SINGULAR VALUE DECOMPOSITION AND LEAST SQUARES ORBIT DETERMINATION

Vladimir Boikov[1], Zakhary N. Khutorovsky[2] and Kyle T. Alfriend[3]

Convergence using least squares for orbit determination is often difficult to assure when a) there are only a small number of measurements with large time intervals between the measurements (sparse measurements), b) the measurement accuracy is poor, c) there are crude or abnormal measurements present, or d) no good initial guess is available. The first case, accurate but sparse measurements, but with a good initial guess, represents a linear least-squares problem. It was shown long ago that such problems are successfully solved by applying singular value decomposition (SVD) to the matrix of the conditional equations. In the nonlinear case, the poor determination of the conditional equations matrix is further aggravated by the difficulties of obtaining a good Hessian approximation in regions which are remote from the solution. In this paper we present an algorithm using SVD in which both of these problems are resolved by choosing the dimension of the minimization subspace for each minimization step. If necessary, minimization in a low dimensionality subspace (2-D or 3-D) can be carried out with improved accuracy Hessian matrices. Results of the application of the algorithm are discussed for typical measurement-based orbit determination. We compare these results with those produced using the Gauss-Newton method. Areas of convergence of the proposed technique are investigated as well.

INTRODUCTION

Orbit determination is typically performed using batch least squares (LS) or the extended Kalman filter (EKF). Both methods are based on linearization about the estimated trajectory and require an a priori state estimate to initiate the estimation. The EKF also needs an a priori covariance. The US and Russia both use LS for maintaining their space object catalogs. Maintenance of the space object catalog is usually a relatively linear problem, consequently the performance of the LS for catalog maintenance is usually satisfactory. However, convergence problems can arise when the nonlinearities become significant. The scenarios in which significant nonlinearities can occur are: a) a poor initial or a priori state estimate, b) a long time interval between measurements, i.e., sparse measurements, and c) the presence of poor or abnormal measurements. Although each of these can arise in the catalog maintenance problem (orbit update) they are more likely to occur when determining the orbit from a single set

[1] Lead Scientist, System Research Center, Vympel Int. Corp., Moscow 123110, Russia.

[2] Manager Section, System Research Center, Vympel Int. Corp., Moscow 123110, Russia, Email: Z.Hutor@g23.relcom.ru

[3] TEES Distinguished Research Professor of Aerospace Engineering, Dept. of Aerospace Eng., H.R. Bright Bldg, Room 701, 3141 TAMU, College Station, TX 77843-3141, Email: alfriend@aero.tamu.edu.

of observations over a short period of time (track), and the subsequent correlation of tracks to establish an accurate orbit and catalog the object. In this case there is no a priori state estimate from previous observations so an initial estimate has to be developed from some of the observations to start the LS process. This is called initial orbit determination (IOD). The a priori state estimate to initiate the LS process is obtained from fitting an orbit through several of the observations. In the US the Herrick Gibbs[1] method is usually used to obtain the a priori state estimate when the track is from radar observations (range, azimuth, elevation); in Russia a different approach is used[2]. In both approaches the position estimate is relatively accurate, but the velocity estimate is likely to be poor. When the observations are range only, or angles only, then the a priori orbit estimate is likely to be poor. As a result of this poor initial state estimate the LS method may fail to converge. After an initial orbit is obtained from a single track, the state and covariance from the single track have to be propagated to the next track to determine if the tracks correlate and, if so, combine the two tracks to establish a more accurate orbit. The covariance is propagated using the 1st variational equations. When there is a long time between tracks, the errors in the dynamic model and the linear propagation of the covariance can generate large errors in both the state and covariance. These, in effect, become the a priori state and covariance for the 2nd track and, if the time interval between the tracks is large, they may be poor estimates and the LS method may not converge.

The presence of poor or abnormal measurements also creates a bigger problem in IOD. An abnormal measurement is one whose error is greater than 3□. In catalog maintenance the abnormal measurements are likely to just be rejected, but in IOD it is difficult to determine which measurements are abnormal, especially with a short track. The result is a poor orbit estimate which may not converge on the 1st track, but, even if it does, the error when propagated to the 2nd track may be very large.

Thus, even though each of these cases can occur in catalog maintenance, they are much more likely to occur in IOD. In all these cases the poor determination of the conditional equations matrix is further aggravated by the difficulty of obtaining a good Hessian approximation in regions that are remote from the solution. In this paper we propose an algorithm using singular value decomposition (SVD), in which both these problems are resolved by choosing the dimension of the minimization subspace for each minimization step. If necessary, minimization in a low dimensionality subspace (2-D or 3-D) can be carried out with improved accuracy Hessian matrices.

The next section presents the theory and the algorithm. Then, the algorithm is applied to several cases to demonstrate its effectiveness and superiority over least squares. This is followed by the Conclusions.

ALGORITHM

Overview

Let $\mathbf{x}$ be the vector representing the state (orbital elements) at epoch that we need to estimate, or determine, and let there be m measurements, each of which has a dimension of k_i. In the method of least squares we want to minimize the function

$$F(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{k_i}\left[\frac{\hat{s}_{ij} - s_{ij}(\mathbf{x})}{\sigma_{ij}}\right]^2 \tag{1}$$

where $\hat{s}_{ij}$ is the *j*-th component of the *i*-th measurement vector, $s_{ij}(\mathbf{x})$ is the corresponding predicted measurement based on the current estimate of the state $\vec{x}$, and the σ_{ij} are the associated standard deviations of the measurement errors. The condition for the minimum is

$$gradF(\mathbf{x}) \equiv g(\mathbf{x}) = 0\,. \tag{2}$$

Now define the normalized measurements

$$\hat{\bar{s}}_{ij} = \frac{\hat{s}_{ij}}{\sigma_{ij}}, \bar{s}_{ij}(\mathbf{x}) = \frac{s_{ij}(\mathbf{x})}{\sigma_{ij}} \tag{3}$$

