct. 2009, pp. 740-747.

[10] N. Blanco-Delgado and F. D. Nunes, "Satellite selection method for multi-constellation GNSS using convex geometry," *IEEE Trans. Veh. Tech.*, vol. 59. no. 9, Nov. 2010, pp. 4289-4297.

[11] M. Wei, J. Wang, and J. Li, "A new satellite selection algorithm for real-time application," *Proc. IEEE Int'l. Cong. Sys. & Informatics*, 2012, pp. 2567-2570.

[12] M. Ranjbar and M. R. Mosavi, "Simulated annealing clustering for optimum GPS satellite selection," *Int'l. Jour. Comp. Sci.*, May 2012.

[13] A. Peng, G. Ou, and G. Li, "Fast satellite selection method for multi-constellation Global Navigation Satellite System under obstacle environments," *IET Radar Sonar Navig.*, 2014.

[14] P. F. Swaszek, R. J. Hartnett, K. C. Seals, and R. M. Swaszek, "A temporal algorithm for satellite subset selection in multi-constellation GNSS," *Proc. ION ITM*, Monterrey CA, Jan. 2017.

[15] N. Bowditch, *The American Practical Navigator*, Bicentennial Edition, National Imagery and Mapping Agency, Bethesda, MD, 2002, p. 264.

[16] T. J. Cutler, *Dutton's Nautical Navigation*, 15th Ed., Naval Institute Press, 2003.

[17] J. A. Van Allen, "Basic principles of celestial navigation," *Amer. Jour. Physics*, Nov. 2004.

[18] J. Karl, *Celestial Navigation in the GPS Age*, Paradise Cay Pub., 2007.

[19] R. W. Severance, "Overdetermined celestial fix by iteration," *Navigation*, 1989.

[20] C. DeWit, "Optimal estimation of a multi-star fix," *Navigation*, 1974.

[21] *The Nautical Almanac*, U.S. Naval Observatory and the United Kingdom Hydrographic Office, updated annually.

[22] G. D. Morrison, "Most probable fix position reduction," *Navigation*, 1981.

APPENDIX – PROOFS OF THEOREMS

Theorem 2: rotation and negation

Proof. We start by checking both conditions for the rotated constellation; the method is primarily trigonometric manipulations. Using hats to indicate the direction components of the rotated constellation, the first condition is

$$
\begin{aligned}
\sum_{k=1}^{m} \widehat{e_k}\, \widehat{n_k} &= \sum_{k=1}^{m} \sin(\theta_k + \phi)\cos(\theta_k + \phi) \\
&= \sum_{k=1}^{m} (\sin\theta_k \cos\phi + \cos\theta_k \sin\phi)(\cos\theta_k \cos\phi - \sin\theta_k \sin\phi) \\
&= (\cos^2\phi - \sin^2\phi)\sum_{k=1}^{m}\sin\theta_k\cos\theta_k - \sin\phi\cos\phi\left(\sum_{k=1}^{m}\sin^2\theta_k - \sum_{k=1}^{m}\cos^2\theta_k\right) \\
&= (\cos^2\phi - \sin^2\phi)\left[\sum_{k=1}^{m}e_k n_k\right] - \sin\phi\cos\phi\left[\sum_{k=1}^{m}(e_k^2 - n_k^2)\right]
\end{aligned}
$$

The terms in square brackets are the balance conditions for the original constellation, both equal to zero by assumption, so this condition is also zero for the rotated constellation.

