Optimal Orbit Determination*

James R Wright†

Abstract

Existing definitions for *optimal orbit determination* are not satisfactory. Any useful definition must explicitly address questions relating to sequential processing, linearization, performance function and its extremalization, state estimate structure completeness, use of physics in applicable stochastic processes, and criteria for validation. Such a definition is presented.

1 Introduction

Orbit determination refers to the estimation of orbits of spacecraft (or natural satellites or binary stars) relative to primary celestial bodies, given applicable measurements. All useful orbit determination methods produce orbit estimates, and all orbit estimates have errors. But what is *optimal* orbit determination?

By itself, the adjective *optimal* refers [21] to most desirable, most favorable, or most satisfactory. But most satisfactory to whom? There are choices to make from available orbit determination methods. The fastest methods are the least accurate. Should we prefer sequential methods to batch methods? Should we minimize the size of measurement residuals or the size of orbit errors? How should we model measurement residuals and orbit errors?

All orbit determination problems are multidimensional and nonlinear. Should we attempt a multidimensional nonlinear solution directly? Or should we use a linearization method? If so, is there a preferred method for linearization?

What is optimal orbit determination? My answer is given in Section 4.7 herein.

2 Orbit Determination Methods

Orbit determination methods are distinguished with three categories according to their inputs, outputs, and accuracy performance:

- Initial orbit determination (IOD)

†Engineer, Analytical Graphics, Inc.

- Batch least squares differential corrections (LS)
- Sequential processing (SP)

Operationally, the order in which these methods are used defines a dependency tree: IOD output is LS input, and LS output is SP input:

$$\text{IOD} \Longrightarrow \text{LS} \Longrightarrow \text{SP}$$

For accuracy performance: IOD produces *crude* orbit estimates, LS produces *refined* orbit estimates in a batch processing mode, and SP produces *refined* orbit estimates in a sequential mode.

IOD methods input tracking measurements with tracking platform locations, and output spacecraft position and velocity estimates. *No* a priori orbit estimate is required. IOD methods are characterized by the use of approximations and/or recursive algorithms operating on nonlinear *two-body* dynamics. IOD methods were derived by various authors[16]: LaPlace, Gauss, Lagrange, Lambert, Gibbs, Herrick, Williams, Stumpp, Lancaster, Blanchard, Gooding, and Smith. Operationally, the orbit determination process is frequently begun, or restarted, with IOD.

LS methods input tracking measurements, tracking platform locations, an a priori orbit estimate, and other a priori state parameter values. The LS output contains a refined orbit estimate. An a priori orbit estimate *is* required. LS methods consist of a sequence of linear LS corrections. At each correction, the LS algorithm processes a subset of a fixed batch of tracking measurements simultaneously. Each correction is characterized by the minimization of the sum of squares of measurement residuals, and by the use of definitive deterministic force models. The LS method was derived first by Gauss in 1795, and then independently by Legendre. Gauss' English translation was first available in 1857[1].

SP methods input tracking measurements with tracking platform locations, an a priori state estimate (inclusive of orbit estimate), and an a priori state error covariance matrix. An a priori state estimate *is* required, and an a priori state error covariance matrix *is* required. SP methods output state estimates and state error covariance matrices. SP methods are forward-time recursive sequential machines consisting of a repeating pattern of *time update* of the state estimate and *measurement update* of the state estimate. The *time update* propagates the state estimate and its error covariance forward, and the *measurement update* incorporates the next measurement. State estimate error magnitudes grow during the time update interval, and are reduced at times of measurement update. The growth of state estimate error magnitudes is due most significantly to errors in force models: gravity, air-drag, solar photon pressure, spacecraft thrusting for trajectory and attitude maneuvers, outgassing of stored fuel, and thermal radiation. SP methods account for the recursive growth and reduction of state estimate errors according to the time order in which they are realized. SP methods were developed and proposed by Swerling[9], Kalman[4], Bucy[7], and others.