Define the normalized measurement vector to be

$$\bar{\mathbf{s}} = \left(\bar{s}_{11},...,\bar{s}_{1k_1},\bar{s}_{21},...,\bar{s}_{2k_2},...,\bar{s}_{m1},...,\bar{s}_{mk_m}\right)^T \tag{4}$$

and similarly for $\hat{\bar{\mathbf{s}}}$. The dimension of these normalized vectors is $M = \sum_{i=1}^{m} k_i$. Equation (1) reduces to

$$F(\mathbf{x}) == \frac{1}{2}\left[\hat{\bar{\mathbf{s}}} - \bar{\mathbf{s}}(\mathbf{x})\right]^T\left[\hat{\bar{\mathbf{s}}} - \bar{\mathbf{s}}(\mathbf{x})\right] = \frac{1}{2}\mathbf{b}^T\mathbf{b}, \mathbf{b} = \left[\hat{\bar{\mathbf{s}}} - \bar{\mathbf{s}}(\mathbf{x})\right] \tag{5}$$

where $\mathbf{b}$ is the normalized residual vector. The gradient and Hessian of Eq. (5) are

$$\begin{aligned} g(\mathbf{x}) = &\quad F(\mathbf{x}) = \frac{\partial \mathbf{b}^T}{\partial \mathbf{x}}\mathbf{b}, \frac{\partial \mathbf{b}^T}{\partial \mathbf{x}} = -\frac{\partial \bar{\mathbf{s}}^T}{\partial \mathbf{x}} \\ G(\mathbf{x}) = &\quad {}^2F(\mathbf{x}) = \frac{\partial \mathbf{b}^T}{\partial \mathbf{x}}\frac{\partial \mathbf{b}}{\partial \mathbf{x}} + \sum_{i=1}^{M}\frac{\partial^2 b_i}{\partial^2 \mathbf{x}}b_i(\mathbf{x}) = \frac{\partial \bar{\mathbf{s}}^T}{\partial \mathbf{x}}\frac{\partial \bar{\mathbf{s}}}{\partial \mathbf{x}} - \sum_{i=1}^{M}\frac{\partial^2 \hat{\bar{s}}_i}{\partial^2 \mathbf{x}}b_i(\mathbf{x}) \end{aligned} \tag{6}$$

Using the following notation for the gradient and Hessian

$$\begin{aligned} A &= -\frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \bar{\mathbf{s}}(\mathbf{x})}{\partial \mathbf{x}} \\ B(\mathbf{x}) &= -\sum_{i=1}^{M}\frac{\partial^2 b_i}{\partial^2 \mathbf{x}}b_i(\vec{x}) = \sum_{i=1}^{M}\frac{\partial^2 \bar{s}_i}{\partial^2 \mathbf{x}}b_i(\vec{x}) \end{aligned} \tag{7}$$

where *A* is the Jacobian. Then the minimization problem reduces to

$$\left(A^T A + B\right)\Delta\mathbf{x} = A^T\mathbf{b} \tag{8}$$

Now consider the brackets in the left-hand side. Components of the residuals enter as factors into the matrix *B*. If the ***residuals*** are small, then the whole matrix *B* is also small and we can neglect *B* with respect to the matrix $A^T A$. Approximating the Hessian by $A^T A$ is called the Gauss-Newton method, and the solution is

$$\left(A^T A\right)\Delta\mathbf{x}_{GN} = A^T\mathbf{b} \tag{9}$$

where the subscript *GN* denotes the Gauss-Newton method. The advantage here, of course, is that only the 1st derivatives are needed. However, if the residuals are large, the neglected term can become dominant, and then the full equations need to be solved.

In practice, the Jacobian is often poorly determined. This means that both very big and very small eigenvalues are present. If we move in the direction of the eigenvectors that correspond to the big eigenvalues, we can neglect the *B* matrix, or residuals. If we follow the small eigenvalues direction, then we should not neglect the residuals. On the other hand, the norm of *B* decreases as we approach the minimum, therefore, the Gauss-Newton technique will become more and more efficient. This is the basis for our proposed method. We propose that there is an optimal strategy, which will initially move along the directions for which *B* is not significant, until it reaches the area of small residuals, where *B* can then be disregarded.

If it is not possible to create such a strategy, we should try to make the area, in which the full system has to be solved, as small as possible. We also have to make sure that the dimensionality of the subspace, in which the components of *B* have to be taken into account, is as small as possible.

As just stated we propose an optimal strategy of searching along the directions for which *B* is not significant. This raises some important questions: How do you determine what direction to search, and when the Hessian becomes degenerate, and the SVD becomes ineffective, what modification is necessary? The answers to these questions are addressed in the next two sections. We use singular value decomposition (SVD) to determine the direction to search and propose a modification of the least squares method when the SVD is ineffective.

Singular Value Decomposition

For the SVD introduce

$$\begin{aligned} &A = USV^T \\ &U^T U = V^T V = I \\ &S = diag\left(s_1, s_2, \ldots, s_n\right) \end{aligned} \tag{10}$$

where U consists of the n orthonormal eigenvectors corresponding to the n largest eigenvalues of AA^T, and V is the matrix of orthonormal eigenvectors vectors of the matrix $A^T A$. The diagonal elements of S are the square roots of the non-negative values of the eigenvalues of $A^T A$. The s_i are called the singular values. Now introduce the vectors

$$\mathbf{g} = U^T \mathbf{b}, \mathbf{y} = V^T \mathbf{x} \tag{11}$$

then Eq. (9) reduces to

$$S\mathbf{y} = \mathbf{g} \tag{12}$$

Since S is a diagonal matrix, the influence of each component of $\mathbf{y}$ can be observed immediately. Introducing the component $\tilde{y}_j = g_j / s_j$ into the solution, we reduce the square of the norm of the residual by g_j^2. Now, let the singular values be in descending order and consider the "probe" solutions

$$\mathbf{y}^{(k)} = \left(\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_k, 0, ..., 0\right)^T \tag{13}$$