For the rotated constellation the second balance condition is

$$
\begin{aligned}
\sum_{k=1}^{m}\left(\widehat{e_k}^2 - \widehat{n_k}^2\right) &= \sum_{k=1}^{m}\left(\sin^2(\theta_k + \phi) - \cos^2(\theta_k + \phi)\right) \\
&= \frac{1}{2}\left(\sum_{k=1}^{m}(1 - \cos 2(\theta_k + \phi) - 1 - \cos 2(\theta_k + \phi))\right) \\
&= -\sum_{k=1}^{m}\cos 2(\theta_k + \phi) \\
&= -\sum_{k=1}^{m}(\cos 2\theta_k \cos 2\phi - \sin 2\theta_k \sin 2\phi) \\
&= -\cos 2\phi\sum_{k=1}^{m}\cos 2\theta_k + \sin 2\phi\sum_{k=1}^{m}\sin 2\theta_k \\
&= -\cos 2\phi\sum_{k=1}^{m}\left(\cos^2\theta_k - \sin^2\theta_k\right) + \sin 2\phi\sum_{k=1}^{m}2\sin\theta_k\cos\theta_k \\
&= \cos 2\phi\left[\sum_{k=1}^{m}(e_k^2 - n_k^2)\right] + 2\sin 2\phi\left[\sum_{k=1}^{m}e_k n_k\right]
\end{aligned}
$$

Again the bracketed terms are zero due to the balance of the original constellation; hence, the rotated constellation is also balanced.

For the reflection operation note that all of the sine terms in the conditions are negated, $\sin(-\theta_k) = -\sin\theta$, while the cosine terms are unchanged, $\cos(-\theta_k) = \cos\theta$. Being linear in the sine, the first balance calculation still sums to zero. The second balance condition only depends upon the square of the sine terms; hence, is also unchanged. The result is that the reflected constellation is also balanced $\qquad\square$

Theorem 3: $m \geq 3$ result

Proof. From Euler's identity

$$e^{j\theta_k} = \cos\theta_k + j\sin\theta_k$$

we have

$$e^{j2\theta_k} = \left(e^{j\theta_k}\right)^2 = (\cos\theta_k + j\sin\theta_k)^2 = \cos^2\theta_k - \sin^2\theta_k + j2\sin\theta_k\cos\theta_k$$

Summing over k

$$\sum_{k=1}^{m} e^{j2\theta_k} = \underbrace{\sum_{k=1}^{m}\cos^2\theta_k - \sum_{k=1}^{m}\sin^2\theta_k}_{\text{real part}} + j\,2\underbrace{\sum_{k=1}^{m}\sin\theta_k\cos\theta_k}_{\text{imaginary part}}$$

in which the real and imaginary parts correspond to the two necessary conditions for balance; in other words, a balanced constellation has

$$\sum_{k=1}^{m} e^{j2\theta_k} = 0$$

In units of radians the equal angle constellation has

$$\theta_k = (k-1)\frac{2\pi}{m}$$

so this condition is

$$\sum_{k=1}^{m} e^{j2\theta_k} = \sum_{k=1}^{m} e^{j2(k-1)\frac{2\pi}{m}} = \sum_{k=1}^{m}\left(e^{j\frac{4\pi}{m}}\right)^{(k-1)} = \sum_{k=0}^{m-1}\alpha^k$$

with

$$\alpha = e^{j\frac{4\pi}{m}}$$

Simplifying the finite power series

$$\sum_{k=0}^{m-1}\alpha^k = \frac{1-\alpha^m}{1-\alpha}$$

But since

$$1-\alpha^m = 1 - e^{j\frac{4\pi}{m}m} = 1 - e^{j4\pi} = 1 - 1 = 0 \qquad \text{and} \qquad \alpha = e^{j\frac{4\pi}{m}} \neq 1$$

the summation is identically zero and the necessary conditions are met for all $m \geq 3$.

$\qquad\square$

Theorem 4: unions

Proof. Let m_1 and m_2 represent the number of stars in each of the two constellations, respectively. Index the stars in the union by $1, 2, \ldots m_1$ for those from the first constellation and $m_1 + 1, m_1 + 2, \ldots m_1 + m_2$ for those from the second. The first balance condition is

$$\sum_{k=1}^{m_1+m_2} e_k n_k = \sum_{k=1}^{m_1} e_k n_k + \sum_{k=m_1+1}^{m_1+m_2} e_k n_k$$

Since each individual constellation is balanced, each of these summations is equal to zero and the overall condition is met. In a similar fashion the second balance condition is satisfied. $\qquad\square$