In summary, IOD methods produce crude state estimates $\hat{X}$ from measurements alone, whereas LS and SP methods produce refined corrections $\Delta\hat{X}$ to a priori state estimates $\hat{X}$.

3 Mathematical Operators for SP Methods

3.1 Subscript Notation

3.1.1 State Matrices

The state estimate $\hat{X}$ is referred to two times with the notation[5]:

$$\hat{X}_{j|i} \equiv \hat{X}\left(t_j|t_i\right) \tag{1}$$

where i, $j \in \{0, 1, 2, \ldots\}$. The time t_j to the left of the vertical bar denotes the epoch for $\hat{X}$, and is driven by the filter time update function. The time t_i to the right of the bar denotes the time-tag of the last measurement processed to form $\hat{X}$, and is driven by the filter measurement update function. Examples: $\hat{X}_{7|6}$ refers to the state estimate at time t_7, given the last measurement processed at time t_6, whereas $\hat{X}_{7|7}$ refers to the state estimate at time t_7, given the last measurement processed at time t_7. Evidently, $\hat{X}_{7|6}$ was obtained by filter time update of $\hat{X}_{6|6}$ from t_6 to t_7.

Similar notation is used for the state estimate correction:

$$\Delta\hat{X}_{j|i} \equiv \Delta\hat{X}\left(t_j|t_i\right) \tag{2}$$

and the state estimate error covariance matrix:

$$P_{j|i} \equiv P\left(t_j|t_i\right) \tag{3}$$

3.1.2 Measurement Matrices

Denote a measurement at time t_j with y_j, and denote a measurement estimate (representation) at time t_j with $\hat{y}_{j|h}$.

3.2 Nonlinear Operators

Nonlinear operators are required in the state estimate time update and the state estimate measurement update for SP methods of orbit determination.

3.2.1 State Propagation

Let φ denote a nonlinear operator that propagates the state estimate $\hat{X}_{i|h}$ from time t_i to time t_j :

$$\hat{X}_{j|h} = \varphi\left\{t_j; \hat{X}_{i|h}, t_i\right\} \tag{4}$$

3.2.2 Measurement Representation

Let $y(\cdot)$ denote a nonlinear operator that calculates the measurement representation $\hat{y}_{j|h}$, given the state estimate $\hat{X}_{j|h}$:

$$\hat{y}_{j|h} = y\left(\hat{X}_{j|h}, t_j\right) \tag{5}$$

3.3 Linear Operators

3.3.1 State Estimate Error Propagation

Let $\Phi_{j,i} \equiv \Phi(t_j, t_i)$ denote the linear operator that propagates the state error estimate $\Delta\hat{X}_{i|h}$ from time t_i to time t_j

$$\Delta\hat{X}_{j|h} = \Phi_{j,i}\Delta\hat{X}_{i|h} \tag{6}$$

where

$$\Phi_{j,i} = \left[\frac{\partial X_j}{\partial X_i}\right]_{\hat{X}_{j|h}} \tag{7}$$

and where evaluation derives from $\hat{X}_{j|h}$.

3.3.2 Measurement Residual

Let Δy_j denote the linear operator that defines the measurement residual at time t_j:

$$\Delta y_j = y_j - \hat{y}_{j|h} \tag{8}$$

3.3.3 Measurement/State Partials Jacobian

Let H_j denote the jacobian of measurement to state partial derivatives at time t_j:

$$H_j = \left[\frac{\partial y_j}{\partial X_j}\right]_{\hat{X}_{j|h}} \tag{9}$$

where evaluation derives from $\hat{X}_{j|h}$.

4 Definitions

4.1 State Estimate Reference for Linearization of SP Methods

Evaluation of the measurement representation $\hat{y}_{j|h}$ defined by Eq. 5 requires the use of some a priori state estimate reference $\hat{X}_{j|h}$, where $t_j \geq t_h$. A similar

requirement is associated with Eqs. 7 and 9. I refer to $\hat{X}_{j|h}$ as the state estimate *reference* for linearization.

