LS GEO IOD

James R Wright*

Abstract

I have constructed and demonstrated a new fast-running algorithm to perform refined orbit determination for any spacecraft in GEO (geosynchronous earth orbit), without *user* requirement for an a priori orbit estimate. We now have a *refined* method for GEO initial orbit determination (IOD). This is enabled by the use of equinoctial orbit elements, by a one-dimensional search in the equinoctial orbit element *mean argument of orbit longitude*, by the use of sensor locations to reduce the one-dimensional search, and by convergence boundaries of the nonlinear least squares (LS) algorithm. I shall refer herein to the new algorithm as LS_GEO_IOD. Algorithm LS_GEO_IOD employs any desired perturbative acceleration model for numerical trajectory integration, and it's success appears to be independent of input measurement type.

BACKGROUND

Initial orbit determination, also called preliminary orbit determination (Herrick[10]), refers to a class of orbit determination methods that derive initial orbit estimates from sensor measurements and sensor locations, without *user* requirement for an a priori orbit estimate. But the use of two-body orbit mechanics, and the existence of significant white noise[1] on minimal sets of measurements in geocentric applications, has always created IOD estimates with very large estimation error magnitudes. Existing IOD algorithms are unlike each other, distinguished by distinct measurement types and by distinct methods to address nonlinearity. They are disparate. And some IOD algorithms produce multiple distinct solutions (e.g., see Gooding[8]). Historically, the two-body IOD algorithms are associated with the names of Laplace (1780), Lagrange(1778,1783), Gauss(1809), Gibbs(1889), Herrick(1940), Gooding(1993), and others (see Herrick[10] and Gooding[8]).

Two-body IOD estimates have been used to seed iterative batch least squares (LS) algorithms so as to calculate refined LS orbit estimates. LS algorithms use complete acceleration models and overdetermined sets of measurements to accomplish the *refinement* and to provide the *unique* orbit estimate. Existing nonlinear LS algorithms, also known as Gauss-Newton algorithms, have always required an a priori orbit estimate for initialization.

ALGORITHM SUMMARY

Algorithm LS_GEO_IOD is unified for distinct measurement types with a nonlinear least squares algorithm, with a one-dimensional search in *mean argument of longitude* (an element in the set of six equinoctial orbit elements[2]), a user interface in Kepler orbit elements, LS differential corrections in ECI position and velocity components, and with LS iterative convergence defined by Cauchy. Six-dimensional nonlinear transformations between equinoctial orbit elements, Kepler orbit elements, and ECI position and velocity components are employed. When the GEO is observable from an

*ODTK Architect, Analytical Graphics, Inc., 220 Valley Creek Blvd, Exton, PA, 19341

[1] Laplace's method for angles measurements fails for most applications to geocentric orbits due to white noise embedded in the angles measurements combined with the second-order Taylor's series approximation that requires measurement time-tags to be close together and from the same sensor.

[2] A short history of development of the equinoctial orbit elements is presented in Appendix A.

overdetermined measurement set, LS_GEO_IOD calculates a unique and refined orbit estimate without *user* requirement for an a priori GEO estimate.

WHY LS_GEO_IOD SUCCEEDS

Denote Greenwich sidereal time with $\theta = \theta_G(t)$, and denote spacecraft earth-fixed longitude with $\lambda = \lambda(t)$. Right ascension $\alpha = \alpha(t)$ of the spacecraft has the representation

$$\alpha = \theta + \lambda \tag{1}$$

so that

$$\lambda = \alpha - \theta \tag{2}$$

Now $\theta = \theta_G(t)$ is a function only of the date and time t. So given the spacecraft right ascension $\alpha = \alpha(t)$, one can know the spacecraft earth-fixed longitude $\lambda = \lambda(t)$ according to Equation 2. Denote the spacecraft position vector with $\mathbf{r}$, and the position unit vector with $\mathbf{U} = \mathbf{r}/r$, where $r = \sqrt{\mathbf{r} \cdot \mathbf{r}}$, to visualize the relations between $\mathbf{r}$, θ, λ, and α in Figure 1.

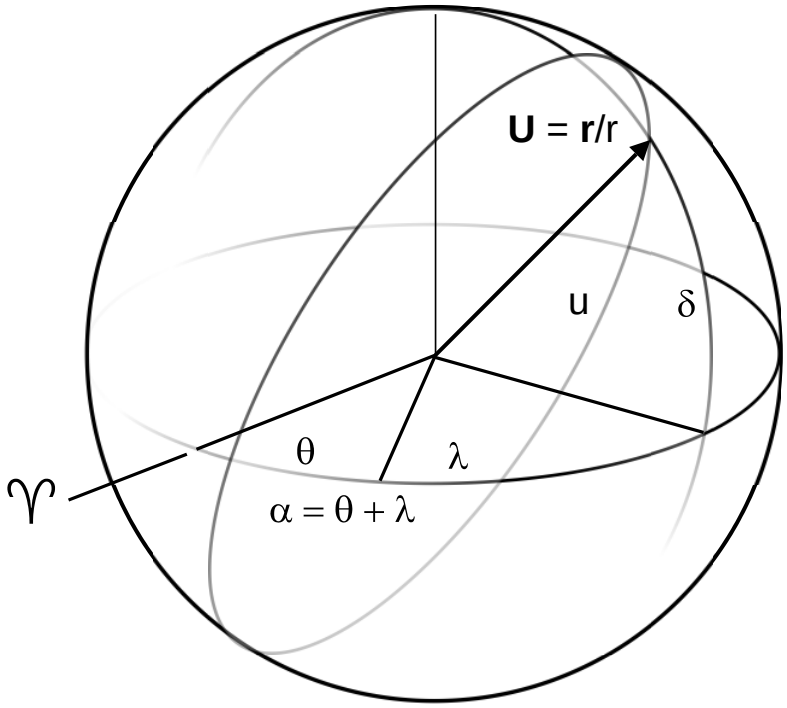

Figure 1: Orbit Unit Sphere for Lambda

Figure 2 relates the true argument of latitude u and inclination i to α and λ geometrically. Let Ω denote spacecraft right ascension of ascending node (the node), and define β with

$$\beta = \alpha - \Omega \tag{3}$$

The right spherical triangle defined by angle i, and sides β and u, provides this relation.