The probe vector $\mathbf{y}^{(k)}$ is the normal pseudo-solution of the least-squares problem, if we disregard the singular values $s_j, j > k$ and consider them equal to zero. From the probe vectors we obtain the probe solutions vector $\mathbf{x}^{(k)}$ from

$$\mathbf{x}^{(k)} = V\mathbf{y}^{(k)} = \sum_{j=1}^{k} \tilde{y}_j v^{(j)} \tag{14}$$

where $v^{(j)}$ is the j-th column of V. The corresponding square of the norm of the residual is

$$\rho_k^2 = \left\| \mathbf{b} - A\mathbf{x}^{(k)} \right\| = \sum_{j=k+1}^{m} g_j^2 \tag{15}$$

Now assume that A is poorly defined, that is, some of the singular values are widely separated. The corresponding $\mathbf{y}^{(k)}$ may be too large due to the small singular values. Thus, one needs to find an index k such that the norm of the probe vector and the norm of the residual for this probe solution are small enough. With the singular values in descending order the procedure is:

1. Develop the matrix of trial vectors.

$$\mathbf{x}^{(k)} = \sum_{j=1}^{k} V^{(j)} \frac{g_j}{w_j} \tag{16}$$

where $V^{(j)}$ is the j-th column of V.

2. For each trial vector compute the expected decrease of the least squares error function using

$$\rho_k^2 = \sum_{j=k+1}^{n} g_j^2 \tag{17}$$

3. For $k = 1,...,n$ check the acceptability of the trial vectors. To accomplish this, check each of their components and determine if the following inequality is satisfied.

$$\left| x_j^{(k)} \right| < \Delta c_j \tag{18}$$

That is, determine if the changes in each of the elements are less than some prescribed amount. If the inequality, eq. (18), is not satisfied for some *j,* calculate the required coefficient of decreasing step size

$$d_j^f = \frac{\Delta c_j}{\left| x_j^{(k)} \right|} \tag{19}$$

After checking all the components set

$$d_{\min} = \min_j \left(d_j^f \right) \tag{20}$$

If the inequality is satisfied for all components go to the next step. If not, the trial vector $\mathbf{x}^{(k)}$ is normalized by multiplying it by $d_{\min}$.

4. Now check the relative decrease of the SVD method as we go to the next trial vector. If

$$\left[\frac{\left(\rho_{k-1}^2 - \rho_k^2 \right)}{\rho_{k-1}^2} \right] > C_\rho \tag{21}$$

then the trial vector $\mathbf{x}^{(k)}$ is taken as the next iteration.

5. If the inequality in Step 4 is not satisfied, then the previous trial vector is used.

6. After computing the least squares function with $\mathbf{x}^{(k)}$ determine, if the SVD method is converging sufficiently. Determine if

$$\left[\frac{\left(F_{k-1} - F_k \right)}{F_{k-1}} \right] > C_{\Delta F} \tag{22}$$

If this inequality is satisfied, go to the next step. If it is not satisfied, then the modified least squares method is used, because the Hessian is degenerate and the residuals need to be considered.

Modified Least Squares

The use of the Modified Least squares method uses eq. (8), but with B not equal to zero. The challenge is obtaining B. There are three methods for obtaining or approximating B. The first is analytical. Theoretically, if the gradient is obtained analytically, then the matrix of 2nd partials can be obtained also. However, it is quite complex algebraically with the orbit theory used. The 2nd approach is to use finite differences, and the 3rd is to estimate B using a quasi-Newton scheme. We used the 2nd approach, finite differences. The approach, regardless of the method used to obtain B, is now described.

From eq. (8) we wish to determine the vector $\mathbf{p}_N$ which satisfies

$$\left(A^T A + B\right)\mathbf{p}_N = A^T\mathbf{b} \tag{23}$$

Using the singular-value decomposition of A given previously gives

$$\left(VS^2V^T + B\right)\mathbf{p}_N = V\begin{bmatrix}S & 0\end{bmatrix}U^T\mathbf{b} \tag{24}$$

Pre-multiplying by $\boldsymbol{V^T}$ gives

$$\left(S^2V^T + V^TB\right)\mathbf{p}_N = \begin{bmatrix}S & 0\end{bmatrix}U^T\mathbf{b} \tag{25}$$

The vector $\mathbf{p}_N$ can be written as a linear combination of the n-linearly independent columns of $\boldsymbol{V}$, with

$$\mathbf{p}_N = V\mathbf{z} \tag{26}$$

Substituting for $\mathbf{p}_N$ in eq. (25) and performing some elementary rearrangement gives

$$\left(S^2 + V^TBV\right)\mathbf{z} = \begin{bmatrix}S & 0\end{bmatrix}U^T\mathbf{b} \tag{27}$$

with $\boldsymbol{p_N}$ being recovered from (1.27). The vector $\boldsymbol{z}$ can be found by determining the LDL^T factorization of $\left(S^2 + V^TBV\right)$, but this method proves unsatisfactory for repeated application unless $\left(S^2 + V^TBV\right)$is well-conditioned. It is often the case with least square problems, especially those derived from data-fitting applications, that each of the b_i (and hence the matrix B) is small. Furthermore, these problems are such that A is often by nature ill-conditioned, and this ill-conditioning must be reflected in the matrix $\left(S^2 + V^TBV\right)$, when $\|B\|$ is small compared to $\left\|A^TA\right\|$.

These numerical problems can be avoided by computing $\left(S^2 + V^T BV\right)$ as a sum of two components, the first in the subspace spanned by the columns of $\boldsymbol{V}$ corresponding to the larger singular values, and the second in the subspace spanned by columns of $\boldsymbol{V}$ associated with the smaller singular values.