4.2 Local Linearization

Let t_k and $t_{k+1} \geq t_k$ be the time tags of adjacent ordered measurements y_k and y_{k+1}, for $k \in \{0, 1, 2, \ldots\}$. That is, there exist no measurements between y_k and y_{k+1}. Then the use of $\hat{X}_{k+1|k}$ as the state estimate *reference* for all linearizations at time t_{k+1} defines *local* linearization at time t_{k+1}.

The use of any state estimate reference other than $\hat{X}_{k+1|k}$ at time t_{k+1} for linearization is a non-local linearization at time t_{k+1}.

4.3 Global Linearization

Given the integer variable $k \in \{0, 1, 2, \ldots\}$ and any fixed non-negative integer j, then the use of $\hat{X}_{k|j}$ as the state estimate *reference* for linearization at time t_k for each k defines *global* linearization.

4.4 Observability

A particular state estimate parameter is *observable* to a particular measurement if and only if the sequential processing of that measurement reduces the estimate error variance on that parameter.

4.5 Completeness

The state estimate structure is *complete* if and only if all parameters, that are both unknown and observable, are contained in the state estimate structure.

4.6 McReynolds' Filter-Smoother Consistency Test

Given the matrix response values to a forward running sequential filter and an associated backward running sequential fixed interval smoother for times t_k, $k \in \{0, 1, 2, \ldots, L\}$, a rigorous consistency test was defined[6] at each time and on each of the N state estimate elements.

Calculate the $N \times N$ difference matrix $\bar{P}_{k|L}$ between the filtered covariance matrix $\hat{P}_{k|k}$ and the smoothed covariance matrix $\tilde{P}_{k|L}$ for time t_k:

$$\bar{P}_{k|L} = \hat{P}_{k|k} - \tilde{P}_{k|L} \tag{10}$$

for each $k \in \{0, 1, 2, \ldots, L\}$. The difference matrix $\bar{P}_{k|L}$ has no negative eigenvalues. Denote the square root of the i^{th} main diagonal element of the $N \times N$ difference matrix $\bar{P}_{k|L}$ as $\sigma^i_{k|L}$. Also calculate the $N \times 1$ difference matrix $\bar{X}_{k|L}$ between filtered state estimate $\hat{X}_{k|k}$ and smoothed state estimate $\tilde{X}_{k|L}$ for time t_k:

$$\bar{X}_{k|L} = \hat{X}_{k|k} - \check{X}_{k|L} \tag{11}$$

Denote the i^{th} element of the $N \times 1$ difference matrix $\bar{X}_{k|L}$ as $\bar{X}^i_{k|L}$. Now calculate and graph the ratio :

$$R^i_{k|L} = \bar{X}^i_{k|L} / \sigma^i_{k|L} \tag{12}$$

for each $i \in \{1, 2, \ldots, N\}$ and for each $k \in \{0, 1, 2, \ldots, L\}$.

4.6.1 Test

If for each $i \in \{1, 2, \ldots, N\}$ and for each $k \in \{0, 1, 2, \ldots, L\}$ we have:

$$\left| R^i_{k|L} \right| \leq 3 \tag{13}$$

then McReynolds' filter-smoother test is satisfied globally. If for each $i \in \{1, 2, \ldots, N\}$ and for each $k \in \{0, 1, 2, \ldots, L\}$ we have:

$$\left| R^i_{k|L} \right| > 3 \tag{14}$$

then McReynolds' filter-smoother test is failed globally. For each i for which inequality 13 is satisfied McReynolds' filter-smoother test is passed for that state estimate element, and for each i for which inequality 14 is satisfied McReynolds' filter-smoother test is failed for that state estimate element.

4.7 Optimal Orbit Determination

By optimal orbit determination, I mean that the method used to calculate the state estimate (containing the orbit estimate) satisfies the following eight requirements:

1. Sequential processing (SP) is used to account for force modeling errors and measurement information in the time order in which they are realized.

2. The optimal state error estimate $\Delta\hat{X}$ is the expectation of the state error ΔX given the measurement residual Δy. That is: $\Delta\hat{X} = E\{\Delta X | \Delta y\}$. This is Sherman's Theorem [2][3][4][5].