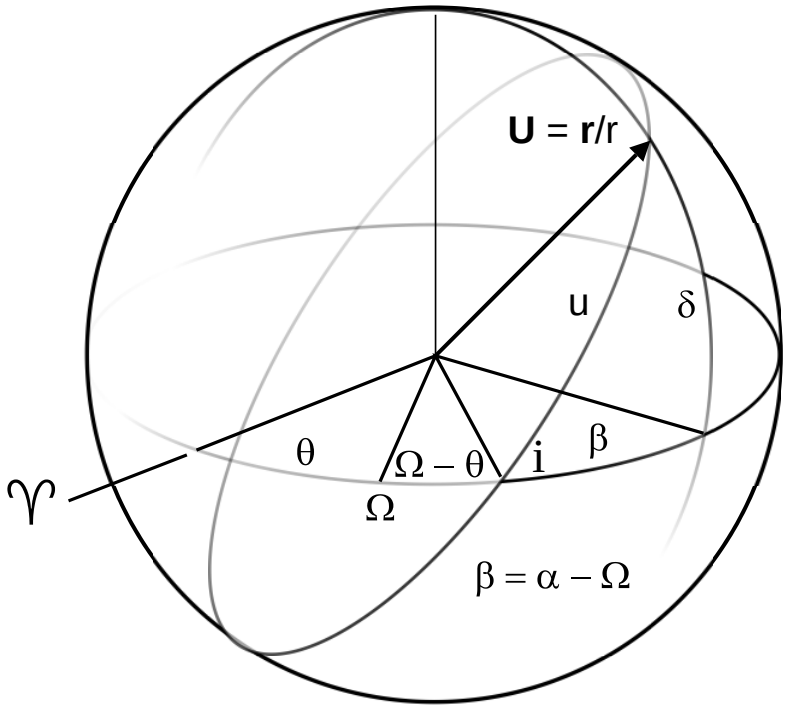

Figure 2: Orbit Unit Sphere for Beta

Spacecraft Longitude

Equation 2 defines the spacecraft earth-fixed longitude λ. Now let l denote the spacecraft true argument of orbit longitude defined by

$$l = u + \Omega \tag{4}$$

(see Herrick[10] page 63) in terms of true argument of orbit latitude u in the orbit plane and node Ω in the equatorial plane. The true argument of longitude l has the vernal equinox as inertial origin.

Equatorial GEO

Notice when inclination $i \to 0$ that u and Ω share the same orbit-equatorial plane (Figure 2). Thus

$$(i = 0) \Longrightarrow (\alpha = l) \tag{5}$$

Insert Equations 1 and 4 into Equation 5 to see that

$$\theta_G + \lambda = u + \Omega$$

or

$$\lambda = u + \Omega - \theta_G = l - \theta_G \tag{6}$$

If $i = 0$, then λ is indeed an Earth-fixed spacecraft longitude.

Ideal Fictitious GEO

By ideal fictitious GEO I assume the value for semi-major axis a to be perfectly known, and that $i = 0$ and $e = 0$. It is useful to examine the Kepler orbit elements for such a GEO

$$\begin{bmatrix} \text{semi-major axis} \\ \text{eccentricity} \\ \text{mean anomaly} \\ \text{inclination} \\ \text{node} \\ \text{argument of perigee} \end{bmatrix} = \begin{bmatrix} a \\ e \\ M \\ i \\ \Omega \\ \omega \end{bmatrix} = \begin{bmatrix} \text{known} \\ 0 \\ \text{undefined} \\ 0 \\ \text{undefined} \\ \text{undefined} \end{bmatrix} \tag{7}$$

Equation 7 presents the resulting undefined GEO: Kepler elements M, Ω, and ω are undefined. Define the equinoctial orbit element mean longitude L with

$$L = \Omega + \omega + M \tag{8}$$

The mean longitude L is a scalar angular sum defined in two planes, and has the vernal equinox as inertial origin. The node Ω is defined in the earth's equator, and the sum $\omega + M$ is defined in the orbit plane. The six equinoctial orbit elements, presented and defined in Appendix A with Equation 33, were first defined by Arsenault and Koskela[13] (1967). Write "$L = l$ if $e = 0$" to mean $\lim_{e\to 0} L = \lim_{e\to 0} l$, where e is the osculating eccentricity. Then Equation 6 can be written[3]

$$\lambda = M + \omega + \Omega - \theta_G = L - \theta_G \tag{9}$$

If one knew a priori that an orbit in question is an "ideal fictitious" GEO with $i = 0$, $e = 0$, and $a = 6.610571137$ er, then use of Equation 33 provides

$$\begin{bmatrix} a_f \\ a_g \\ n \\ L \\ \chi \\ \psi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \text{known} \\ \text{unknown} \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

where the orbit is well-defined and the only unknown is L. Compare Equations 10 and 7 to appreciate the advantage to using equinoctial orbit elements for GEO. Given an appropriate simulated measurement set for which a GEO is observable, perform a one-dimensional search on L, using the minimum measurement residual RMS to identify the orbit.

Operational GEO

For an operational GEO to be geosynchronous, one must have $a = 6.610571137$ er approximately. And when $i \approx 0$ for $0 \le i \le \pi$, and $e \approx 0$ for $0 \le e < 1$, then the LS (Gauss-Newton) convergence boundaries are sufficiently large to capture all operational GEO with a one-dimensional search on L. See the TDOA EMPIRICAL RESULTS below**.**

REDUCE THE ONE-DIMENSIONAL SEARCH

Denote an earth-centered sensor position vector with $\mathbf{s}$, and define the instantaneous range vector $\boldsymbol{\rho}$

$$\boldsymbol{\rho} = \mathbf{r} - \mathbf{s} \tag{11}$$

Define the instantaneous one-way range scalar ρ with

$$\rho = \sqrt{\boldsymbol{\rho} \cdot \boldsymbol{\rho}} \tag{12}$$

Vectors $\mathbf{r}$ and $\mathbf{s}$ define a plane in $\mathbb{R}^3$; e.g., the plane of Figure 3. Extrema for visability limits of point $\mathbf{r}$ from point $\mathbf{s}$, for a spherical earth model, are found from

[3] See Kaula[12] page 51 Equation 3.125 for his definition of λ_A.