Consider a partition of the diagonal matrix S such that $S_1 = diag\left(s_1,...,s_r\right)$ and $S_2 = diag\left(s_{r+1},...,s_n\right)$. We shall define $\boldsymbol{r}$ as the ***grade*** of the matrix A. The criterion for fixing the grade of $\boldsymbol{A}$ will depend upon the progress of the optimization algorithm. The value of $\boldsymbol{r}$ is generally equal to the number of dominant singular values of A and is rarely significantly less than $\boldsymbol{n}$. For the moment we require only that $\boldsymbol{S_1}$ be nonsingular. Unless $\left\|A^T A\right\| = 0$ a choice of $r \geq 1$ is always possible,since $s_1 > 0$. We shall define the zero grade of A as $S_1 = 0, S_2 = diag\left(s_1,...,s_n\right)$.
The partition of $\boldsymbol{S}$ into S_1 and S_2 implies a similar partition of $\boldsymbol{V}$ such that

$$V = \left[V_1 \quad V_2\right] \tag{28}$$

where V_1 is an n x r matrix and V_2 an n x (n -r) matrix, we can write

$$\mathbf{p}_N = V_1 \mathbf{w}_N + v_2 \mathbf{y}_N \tag{29}$$

where $\mathbf{w}_N$ is an r vector and $\mathbf{y}_N$ is an (n -r) vector. Substituting for $\mathbf{p}_N$ in eq. (27) and using the fact that

$$V^T V_1 = \begin{bmatrix} I_r \\ 0 \end{bmatrix}, V^T V_2 = \begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix} \tag{30}$$

gives

$$S^2 \begin{bmatrix} I_r \\ 0 \end{bmatrix} \mathbf{w}_N + V^T BV_1 \mathbf{w}_N + S^2 \begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix} \mathbf{y}_N + V^T BV_2 \mathbf{y}_N = \left[S \quad 0\right] U^T \mathbf{b} \tag{31}$$

From this equation the following two systems can be obtained

$$\begin{aligned} S_1^2 \mathbf{w}_N + V_1^T BV_1 \mathbf{w}_N + V_1^T BV_2 \mathbf{y}_N &= S_1 \mathbf{b}_i \\ S_2^2 \mathbf{y}_N + V_2^T BV_2 \mathbf{y}_N + V_2^T BV_1 \mathbf{W}_N &= S_2 \mathbf{b}_2 \end{aligned} \tag{32}$$

where $\mathbf{b}_1, \mathbf{b}_2$ and $\mathbf{b}$ are r, (n -r) and (m-n)-vectors respectively such that $\mathbf{b}^T U = \left[\mathbf{b}_1^T \quad \mathbf{b}_2^T \quad \mathbf{b}^T\right]$.

The significance of these equations is best illustrated by substituting $B = \varepsilon \bar{B}$ where $\|\bar{B}\| = 1$ and ε is a scalar, and writing $\mathbf{p}_N = V_1 \mathbf{w}_N + V_2 \mathbf{y}_N$ in eq. (32). Thus

$$\begin{aligned} S_1^2 \mathbf{w}_n + \varepsilon V_1^T \bar{B} V_1 \mathbf{p}_N &= S_1 \mathbf{b}_1 \\ S_2^2 \mathbf{y}_N + \varepsilon V_2^T \bar{B} V_2 \mathbf{y}_N + \varepsilon V_2^T \bar{B} V_1 \mathbf{w}_N &= S_2 \mathbf{b}_2 \end{aligned} \tag{33}$$

An approximation $\bar{\mathbf{w}}_N$ to $\mathbf{w}_N$ can be obtained by neglecting terms of order ε in the first set of equations. This gives $\bar{\mathbf{w}}_N = S_1^{-1} \mathbf{b}_1$. If $\bar{\mathbf{w}}_N$ is substituted for $\mathbf{w}_N$ in the second set of equations, we obtain

$$\left(S_2^2 + \varepsilon V_2^T B V_2\right) \bar{\mathbf{y}}_N = S_2 \mathbf{b}_2 - \varepsilon V_2^T \bar{B} V_1 \bar{\mathbf{w}}_n \tag{34}$$

This idea can be extended to give the following iterative scheme for $\mathbf{p}_N$.

Iterative Algorithm for $\mathbf{p}_N$:

(i) Define $\bar{\mathbf{p}}_N = 0$;

(ii) for j =0, 1,2, . . . ,compute $\bar{\mathbf{w}}_N^{(j+1)}$ and $\bar{\mathbf{y}}_N^{(j+1)}$ such that

$$\begin{aligned} S_1^2 \bar{\mathbf{w}}_N^{(j+1)} &= S_1 \mathbf{b}_1 - V_1^T B \bar{\mathbf{p}}_N^{(j)} \\ \left(S_2^2 + V_2^T B V_2\right) \bar{\mathbf{y}}_N^{(j+1)} &= S_2 \mathbf{b}_2 - V_2^T B V_1 \bar{\mathbf{w}}_N^{(j+1)} \end{aligned} \tag{35}$$

and set

$$\bar{\mathbf{p}}_1^{(j+1)} = V_1 \bar{\mathbf{w}}_N^{(j+1)}, \mathbf{p}_2^{(j+1)} = V_2 \bar{\mathbf{y}}_N^{(j+1)}, \mathbf{p}_N^{(j+1)} = \bar{\mathbf{p}}_1^{(j+1)} + \bar{\mathbf{p}}_2^{(j+1)} \tag{36}$$

Since, by assumption, the matrix $\left(A^T A + B\right)$is positive definite, so is the matrix $\left(S_2^2 + V_2^T B V_2\right)$. Furthermore, $\left(S_2^2 + V_2^T B V_2\right)$ is not ill-conditioned because S_2 does not contain the large singular values of $\boldsymbol{S}$. Accordingly the LDL^T factorization is an efficient and numerically-stable method for solving the latter of eqs. (35) for $\bar{\mathbf{y}}_N^{(j+1)}$.

The following theorem gives some indication of the convergence of this approach.