3. Linearization of state estimate time transition and state to measurement representation is *local* in time, *not global.*

4. The state estimate structure is *complete*

5. All state estimate models and state estimate error model approximations are derived from appropriate force modeling *physics,* and measurement sensor performance

6. All measurement models and measurement error model approximations are derived from appropriate sensor hardware definition and associated physics, and measurement sensor performance

7. Necessary conditions for real data:

   - Measurement residuals approximate Gaussian white noise[5][8]
   - McReynold's filter-smoother consistency test is satisfied with probability 0.99

8. Sufficient conditions for simulated data: The state estimate errors agree with the state estimate error covariance function.

The first six requirements define standards for optimal algorithm design, and the creation of a *realistic* state estimate error covariance function. The last two requirements enable validation: They define realizable test criteria for optimality. The last requirement implies the development and use of a physically realistic measurement simulator.

5 Discussion

5.1 SP Initialization

Successful initialization of sequential processing requires an a priori state estimate (and error covariance) that is within the capture region of the sequential processor.

Capture refers to the subsequent non-divergence of the sequential processor. *Divergence* refers to the unbounded separation between the state estimate and the true state. Divergence is most directly identified by the autonomous sequential rejection of all measurements.

5.2 Sherman's Theorem

Sherman's Theorem is applicable to a *linear* state estimate; i.e., a condition where the state estimate is a linear combination of available measurements. But in all orbit determination problems, the orbit (substate) estimate is a *nonlinear* function of available measurements. We must linearize in order to use Sherman's Theorem.

Optimal orbit determination requires measurement linearization about the local state estimate $\hat{X}_{k+1|k}$ to produce a local measurement *residual*:

$$\Delta y_{k+1} = y_{k+1} - y\left(\hat{X}_{k+1|k}\right)$$

and linearization about the same local state estimate $\hat{X}_{k+1|k}$ to produce a local state *error* estimate $\Delta\hat{X}_{k+1|k+1}$, given Δy_{k+1}. Local linearization enables a

linear relation between each state error estimate $\Delta\hat{X}_{k+1|k+1}$ and each measurement residual Δy_{k+1}. With *local* linearization, one applies Sherman's Theorem anew to each scalar measurement residual, never simultaneously to a batch of measurement residuals.

Let $\Delta X_{k+1|k} = X_{k+1|k} - \hat{X}_{k+1|k}$ define the error in state estimate $\hat{X}_{k+1|k}$, and let $\delta X_{k+1|k} = \Delta X_{k+1|k} - \Delta\hat{X}_{k+1|k}$ define the error in $\Delta\hat{X}_{k+1|k}$. Ideally $\delta X_{k+1|k} = 0$, and the state error estimate $\Delta\hat{X}_{k+1|k}$ is perfect. When $\delta X_{k+1|k} \neq 0$, assign a penalty (or loss) function $L = L\left(\delta X_{k+1|k}\right)$ with four admissibility requirements:

- L is a scalar-valued function of the N state estimate variables.
- $L(0) = 0$, where the first 0 is an $N \times 1$ matrix of zeros. No loss is assigned when the state error estimate is perfect.
- $L\left(\delta X^a_{k+1|k}\right) \geq L\left(\delta X^b_{k+1|k}\right)$ whenever $\rho\left(\delta X^a_{k+1|k}\right) \geq \rho\left(\delta X^b_{k+1|k}\right)$, where ρ is a scalar-valued, non-negative, convex function of N variables. Thus L is defined to be a non-decreasing function of distance ρ from the origin. The closer $\delta X_{k+1|k}$ is to zero, the smaller the loss.
- $L\left(\delta X_{k+1|k}\right) = L\left(-\delta X_{k+1|k}\right)$. That is, $L(\cdot)$ is symmetric about the origin.