$$\mathbf{s} \cdot \boldsymbol{\rho}_T = 0 \tag{13}$$

where range vector $\boldsymbol{\rho}_T$ is tangent to the earth sphere, and $\mathbf{s}$ is held fixed. Let $\mathbf{r} = \mathbf{r}_T$ denote a spacecraft position vector for which $\boldsymbol{\rho} = \boldsymbol{\rho}_T$; that is

$$\boldsymbol{\rho}_T = \mathbf{r}_T - \mathbf{s} \tag{14}$$

Then

$$\mathbf{s} \cdot \boldsymbol{\rho}_T = \mathbf{s} \cdot (\mathbf{r}_T - \mathbf{s}) = 0$$

and

$$\mathbf{r}_T \cdot \mathbf{s} = \mathbf{s} \cdot \mathbf{s}$$

or

$$\hat{\mathbf{r}}_T \cdot \hat{\mathbf{s}} = s/r_T \tag{15}$$

for unit vectors $\hat{\mathbf{r}}_T$ and $\hat{\mathbf{s}}$ and lengths r_T and s of vectors $\mathbf{r}_T$ and $\mathbf{s}$. Define θ_{rs} as the angle between $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$

$$\theta_{rs} = (\hat{\mathbf{r}}_T, \hat{\mathbf{s}}) \tag{16}$$

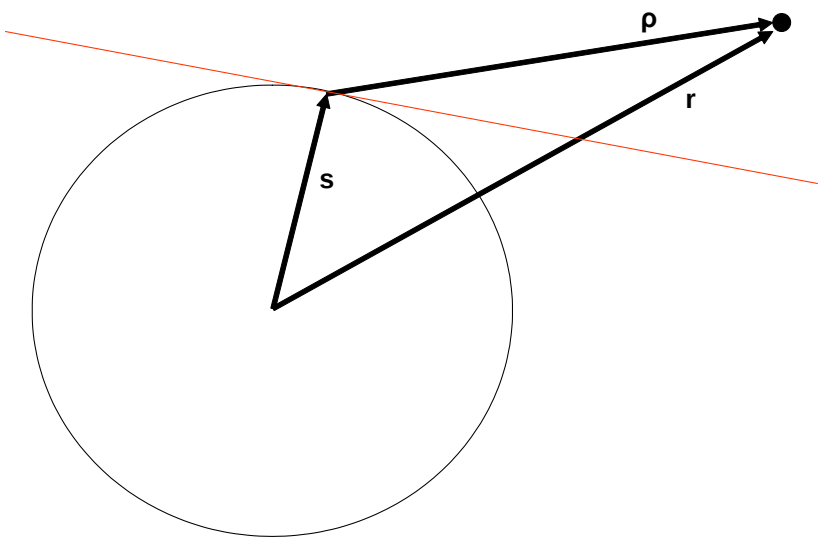

Figure 3: Range Vector Triangle

Then

$$\cos\theta_{rs} = \cos(\hat{\mathbf{r}}_T, \hat{\mathbf{s}}) = \frac{s}{r_T} \tag{17}$$

Let φ_s and λ_s denote the known geocentric earth-fixed ground station sensor geocentric latitude and East longitude, and let φ denote declination of the spacecraft. Particularize the oblique spherical triangle of Figure 4 to unit vectors of spacecraft and sensor positions. Invoke the identity (e.g., Todhunter[19] page 18)

$$\cos a = \cos b \cos c + \sin b \sin c \cos A \tag{18}$$

Assign

$$A = \Delta\lambda \tag{19}$$

$$a = \theta_{rs} \tag{20}$$

$$b = \pi/2 - \varphi_s \tag{21}$$

$$c = \pi/2 - \varphi \tag{22}$$

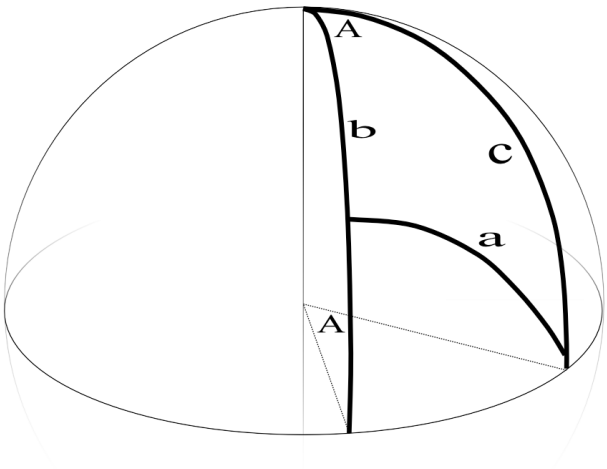

Figure 4: Oblique Spherical Triangle

Then

$$\cos(\Delta\lambda) = \frac{\cos\theta_{rs} - \sin\varphi_s \sin\varphi}{\cos\varphi_s \cos\varphi} \tag{23}$$

We know the station geocentric latitude φ_s and we have a useful approximation for the maximum value of θ_{rs}. We do not have an exact value for the spacecraft declination φ, but we are given that the orbit is a GEO. Then adopt the useful approximation

$$i \approx 0.0 \tag{24}$$

to write

$$\varphi \approx 0 \tag{25}$$

Equation 23 becomes

$$\cos(\Delta\lambda) \approx \frac{\cos\theta_{rs}}{\cos\varphi_s} \tag{26}$$

Define angle $\zeta(\varphi_s)$

$$\zeta(\varphi_s) = \max|\lambda_s - \lambda| \tag{27}$$

Insert Equation 17 into Equation 26 to get

$$\cos(\max|\lambda_s - \lambda|) \approx \frac{s}{r_T \cos\varphi_s} \tag{28}$$

where all GEO radii r (inclusive of r_T) are in the neighborhood of 6.610571137 er. Then

$$\zeta(\varphi_s) \approx \cos^{-1}\left[\frac{s}{r_T \cos\varphi_s}\right] \tag{29}$$

Angle $|\lambda_s - \lambda|$ is maximal due to Equation 13. Then approximately

$$\lambda_s - \zeta(\varphi_s) \ < \lambda < \lambda_s + \zeta(\varphi_s) \tag{30}$$

Search Bounds for Mean Longitude L

Insert Equation 9 into Equation 30 and rearrange to get

$$\theta_G(t) + [\lambda_s - \zeta(\varphi_s)] \ < L(t) < \theta_G(t) + [\lambda_s + \zeta(\varphi_s)] \tag{31}$$

Values for the station latitude φ_s and longitude λ_s are given, and $\theta_G(t)$ can be evaluated from t with an existing software routine (or looked up in an almanac table). The search for GEO mean longitude $L(t) = M(t) + \omega + \Omega = L(t_0) + n[t - t_0]$ at any time t is bounded by Inequality 31. For measurements from multiple ground stations the search for GEO mean longitude $L(t)$ is further reduced by averaging the centers of station intervals on L to obtain an initial candidate for LS.