Theorem

If $\bar{\mathbf{p}}_N^{(j)}$ and $\bar{\mathbf{p}}_N^{(j+1)}$ are two approximations of $\mathbf{p}_N$ obtained using the iterative scheme just described, then their relative errors satisfy the inequality

$$\frac{\left\|\mathbf{p}_n - \bar{\mathbf{p}}_N^{(j+1)}\right\|}{\left\|\mathbf{p}_n\right\|} \le \varepsilon\sigma\left(1+\varepsilon\eta\right)\frac{\left\|\mathbf{p}_n - \bar{\mathbf{p}}_N^{(j)}\right\|}{\left\|\mathbf{p}_n\right\|}$$

where (37)

$$\varepsilon = \|B\|, \sigma = \left\|S_1^{-2}\right\|, \eta = \left\|\left(S_2^2 + V_2^T B V_2\right)^{-1}\right\|$$

The proof is not provided, but the theorem shows that the iterative scheme converges, if $\varepsilon\sigma\left(1+\varepsilon\eta\right)<1$. If any of the singular values of A are less than $\varepsilon^{-1/2}$, then $\eta = O(1/\varepsilon)$ and we essentially require the partition of S to be such that $\varepsilon\sigma<1$. If we allow $S_1 = 0$ and $S_2 = S$ in our set of allowable partitions, there must exist a partition for which the iterative scheme given by eq. (35) converges since, if $V_2 = V$, the computation of $\mathbf{p}_N$ trivially reduces to the solution of the equations (1.28). For a given partition, if it is evident from the evaluation of the sequence $\left\{\bar{\mathbf{p}}_N^{(j)}\right\}$ that it is not converging to $\mathbf{p}_N$ sufficiently quickly, more elements of S_1 can be included in S_2 and the iterative scheme restarted. This automatically ensures that $\left\|S_1^{-2}\right\|$ is reduced (since the singular values of A are arranged in order of decreasing magnitude), but makes eq. (35) less well-conditioned.

RESULTS

To demonstrate the performance of the SVD algorithm, it is applied to the orbit determination of 24-hour satellites with angles only. The orbit theory is the semi-analytic theory used in the Russian Space Surveillance Center[3] (RSSC). It uses the following set of nonsingular variables

$$\begin{aligned}
&\lambda = M + \omega + \Omega,\ L = \sqrt{\mu a} \\
&p = \sin\left(i/2\right)\cos\Omega, q = \sin\left(i/2\right)\sin\Omega \\
&\tilde{h} = e\sin\left(\omega+\Omega\right), \tilde{k} = e\cos\left(\omega+\Omega\right)
\end{aligned} \tag{38}$$

The Δc_j parameters used were

$$\begin{aligned}
&\Delta c_\lambda = 0.3 \\
&\Delta c_L = 1000 \\
&\Delta c_p = \Delta c_q = \Delta c_{\tilde{h}} = \Delta c_{\tilde{k}} = 0.2
\end{aligned} \tag{39}$$

Figures 1-4 compare the residuals using the SVD algorithm to those from the least squares method used in the RSSC for four different orbit determinations. In these figures 1985 denotes the standard least squares method used at the time in the RSSC (the green and pink curves) and 2005 indicates the SVD algorithm (blue and black curves). Right ascension and declination are indicated by α and δ, respectively. These plots show the superior performance of the SVD algorithm for cases with degenerate Hessians.

To demonstrate the convergence performance of the SVD algorithm, four examples are presented. Table 1 shows the number of observations and the time span of observations for each of the four cases. The initial orbital parameters of the object obtained using Laplace's method are given in Table 2.