Performance $J\left[\delta X_{k+1|k}\right]$ is defined as the of the expectation of the loss:

$$J\left[\delta X_{k+1|k}\right] = E\left\{L\left(\delta X_{k+1|k}\right)\right\} \tag{15}$$

That is, our measure of performance is defined by the mean value of loss. Our goal is to minimize $J\left[\delta X_{k+1|k}\right]$.; i.e., to minimize the mean value of loss.

Denote the conditional probability distribution function of $\Delta X_{k+1|k}$ given Δy_{k+1} by:

$$P\left\{\Delta X_{k+1|k} \leq \xi | \Delta y_{k+1}\right\} = F\left\{\xi | \Delta y_{k+1}\right\} \tag{16}$$

The reader is referred to Chapter 5.0, Section 5.2, of Meditch [5] for the following theorems.

5.2.1 Most General Form

Given any admissible loss function $L\left(\delta X_{k+1|k}\right)$, and any conditional probability distribution function $F\left\{\xi | \Delta y_{k+1}\right\}$ such that $F\left\{\xi | \Delta y_{k+1}\right\}$ is:

- Symmetric about its mean $\bar{\xi}$
- Convex for all $\xi \leq \bar{\xi}$

then:

$$\Delta\hat{X}_{k+1|k+1} = E\left\{\Delta X_{k+1|k} | \Delta y_{k+1}\right\} \tag{17}$$

Application of the conditional mean $E\left\{\Delta X_{k+1|k}|\Delta y_{k+1}\right\}$ generates a global minimum to the performance function $J\left[\delta X_{k+1|k}\right]$. This is true for all combinations of admissible loss functions and symmetric and convex conditional probability distribution functions. Proof is due to Sherman[2][3].

5.2.2 Gaussian Distribution

Given any admissible loss function $L\left(\delta X_{k+1|k}\right)$, and Gaussian random variables $\Delta X_{k+1|k}$ and Δy_{k+1}, then:

$$\Delta\hat{X}_{k+1|k+1} = E\left\{\Delta X_{k+1|k}|\Delta y_{k+1}\right\} \tag{18}$$

Application of the conditional mean $E\left\{\Delta X_{k+1|k}|\Delta y_{k+1}\right\}$ generates a global minimum to the performance function $J\left[\delta X_{k+1|k}\right]$, even for asymmetric loss functions. Proof is due to Doob[14].

5.2.3 Mean Square Error

If $L\left(\delta X_{k+1|k}\right) = \left(\delta X_{k+1|k}\right)^T\left(\delta X_{k+1|k}\right)$, then:

$$\Delta\hat{X}_{k+1|k+1} = E\left\{\Delta X_{k+1|k}|\Delta y_{k+1}\right\} \tag{19}$$

The loss function $\left(\delta X_{k+1|k}\right)^T\left(\delta X_{k+1|k}\right)$ is referred to as the mean square state error. Minimization of the performance function $E\left\{\left(\delta X_{k+1|k}\right)^T\left(\delta X_{k+1|k}\right)\right\}$ results, in part, in the minimization of mean square orbit error. Application of the conditional mean $E\left\{\Delta X_{k+1|k}|\Delta y_{k+1}\right\}$ generates a global minimum to the performance function. In this case the conditional probability distribution function need not be either symmetric or convex. Proof is due to Doob[14].

5.3 Complete State Estimate

Consider any case where the state estimate structure is incomplete. The observable parameter neglected in the state estimate structure will *alias* into the estimated orbit elements, significantly degrading them. Thus one needs an appropriate place in the state estimate structure to put every observable effect.

5.4 Gaussian White Noise

5.4.1 What is it?

In one dimension, I think of a Gaussian white noise sequence as a sequence of ratios from a random walk sequence R_j, $j \in \{0, 1, 2, \ldots\}$. The numerator in each ratio is the difference $(R_{j+1} - R_j)$ in the random walk functional across a specified time interval $[t_j, t_{j+1}]$, and the denomenator is the associated time difference $(t_{j+1} - t_j)$. The ratio limit $(t_{j+1} - t_j) \longrightarrow 0$ does not exist[14]. Thus

one must always use Gaussian white noise in a granular manner. The Wiener-Levi (random walk) sequence developed in Papoulis [11] provides appropriate useful results for application.