LS CONVERGENCE CRITERION

The operator selects a threshold N_{MAX} for maximum number of least squares iterations without convergence, with default $N_{MAX} = 15$. Iterative least squares RMS *convergence* is defined for scalar measurements after iteration k if the root-mean-square $RMS^{(k)}$ on scalar measurement residuals $\Delta y_j^{(k)}$, $j \in \{1, 2, \ldots\}$, is reduced to less than a small predefined positive constant ϵ_{RMS}

$$RMS^{(k)} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left[\Delta y_j^2\right]^{(k)}} < \epsilon_{RMS} \tag{32}$$

Otherwise RMS divergence is defined. If multiple measurements are to be processed at each measurement time-tag, then the scalar $RMS^{(k)}$ is replaced with the scalar *weighted* root-mean-square $WRMS^{(k)}$, where each squared measurement residual Δy_j^2 of type i is replaced by $W^i \Delta y_j^2$ in Expression 53, such that $\sum_i W^i = 1$.

Cauchy Convergence

A necessary and sufficient condition[4] for the sequence $\left\{RMS^{(k)}\right\}$, for $k \in \{1, 2, 3, \ldots\}$, to have a limit is that the absolute difference $\left|RMS^{(k)} - RMS^{(j)}\right|$ approach 0 as j and $k \to \infty$, where $j \in \{k-1, k-2, \ldots\}$. That is, for each positive ϵ_{RMS} there is some positive integer N such that $\left|RMS^{(k)} - RMS^{(j)}\right| < \epsilon_{RMS}$ if $N \leq k$ and $N \leq j$. We can approximate use of the rigorous Cauchy convergence criterion using a sequence $RMS^{(k)}$ on the measurement residual RMS.

[4] Cauchy convergence is presented in many books on calculus; e.g., see Taylor[18] page 453.

TDOA EMPIRICAL RESULTS

By time-of-arrival (TOA) measurement I refer to measurement of one-way propagation time of a radio wave-front from the spacecraft transmitter to sensor receiver. By time-difference-of-arrival (TDOA) I refer to the difference between two TOA measurements made at two distinct sensors from a common radio wave-front. Numerically integrated spacecraft accelerations were simulated[5] from a (degree,order) = (6,6) geopotential and a spherical spacecraft body for solar photon pressure to create simulated position and velocity ephemerides.

Sensor Name	**Height**	**Latitude**	**E Longitude**
COOK	271.51 m	34.82260940 deg	239.4981480 deg
HULA	428.42 m	21.56228000 deg	201.7578910 deg
GUAM	217.00 m	13.61519420 deg	144.8560742 deg

Table 1: Ground Station Locations

Three AFSCN ground station sensors COOK, HULA, and GUAM were used to simulate TDOA measurements. With three ground stations, and two ground stations per TDOA measurement, we have three sets of TDOA measurements: COOK-HULA, COOK-GUAM, and HULA-GUAM. A time granularity of 10 minutes was selected for each set of TDOA measurements. The least squares fit span selected was 1440 minutes. So each of the three sets has 145 scalar simulated TDOA measurements. The three ground station locations are presented in Table 1.

Simulated truth values for Kepler elements e, i, and Ω were selected to satisfy optimal GEO control according to George Chao[7]. Without stationkeeping maneuvers, the inclination i of a GEO will vary between 0 degrees and 15 degrees in about 26 years due to luni-solar gravitational perturbations. Thus our required search interval bounds for unknown inclination i is: 0 deg $\leq i \leq$ 15 deg. GEO station-keeping maneuvers to maintain *stationary inclination* is achieved with $i \approx 7.3$ degrees and $\Omega \approx 0.0$ degrees. Physical long-term variations in GEO eccentricity are driven by solar photon radiation pressure. GEO eccentricity variations can be minimized with control maneuvers that keep perigee pointed toward the Sun and keep the eccentricity value near to the so-called forced eccentricity e_f. For a typical GEO spacecraft (Chao page 109) $e_f \approx 0.0005$.

Single Element Variation Bounds

Kepler Element	**Symbol**	**Sim Truth**	**Extreme Estimates**	**LS Conv Bounds**
semi-major axis	a	42163 km	[46163 km, 39163 km]	[−4000 km, +3000 km]
eccentricity	e	0.0005	[0.1505, 0.0000]	[−0.1500, +0.0005]
true arg latitude	u	295.4 deg	N/A	not bounded
inclination	i	7.3 deg	[32.3 deg, 0.0 deg]	[−25 deg, +7.3 deg]
node	Ω	0 deg	N/A	not bounded
argument perigee	ω	0 deg	N/A	not bounded

Table 2: Single Element LS Bounds for GEO

Table 2 presents approximate least squares convergence bounds (**LS Conv Bounds**) in Kepler orbit elements for simulated TDOA measurements[6]. The LS convergence bounds for a, e, and i of Table 2 were derived empirically, one at a time; i.e., by input variation of one orbit element with

[5] ODTK (AGI) was extended to derive these results.

[6] Similar results were obtained for two-way range measurements.

no variation in the other five orbit elements. Each of the two values in the 5^{th} column (**LS Conv Bounds**) was derived by subtracting each of the two values in the 4^{th} column (**Extreme Estimates**) from the value in the 3^{th} column (**Sim Truth**). The "not bounded" entry for u in column 5 appears because the one-dimensional search in mean argument of longitude L is also a one-dimensional search in true argument of latitude u.

Successful Multiple Element Variations

Kepler Element	**Symbol**	**Sim Truth**	**Initial Estimate**	**Initial LS Error**
semi-major axis	a	42163 km	[47163 km, 38163 km]	[−5000 km, +4000 km]
eccentricity	e	0.0005	0.1005	−0.10
true arg latitude	u	295.4 deg	115.4 deg	+180 deg
inclination	i	7.3 deg	27.3 deg	−20 deg
node	Ω	0 deg	180 deg	−180 deg
argument perigee	ω	0 deg	180 deg	−180 deg

Table 3: Multiple Element LS Bounds for GEO

LS convergence was demonstrated for many multiple element cases. Two multiple element cases are presented in Table 3, with extreme initial errors in a of −5000 km and +4000 km. Extreme initial condition errors were tested in the six Kepler orbit elements a, e, u, i, Ω, ω simultaneously. GEO LS convergence boundaries are coupled in a, e, u, and i.