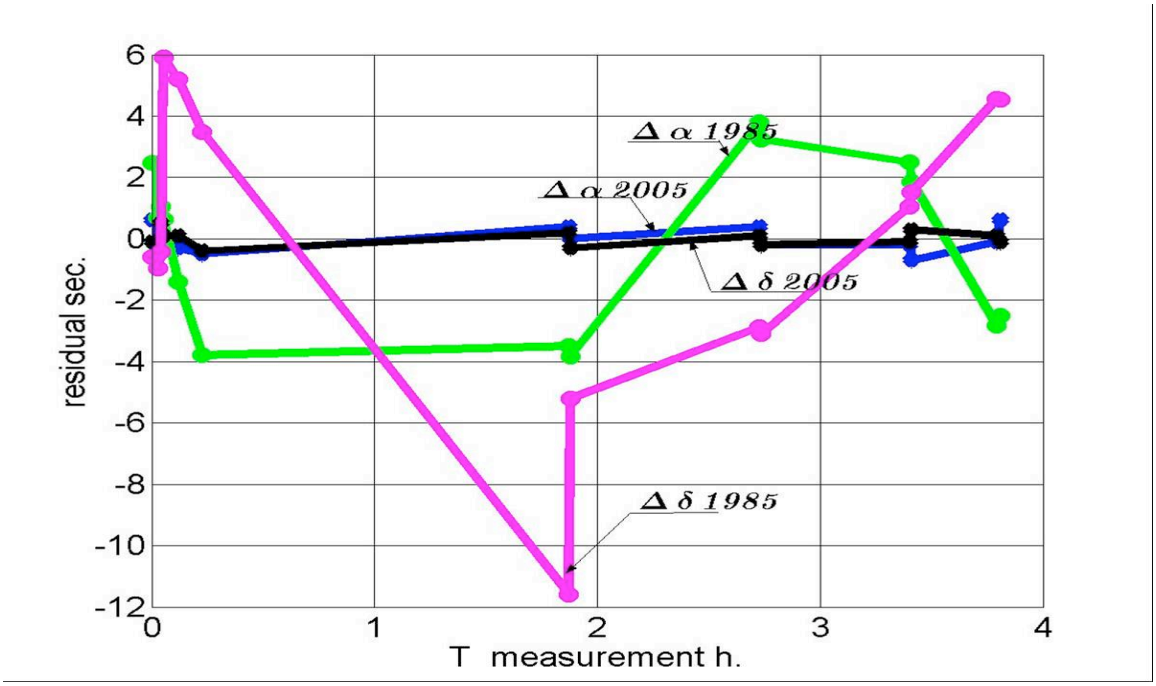

Figure 1 Residual Comparison – Case 1

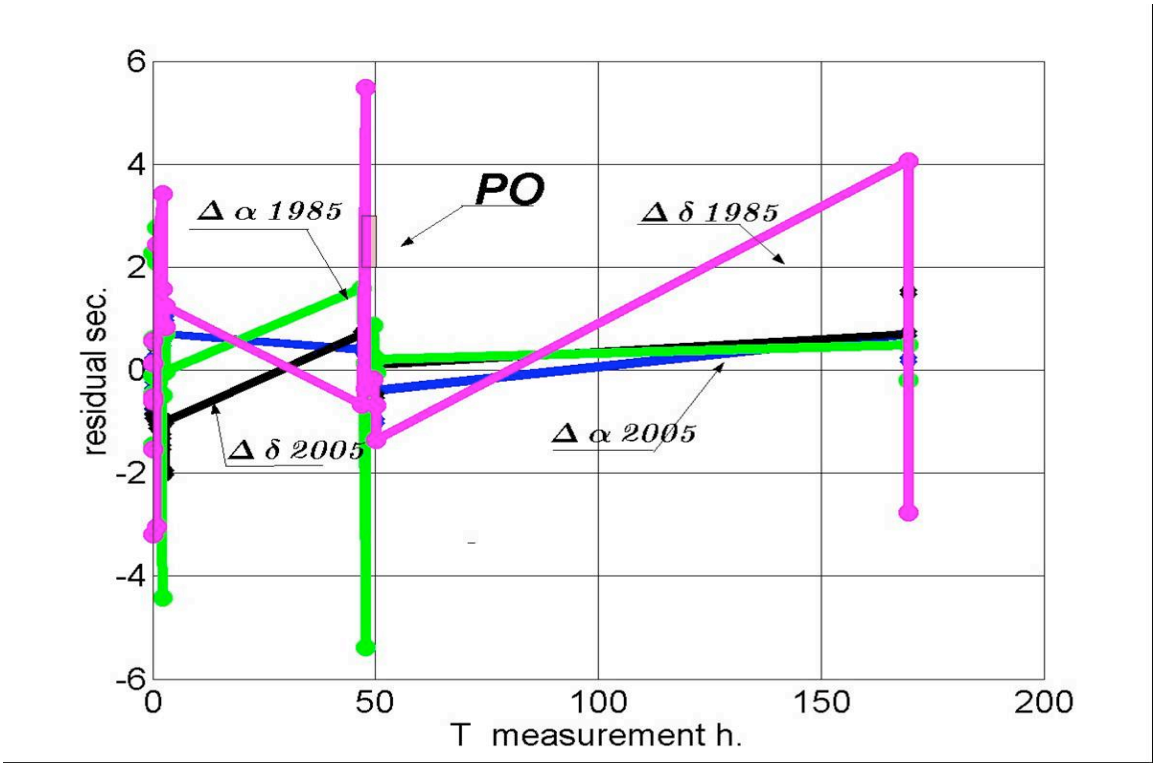

Figure 2 Residual Comparison – Case 2

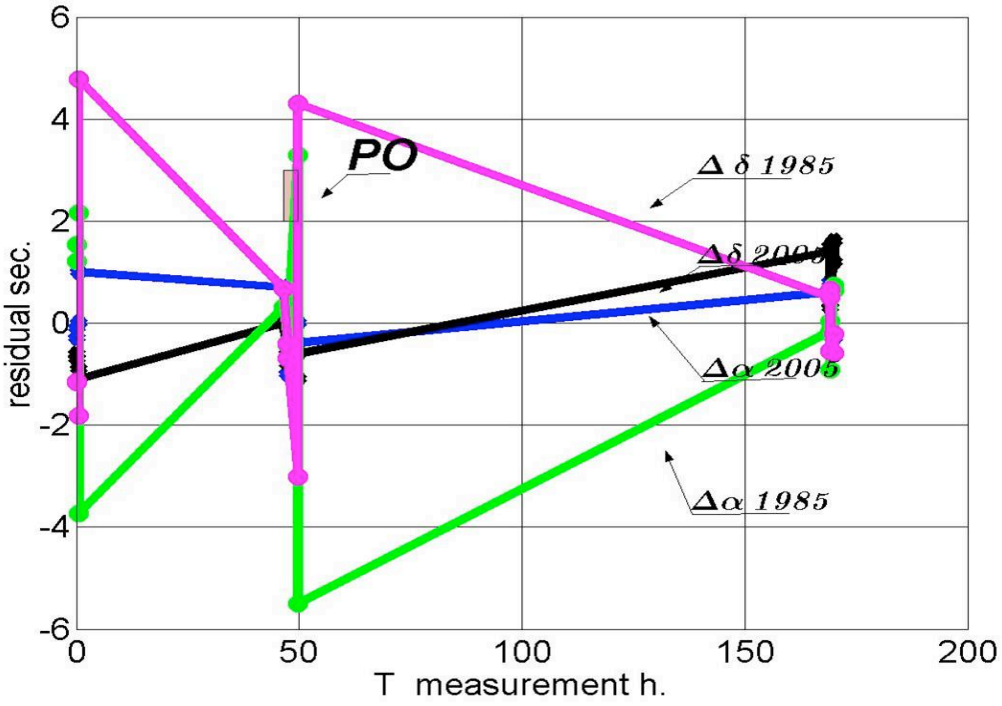

Figure 3 Residual Comparison – Case 3

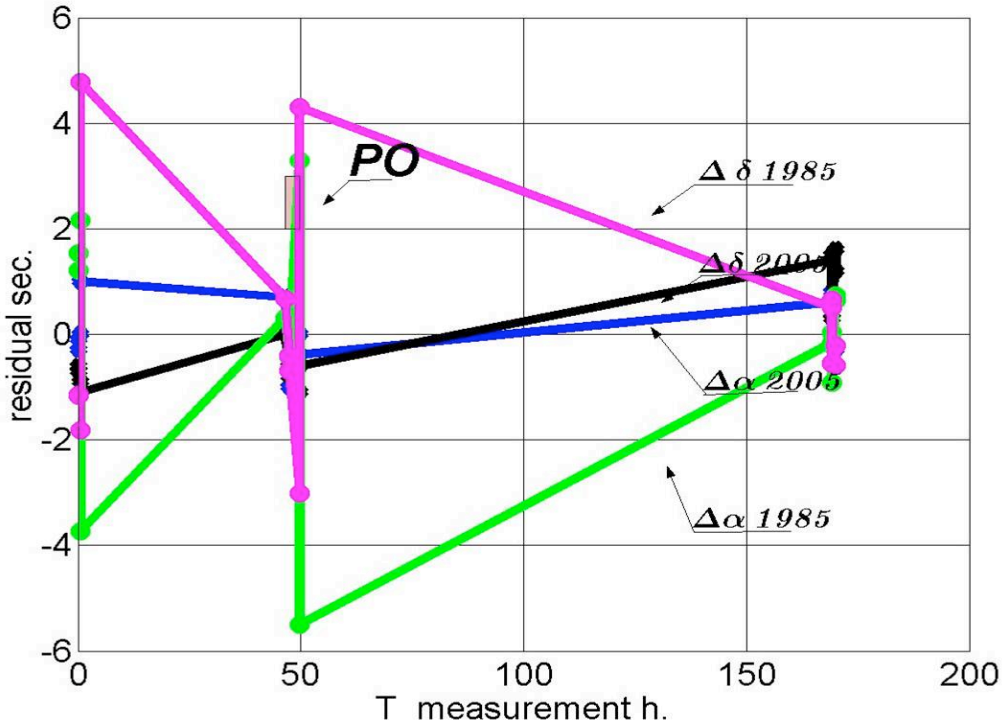

Figure 4 Residual Comparison – Case 4

Table 3 shows the dimension of the subspace of the SVD algorithm as a function of iteration. The function minimization process, as a function of iteration number, is shown in Figs. 5-8. Also provided are the function values. In each example the initial value of the function is $O(10^{11})$ and it reduces to 1-10. The plots show that as the function value decreases the SVD method loses its ability, and we need to transfer to the corrected/modified least squares method. Then, once a critical point is passed, the SVD works again and at high speed, i.e., rapid convergence. Figure 5 is the example with fast convergence, it rapidly converged in six iterations. The first iteration had a subspace dimension of five, and after that, it was six. Thus, the performance in Fig. 5 would be very similar to that which would occur with the standard least squares method. Figure 6 shows an example of intermediate speed convergence. In this case the minimization slowed down between iterations 3 and 6. This is probably caused by a premature change