5.4.2 Measurements

Gaussian white noise models are used appropriately for themal noise when associated with resistance in electronic circuits [12]. Thus Gaussian white noise is used *directly* for modeling stochastic phenomena in clocks, transmitters, receivers, and sensors. Range and Doppler measurements consist of signal, Gaussian white noise, and various biases.

Gauss[1] used a Gaussian white noise model for orbit determination to represent noise in angles right ascension and declination.

5.4.3 Linear Systems

Gaussian white noise is used appropriately, but indirectly, as a linear system input to develop a Gaussian stochastic output functional with particular serial correlation properties. This provides a convenient method to represent stochastic modeling errors in some cases. But in other cases this method is useless; e.g., for acceleration modeling errors that derive from errors in modeling the geopotential.

5.4.4 Not Spacecraft Acceleration Model Errors

Acceleration modeling errors that derive from errors in modeling the geopotential, atmospheric density, and solar photon pressure are random and nonstationary, but they are *not* white.

6 Summary

Optimal orbit determination refers to the satisfaction of the eight requirements given in Section 4.7. It remains to demonstrate mathematically and physically under exactly what set of conditions these requirements are realizable and mutually consistent. For now, I rely on their successful use for the past twenty-two years.

This paper is presented in response to a request from Dave Vallado.

References

[1] Gauss, K. F., *Theoria Motus Corporum Coelestium (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections),* Little, Brown, and Co., 1857, republished by Dover, 1963

[2] Sherman, S., *A Theorem on Convex Sets with Applications*, Ann. Math. Stat., 26, 763-767, 1955.

[3] Sherman, S., *Non-Mean-Square Error Criteria*, IRE Transactions on Information Theory, Vol. IT-4, 1958.

[4] Kalman, R. E., *New Methods in Wiener Filtering Theory,* Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability, edited by J. L. Bogdanoff and F. Kozin, John Wiley & Sons, New York, 1963.

[5] Meditch, J. S., *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, New York, 1969.

[6] McReynolds, S. R., Personal Communications, GE, 1981

[7] Bucy, Richard S., Joseph, Peter D., *Filtering for Stochastic Processes with Applications to Guidance*, Interscience Publishers, John Wiley & Sons, New York, 1968

[8] Anderson, Brian D. O., Moore, John B., *Optimal Filtering*, Prentice-Hall, New Jersy, 1979.

[9] Swerling, P. *A Proposed Stagewise Differential Correction Procedure for Satellite Tracking and Prediction*, P-1292, Rand Corporation, 8 Jan 1958

[10] Ho, Y.C., *The Method of Least Squares and Optimal Filtering Theory*, Rand Corporation, Oct 1962

[11] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

[12] Davenport, Wilbur B., Root, William L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.

[13] Wiberg, D. M., *Theory and Problems of State Space and Linear Systems,* Schaum's Outline Series, McGraw-Hill, New York, 1971.

[14] Doob, J. L., *Stochastic Processes,* John Wiley & Sons, New York, 1953.

[15] Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. 1, John Wiley & Sons, New York, 1950.

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

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

[18] King-Hele, Desmond, *Theory of Satellite Orbits in an Atmosphere*, Butterworths, London, 1964

[19] Wright, J. R., *Sequential Orbit Determination with Auto-Correlated Gravity Modeling Errors,* AIAA, Journal of Guidance and Control, Vol 4, No. 2, May-June 1981, page 304.

[20] Wright, J. R., *Orbit Determination Solution to the Non-Markov Gravity Error Problem,* AAS/AIAA Paper AAS 94-176, AAS/AIAA Spaceflight Mechanics Meeting, Cocoa Beach, FLA, Feb., 1994.

[21] Benton, William (publisher), *Webster's Third New International Dictionary of the English Language Unabridged*, Encyclopaedia Britannica, Inc, Chicago, 1966