Absence of Applicable Theory

The TDOA results presented were obtained by simulation experiments, without a rigorous theory to quantify least squares convergence boundaries as a function of initial condition orbit errors. The successful single element convergence bound magnitudes of Table 2, for the three ground station sensors used, are surprisingly large. And the successful multiple element cases of Table 3 have initial least squares error magnitudes that are surprisingly large. I have run many other multiple element cases and have found none that would seem to prohibit general application of algorithm LS_GEO_IOD to GEO.

It #	**RMS (ns)**
1	4.94×10^6
2	2.23×10^6
3	2.50×10^6
4	2.69×10^5
5	8.17×10^5
6	1.03×10^6
7	2.16×10^4
8	4.76×10^3
9	1.14
10	0.78

Table 4: LS Convergence Pattern

These simulation experiments have frequently demonstrated during the search on mean orbit longitude L that the iterative root-mean-square (RMS) on measurement residuals does not converge

monotonically when it does converge. That is, the RMS frequently increases before decreasing and achieving convergence. Initially the RMS values may be very large. Table 4 presents the TDOA measurement residual RMS convergence sequence associated with the LS bounds of Table 3. Simulated white noise variates of the TDOA measurements were derived from root variance $\sigma_{WN} = 1$ ns (nano-second) and mean zero, and the convergence threshold ϵ_{WN} was set at $\epsilon_{WN} = 1$ ns.

LS divergence, defined by an RMS sequence that increases without bound, is quickly identified when the LS GEO estimate goes hyperbolic. I have been unable to find a rigorous Gauss-Newton convergence theory for use in explanation of the LS orbit convergence patterns experienced with these simulations.

Transitive Partial Derivatives

The reader is referred to Appendix B for a discussion on the critical requirement for appropriate transitive partial derivatives to obtain maximal LS convergence boundaries.

ACKNOWLEDGMENTS

I am appreciative of William Chuba's work in extending ODTK to include the LS_GEO_IOD algorithm. This automatically enables the use of a complete force model, several measurement types that are in current operational use, the use of simulated measurement data for analysis, and the use of real measurement data for operations.

I thank Jim Woodburn for emphasizing the importance of detailed *variational equations* for evaluation of LS transitive partial derivatives. For GEO, I had used a (deg,ord) = (6,6) geopotential for trajectory integration and default values for geopotential variational equations with (deg,ord) = (2,0). The use of (deg,ord) = (6,6) *variational equations*, third-body Sun-Moon *variational equations*, and solar photon pressure partial derivatives produced an enlarged LS convergence boundary for multiple orbit element initial condition errors.

References

[1] Arsenault, J. L., Ford, K. C., Koskela, P. E., *Orbit Determination Using Analytic Partial Derivatives of Perturbed Motion*, AIAA Journal, 8, 4-12, 1970

[2] Arsenault, Jeannine, telephone conversation, 1 June 2009

[3] Arsenault, Jeannine, e-mail correspondence, 2 June 2009

[4] Battin, Richard H., *An Introduction to The Mathematics and Methods of Astrodynamics*, AIAA, 1987

[5] Broucke, R. A., Cefola, P. J., *On the Equinoctial Orbit Elements*, Celestial Mechanics, 5, pp. 303-310, 1972

[6] Brouwer, Dirk, Clemence, Gerald M., *Methods of Celestial Mechanics*, Academic Press, New York, 1961

[7] Chao, Chia-Chun "George", *Applied Perturbation and Maintenance*, The Aerospace Press, El Segundo, 2005

[8] R. H. Gooding, *A New Proceedure for Orbit Determination Based on Three Lines of Sight (Angles Only)*, Technical Report 93004, Defence Research Agency, Farnborough, Hampshire, April 1993

[9] Gratton, S., Lawless, A. S., Nichols, N. K., *Approximate Gauss-Newton Methods for Nonlinear Least Squares Problems*, SIAM J. Optim., Vol. 18, No. 1, pp 106-132, 2007

[10] Herrick, Samuel, *Astrodynamics*, Vol 1, Van Nostrand, London, 1971

[11] Herrick, Samuel, *Astrodynamics*, Vol 2, Van Nostrand, London, 1972

[12] Kaula, William M., *Theory of Satellite Geodesy*, Blaisdell Pub., Mass., 1966, General Publishing Company, Ltd., Toronto, (Gene Kaula), 2000

[13] Paul E. Koskela (L. G. Walters, C. G. Hilton, J. L. Arsenault, R. G. Schinnerer, Chela Varrentzoff, K. C. Ford), *Astrodynamic Analysis for the Advanced Orbit/Ephemeris Subsystem*, Aeronutronic Pub. No. U-4180, DDI Subcontract No. 66-1, AF 04(695)-976, Philco-Ford Aeronutronic Div, 1 Sep., 1967

[14] Montenbruck, Oliver, Gill, Eberhard, *Satellite Orbits*, Springer, 2000

[15] Lagrange, Collected Works, Vols. V and VI., 1766

[16] Lawson, Charles, L., Hanson, Richard J., *Solving Least Squares Problems*, Prentice-Hall, New Jersey, 1974

[17] Moulton, Forest Ray, *An Introduction to Celestial Mechanics*, MacMillan 1902, Dover 1970

[18] Taylor, Angus E., *Calculus*, Prentice-Hall, 1959

[19] Todhunter, I., *Spherical Trigonometry*, MacMillan, 1863, GOOGLE it

[20] Walters, Louis, telephone conversation, 29 May 2009

[21] Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, Oxford Science Pub., 1965

A. EQUINOCTIAL ORBIT ELEMENT HISTORY

Sputnik I was launched on 4 October 1957. Three months later Dr Lou Walters[20] hired six of Professor Samuel Herrick's graduate students from UCLA, and Samuel Herrick, into Aeronutronic Systems, Inc. of Newport-Beach CA, a fully owned subsidiary of the Ford Motor Company[7]. Lou thereby became the astrodynamics manager of what would become one of the most influential and close-knit groups of US astrodynamicists at that time, particularly with respect to US military programs. Jeannine Arsenault and Paul Koskela were members of Lou Walter's group. The naming, definition, and analysis of the *equinoctial* orbit elements would emerge from this group, and be documented and presented by Paul Koskela[13] in 1967. Jeannine Arsenault played a key role in equinoctial orbit element analysis, and she derived their rigorous perturbative time-derivatives (presented on pages 86 and 87 of Koskela[13]).