of the subspace dimensionality from 4 to 5. However, we successfully crossed this area using the SVD method. A similar phenomenon was observed as the dimensionality changed from 4 to 5. Nevertheless, in this case the need is rather to adopt a more careful strategy in increasing the dimensionality of the minimization subspace rather than to use the corrected method.

Figure 7 is a case when the degenerate region is much wider and stretches from the 5th to the 15th iteration. The corrected technique might be beneficial starting at about the 7th iteration, but it was not used. In this example, the subspace dimension starts at one and during the degenerate region it is five. As it starts to converge faster at iteration 11, it transitions to the full dimension of six.

Table 1 Observation Data

Example	1	2	3	4
# of obs	7	5	8	4
Ob Time span (hr:min)	25:22	5:06	25:39	4:21

Table 2 Initial Conditions For the Orbit Determination

Example	Period (min)	Inclination (deg)	Arg. of Per. (deg)	Ascend Node (deg)	Eccentricity
1	1442.78	15.22	117.0	-5.27	0.0046
2	1782.05	9.33	-3.18	48.14	0.1688
3	1436.01	15.21	-122.24	-4.883	0.1029
4	1459.05	11.33	24.45	35.52	0.0090

Table 3 SVD Sub Space Dimension

Example	1	2	3	4
Iter 1	5	1	1	1
2	6	4	1	4
3	6	5	3	4
4	6	5	5	5
5	6	5	5	4
6	6	5	5	5
7		5	5	5
8		6	5	6
9		6	5	6
10		6	5	6
11		6	5	6
12			6	6
13			6	
14			6	
15				