When the orbit eccentricity or orbit inclination of any space object is in the neighborhood of zero, then there are small divisors in the perturbative expressions with Kepler orbit elements. But for these classes of orbits there are no small divisors in perturbative expressions with the equinoctial orbit elements. Thus equinoctial orbit elements are preferred for near-circular or near-equatorial[8] orbits. My first experience (1968) with equinoctial orbit elements was obtained during development and deployment of the USAF Advanced Orbit Ephemeris Subsystem (AOES) while I was employed with System Development Corporation in Santa Monica between 1968 and 1973. The well-written AOES math-spec (1967) by Koskela[13] was used extensively.

But there is confusion as to who named the equinoctial orbit elements, when they were named, how they were defined, and how they were developed. Richard Battin said (see Battin[4] page 492) in 1987 that they were named by Roger Broucke in 1972 , Broucke said (see Broucke[5] page 303) they were named by Jeannine Arsenault[9] in 1970, and Jeannine has recently said[3] that Herrick may have first suggested the name prior to 1967.

[7] When Ford bought Philco in 1963, Aeronutronic Systems, Inc. became known as the Aeronutronic Division of Philco-Ford[3].

[8] When inclination i is in the neighborhood of π, the equinoctial elements are redefined to shift the singularity to $i = 0$.

[9] Broucke cited the 1970 paper[1] by Arsenault, Ford, and Koskela.

Jeannine Arsenault and Paul Koskela

AOES was used operationally between 1969 and 1992 in the USAF Satellite Test Center (STC) in Sunnyvale for the USAF Satellite Control Facility (SCF). The mathematics specification[13] for the AOES was prepared by Paul Koskela, with significant contributions by J. L. Arsenault (Jeannine), L. G. Walters (Lou), C. G. Hilton (Jeff), R. G. Schinnerer (Ralph), P.E. Koskela (Paul), R. F. Olmsted (Richard), K. C. Ford (Ken), and G. A. Mahon (George). Samuel Herrick was acknowledged therein for "primacy" of his "pioneering mathematical investigations". The six equinoctial orbit elements of the set $\{a_f, a_g, n, L, \chi, \psi\}$, presented here with Equation 33, were first defined by Arsenault and Koskela on page 68 of the AOES[10] math-spec[13] (1967)

$$\begin{bmatrix} a_f \\ a_g \\ n \\ L \\ \chi \\ \psi \end{bmatrix} = \begin{bmatrix} e\cos(\Omega+\omega) \\ e\sin(\Omega+\omega) \\ \sqrt{\mu/a^3} \\ \Omega+\omega+M \\ \sin i \sin\Omega/(1+\cos i) \\ \sin i \cos\Omega/(1+\cos i) \end{bmatrix} = \begin{bmatrix} \mathbf{a}\cdot\mathbf{F} \\ \mathbf{a}\cdot\mathbf{G} \\ \text{mean motion} \\ \text{mean longitude} \\ W_X/(1+W_Z) \\ -W_Y/(1+W_Z) \end{bmatrix} \tag{33}$$

as functions of the six Kepler orbit elements

$$\begin{bmatrix} a \\ e \\ M \\ i \\ \Omega \\ \omega \end{bmatrix} = \begin{bmatrix} \text{semi-major axis} \\ \text{eccentricity} \\ \text{mean anomaly} \\ \text{inclination} \\ \text{node} \\ \text{argument of perigee} \end{bmatrix} \tag{34}$$

Let $\mathbf{P}$ denote the unit vector pointing to perigee. The definition[11] for vector $\mathbf{a} = e\mathbf{P}$ in Equation 33 follows[12] Koskela[13] (page 44). ECI components W_X, W_Y, and W_Z are those of the unit vector $\mathbf{W} = \mathbf{h}/h$, where $\mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}}$, $h = \sqrt{\mathbf{h}\cdot\mathbf{h}}$, $\mathbf{r}$ is the Earth centered spacecraft position vector, and the vector time derivative $\dot{\mathbf{r}} = d\mathbf{r}/dt$ for velocity is referred to an inertial frame. ECI components of the right-handed orthonormal vector triad $(\mathbf{F}, \mathbf{G}, \mathbf{W})$ were defined by Herrick[11] page 108. The unit vector $\mathbf{F}$ points to the vernal equinox when the orbit plane coincides with the equator, and for this case is *equinoctial*. Notice that when the equinoctial set is employed with the rigorous theory of variation of parameters (VOP), then n is the osculating time-derivative of L, because $L = L_0 + n\,[t - t_0]$. Thus n is preferred to a or P, and this was the choice by Arsenault and Koskela. This set was selected for AOES orbit determination, in preference to all other possible sets of orbit elements.

From the fourth paragraph of Acknowledgements in Koskela[13](1967) page iii: "The details for the equinoctial (F,G) set of orbit parameters were developed by Mrs. Arsenault ...". The reference to FG elements refers obviously to use of the orthonormal vector triad $(\mathbf{F}, \mathbf{G}, \mathbf{W})$. The equinoctial orbit elements were referred to as FG elements thereafter in the AOES specification, and were known as FG elements by personnel working with the AOES. Arsenault's choices for symbols a_f and a_g from definitions $a_f = \mathbf{a}\cdot\mathbf{F}$ and $a_g = \mathbf{a}\cdot\mathbf{G}$ are descriptive and easy to remember, and her symbols for mean longitude L and mean motion n were used by Herrick and other professors. Theses same equinoctial orbit element definitions were given in the Nomenclature Section of the 1970 paper[1] by Arsenault, Ford, and Koskela.

Samuel Herrick

From Volume I[10], the unit vector $\mathbf{W}$ is defined on pages 49 and 50, and unit vectors $\mathbf{F}$ and $\mathbf{G}$ are defined on page 55, to define the right-handed orthonormal vector triad $(\mathbf{F}, \mathbf{G}, \mathbf{W})$. Mean longitude $L = L_0 + n\,[t - t_0] = M + \omega + \Omega$ is presented on page 494.

[10] This document is available from AGI as a PDF file.

[11] Koskela (1967) departs from Herrick's definition of $\mathbf{a}$. Herrick defines vector $\mathbf{a}$ with $\mathbf{a} = \sqrt{\mu}\, e\mathbf{P}$ (Herrick[10] (4E18) page 65), and refers to $\mathbf{a}$ as a Laplacian integral. So the Laplacian integral vector $\mathbf{a}$ points to perigee and has length $\sqrt{\mu}\, e$.

[12] Battin[4] (1987, page 116) calls the Koskela vector $\mathbf{a}$ the eccentricity vector $\mathbf{e}$ because it has length e.

From Volume II[11], Equations

$$a_F = e\cos\tilde{\omega}$$

$$a_G = e\sin\tilde{\omega}$$

are presented on page 110, with

$$\tilde{\omega} = \omega + \Omega$$

on page 105. Thus the descriptive names a_F and a_G appear to have been introduced by Herrick[13]. Also see page 141 for use of $e\cos\tilde{\omega}$ and $e\sin\tilde{\omega}$. On page 141 Herrick suggests replacing i and Ω with $\sin i\cos\Omega$ and $\sin i\sin\Omega$ for Kepler element expressions with denominator $\sin i$.

The complete selection of the equinoctial elements of Equation 33 had to wait for Arsenault, but they were anticipated in part by Herrick.

Different Names

The equinoctial orbit elements were renamed by Roger Broucke and Paul Cefola[5] in 1972

$$\begin{bmatrix} a \\ h \\ k \\ \lambda_0 \\ p \\ q \end{bmatrix} = \begin{bmatrix} a \\ e\sin(\omega+\Omega) \\ e\cos(\omega+\Omega) \\ M_0+\omega+\Omega \\ [\tan(i/2)]\sin\Omega \\ [\tan(i/2)]\cos\Omega \end{bmatrix} \tag{35}$$

where:

$$\tan(i/2) = \sin i/(1+\cos i) \tag{36}$$

Broucke and Cefola say that the elements given by Equation 35 "are the same as those used by Arsenault"; i.e., the elements of Equation 33. Notation $k = e\cos\tilde{\omega}$ and $h = e\sin\tilde{\omega}$ adopted by Broucke and Cefola is the same as that found in Brouwer and Clemence[6] (page 287, 1961). Forest Ray Moulton (1902) presented $h = e\sin\pi$ and $l = e\cos\pi$ ($\pi = \omega+\Omega$?) on page 421[17] in his discussion of secular terms in finite expansions of certain functions, and Moulton refers to Lagrange[15](1766) as source for these expressions. Perhaps this prompts the Broucke and Cefola claim that "These orbit elements were used more than a century ago (by Lagrange) ..."

Relative to Arsenault and Koskela[13] (1967), Broucke and Cefola[5] (1972) replaced the mean motion n with the semi-major axis a, they used the half-angle tangent function identity of Equation 36 (presented by Arsenault[1], Ford and Koskela in 1970) for the last two equinoctial elements, and they assigned different symbols to the last five elements of the right-hand column of Equation 35.

Broucke and Cefola presented partial derivatives of position and velocity components with respect to equinoctial elements, and they derived and presented the Poisson and Lagrange brackets for the equinoctial elements.

B. NONLINEAR LEAST SQUARES

Algorithms for solution to iterative nonlinear least squares (LS) problems are frequently referred to as *Gauss-Newton* algorithms. But the papers on Gauss-Newton algorithms I have reviewed completely conceal the role of the LS transition function $\Phi_{j,0}$ defined by Equation 40 below. *Size* of six-dimensional LS orbit convergence boundaries depends in a critical manner on the LS transition function $\Phi_{j,0}$. For GEO, transitive partial derivatives for a (degree,order) = (6,6) geopotential are

[13] Herrick's two volumes of Astrodynamics were in possession of his students long before they were published in 1971 and 1972.

required, transitive partial derivatives for Sun and Moon gravitational perturbations are required, and transitive partial derivatives for solar photon pressure are required. The *size* of six-dimensional LS orbit convergence boundaries is much larger with the use of these transitive partial derivatives than without them. The reader is referred to Chapter 3 of Montenbruck and Gill[14] for details on modeling these partial derivatives.

Normal Equation

Let n denote the size of state, and let m denote the number of distinct scalar measurements to be processed at each time t_j, $j \in \{1, 2, \ldots, m\}$. Matrices A_j, Δy, W_j, ΔX_0, have dimensions $m \times n$, $m \times 1$, $m \times m$, $n \times 1$ respectively. The summed form of the matrix LS Normal Equation

$$\sum_{j=1}^{m} A_j^T W_j A_j \ \Delta X_0 = \sum_{j=1}^{m} A_j^T W_j \, \Delta y_j \tag{37}$$

is algebraically identical[14] to the unsummed form of the matrix LS Normal Equation

$$A^T W A \ \Delta X_0 = A^T W \, \Delta y \tag{38}$$

where the $m \times n$ matrix A_j of Equation 37 is defined with

$$A_j = \frac{\partial y_j}{\partial X_j} \frac{\partial X_j}{\partial X_0} \tag{39}$$

The $n \times n$ jacobian matrix of partial derivatives $\partial X_j / \partial X_0$ is formally known as a linear state transition matrix $\Phi_{j,0}$

$$\Phi_{j,0} = \frac{\partial X_j}{\partial X_0} \tag{40}$$

and the $m \times n$ jacobian matrix of partial derivatives $\partial y_j / \partial X_j$ is frequently abreviated with H_j

$$H_j = \frac{\partial y_j}{\partial X_j} \tag{41}$$

Thus

$$A_j = H_j \Phi_{j,0} \tag{42}$$

The measurement residual Δy_j is defined by

$$\Delta y_j = y_j - y\left(\hat{X}_{j|0}\right) \tag{43}$$

where y_j is the measurement value and $y\left(\hat{X}_{j|0}\right)$ is the nonlinear measurement representation from the orbit estimate $\hat{X}_{j|0}$. The weighting matrix W_j has the inverse measurement residual covariance $m \times m$ matrix

$$C_j = W_j^{-1} \tag{44}$$

and is usually treated as a constant diagonal $m \times m$ matrix. Refering to Equation 38, then

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \tag{45}$$

[14] But the two are not operationally equivalent.

$$\Delta y = \begin{bmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_m \end{bmatrix} \tag{46}$$

$$W = \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_m \end{bmatrix} \tag{47}$$

The measurement residual RMS is defined by

$$RMS = \sqrt{\frac{1}{m}\sum_{j=1}^{m} \Delta y_j^2} = \sqrt{\Delta y^T \Delta y / m} \tag{48}$$

The Least Squares Equation

Define

$$\tilde{A} = W^{1/2} A \tag{49}$$

$$\Delta\tilde{y} = W^{1/2} \Delta y \tag{50}$$

where $W = W^{1/2} W^{1/2}$ is an $m \times m$ positive-definite diagonal matrix, $\Delta\tilde{y}$ is an $m \times 1$ matrix, and $\tilde{A}$ is an $m \times n$ matrix with $m \geq n$. Then Equation 38 becomes

$$\tilde{A}^T \tilde{A}\ \Delta X_0 = \tilde{A}^T \Delta\tilde{y} \tag{51}$$

Equation 51 is derived trivially from the equivalent equation

$$\tilde{A}\ \Delta X_0 = \Delta\tilde{y} \tag{52}$$

with left multiplication by $\tilde{A}^T$. Equation 52 is referred to as the *least squares equation* by several authors.

Determinant and Eigenvalues

Given the $m \times n$ matrix $\tilde{A}$ of rank $r \leq n$ where $m \geq n$, and given a consistent $m \times 1$ column matrix $\Delta\tilde{y}$, search success depends on rank $r = n$ for each iterative solution to Equation 52 for ΔX_0. The LS solution to Equation 52 is accomplished with $\tilde{A} = QR$ decomposition[15]. For the QR decomposition of $\tilde{A}$ we have $Q^T Q = I_{n \times n}$, and the $m \times n$ matrix R has an $n \times n$ upper triangular submatrix $\check{R}$ with the N eigenvalues[16] of $\check{R}$ presented on the diagonal of $\check{R}$. After QR decomposition, Equation 52 is solved for ΔX_0 with backward divisions and substitutions beginning with the bottom-right element of $\check{R}$. If any of the diagonal elements of $\check{R}$ is zero, then a divide-by-zero problem is encountered. This process is similar to Gaussian elimination, and is equivalent to inversion of matrix $\check{R}$.

The solution to Equation 38 requires inversion of the LS normal matrix $A^T W A$ (the information matrix). This is most efficiently and economically accomplished with Choleski decomposition[17]. Use of the normal equation solution is twice as fast as QR decomposition of $\tilde{A}$, but suffers from accuracy loss[18]. But for either technique the LS solution fails when rank $r < n$.

[15] See Lawson and Hanson[16] Chapter 18 and Appendix B. Singular value decomposition (SVD) is useful for analysis here but not for speed.

[16] The eigenvalues of $\check{R}$ are also the eigenvalues of $A^T W A$.

[17] See Wilkinson[21] Sections 42 and 44, and Lawson and Hanson[16] Chapter 19

[18] With double precision calculations this is due to the squaring operations to form $A^T W A$.

With an infinite computer word mantissa, each of the N eigenvalues of the LS information matrix $A^T W A$ would be either positive or zero, and solution of Equation 52 would be equivalent to solution of Equation 38. Since the determinant Δ of $A^T W A$ is the product of it's n eigenvalues, then Δ would be either positive or zero. The determinant Δ globally captures the information content of $A^T W A$. Now $\Delta > 0$ if and only if $r = n$. If $\Delta = 0$, then there exists no information on which to perform initial orbit determination.

But in practice, using double precision calculations on a computer (mantissa with 15^+ decimals), one never sees a zero eigenvalue or a zero determinant. In practice it is difficult to distinguish $\Delta = 0$ from $\Delta > 0$ when Δ is very small. And Δ is very small for LS orbit determination.

LS Convergence Criteria

The operator selects a threshold N_{MAX} for maximum number of least squares iterations without convergence, with default $N_{MAX} = 15$. Iterative least squares RMS *convergence* is defined for scalar measurements after iteration k if the root-mean-square $RMS^{(k)}$ on scalar measurement residuals $\Delta y_j^{(k)}$, $j \in \{1, 2, \ldots\}$, is reduced to less than a small predefined positive constant ϵ_{RMS}

$$RMS^{(k)} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left[\Delta y_j^2\right]^{(k)}} < \epsilon_{RMS} \tag{53}$$

Otherwise RMS divergence is defined. But divergence is quickly identified when the GEO LS estimate goes hyperbolic.

Iterative least squares relative-RMS $RRMS$ "convergence" is defined for scalar measurements after iteration k if the relative-root-mean-square $RRMS^{(k)}$ is reduced to less than a small predefined positive constant ϵ_{RRMS}

$$RRMS^{(k)} = \frac{\left|RMS^{(k)} - RMS^{(k-1)}\right|}{RMS^{(k-1)}} < \epsilon_{RRMS} \tag{54}$$

Otherwise $RRMS$ divergence is defined. Operators usually selects one or both of these convergence criteria. We have discovered scalar examples where iterative "convergence" is achieved with the *relative* residual RMS $RRMS^{(k)} < \epsilon_{RRMS}$, and yet the residual RMS $RMS^{(k)}$ is very large; i.e., the orbit estimate is far from the true orbit ($RMS^{(k)} >> \epsilon_{RMS}$). And we have discovered scalar examples where the iterative "convergence" is *not* achieved with the *relative* residual RMS $RRMS^{(k)}$, and yet the residual RMS $RMS^{(k)} < \epsilon_{RMS}$ indicates convergence. Thus we recommend use only of the scalar measurement $RMS^{(k)} < \epsilon_{RMS}$ convergence criterion.

If multiple measurements are to be processed at each measurement time-tag, then the scalar $RMS^{(k)}$ is replaced with the scalar *weighted* root-mean-square $WRMS^{(k)}$, where each squared measurement residual Δy_j^2 of type i is replaced by $W^i \Delta y_j^2$ in Expression 53, such that $\sum_i W^i = 1$. Notation $RRMS^{(k)}$ is replaced with scalar relative-weighted-root-mean-square $RWRMS^{(k)}$.

Cauchy Convergence

A necessary and sufficient condition[19] for the sequence $\left\{RMS^{(k)}\right\}$, for $k \in \{1, 2, 3, \ldots\}$, to have a limit is that the absolute difference $\left|RMS^{(k)} - RMS^{(j)}\right|$ approach 0 as j and $k \to \infty$, where $j \in \{k-1, k-2, \ldots\}$. That is, for each positive ϵ_{RMS} there is some positive integer N such that $\left|RMS^{(k)} - RMS^{(j)}\right| < \epsilon_{RMS}$ if $N \leq k$ and $N \leq j$. We can approximate the rigorous Cauchy convergence criterion using the residual RMS, but *not* using the residual relative-RMS.

[19] Cauchy convergence is presented in many books on calculus; e.g., see Taylor[18] page 453.