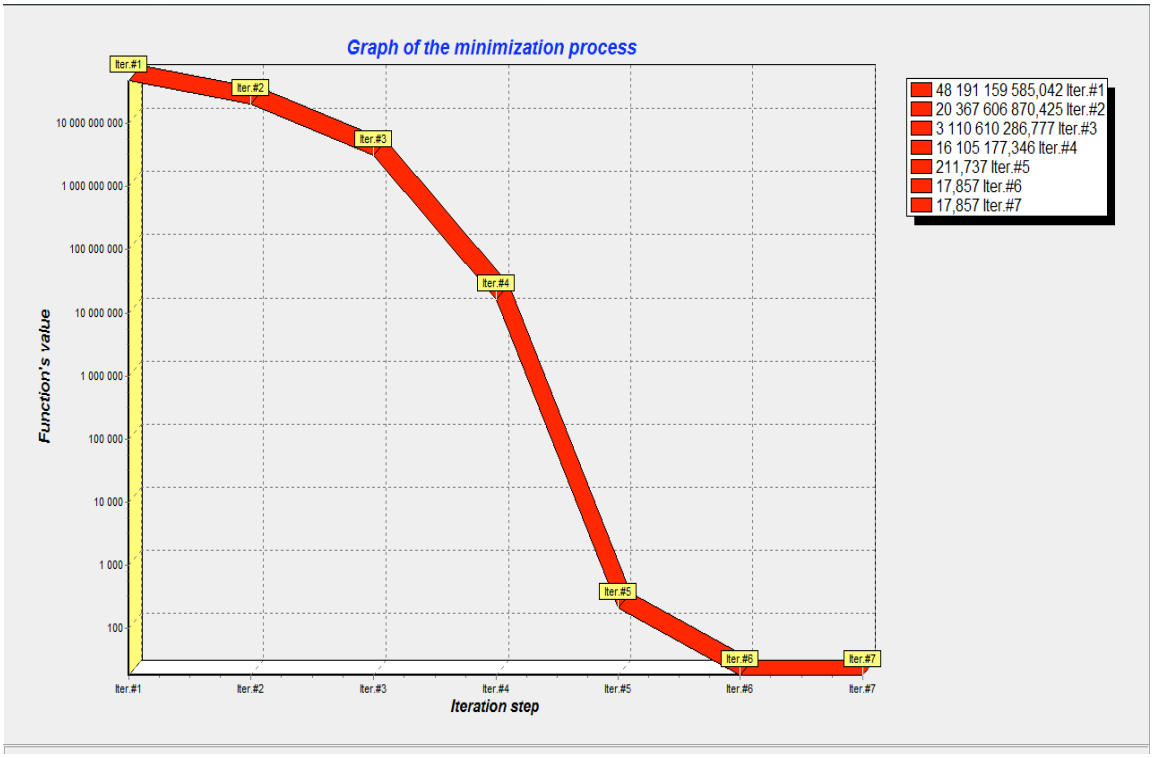

Figure 5 SVD Algorithm Function Minimization Process – Example 1-Fast

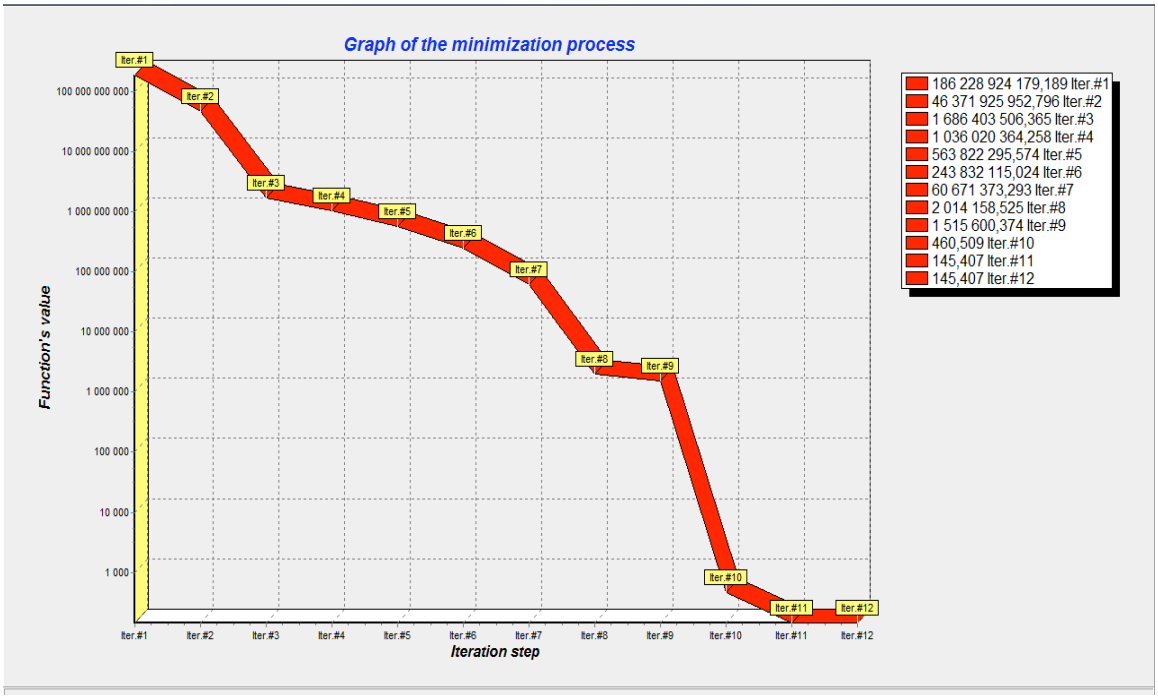

Figure 6 SVD Algorithm Function Minimization Process – Example 2-Intermediate

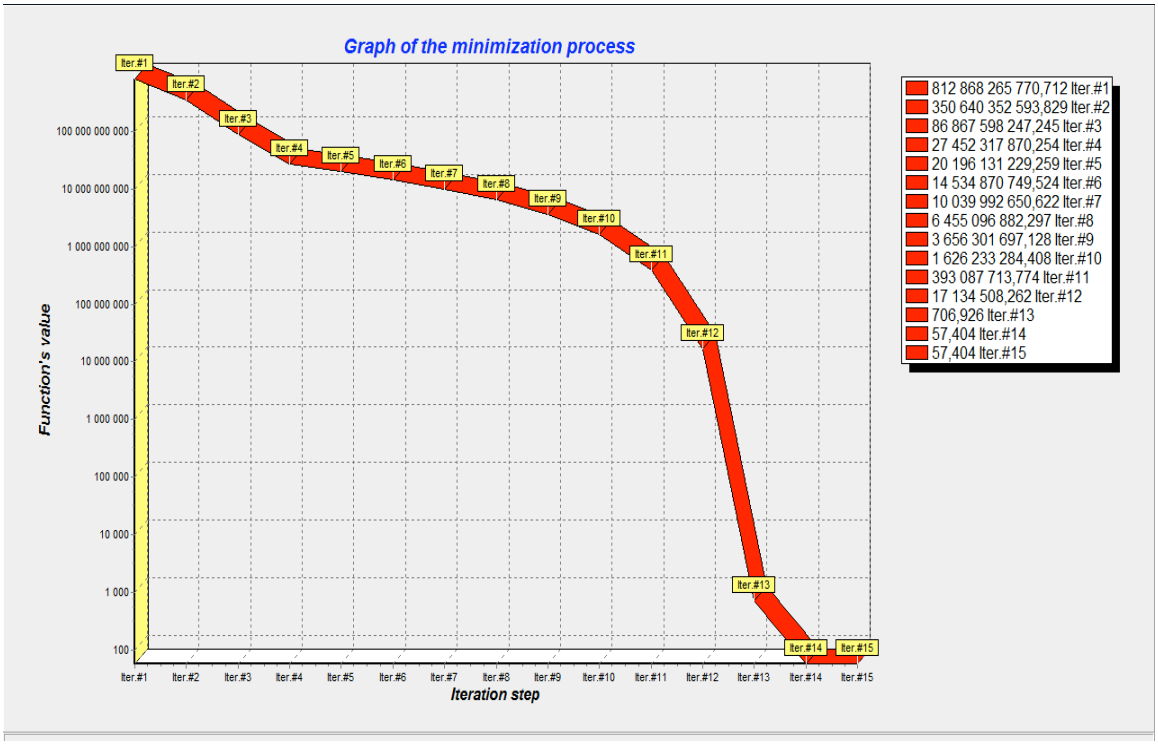

Figure 7 SVD Algorithm Function Minimization Process – Example 3-Slow

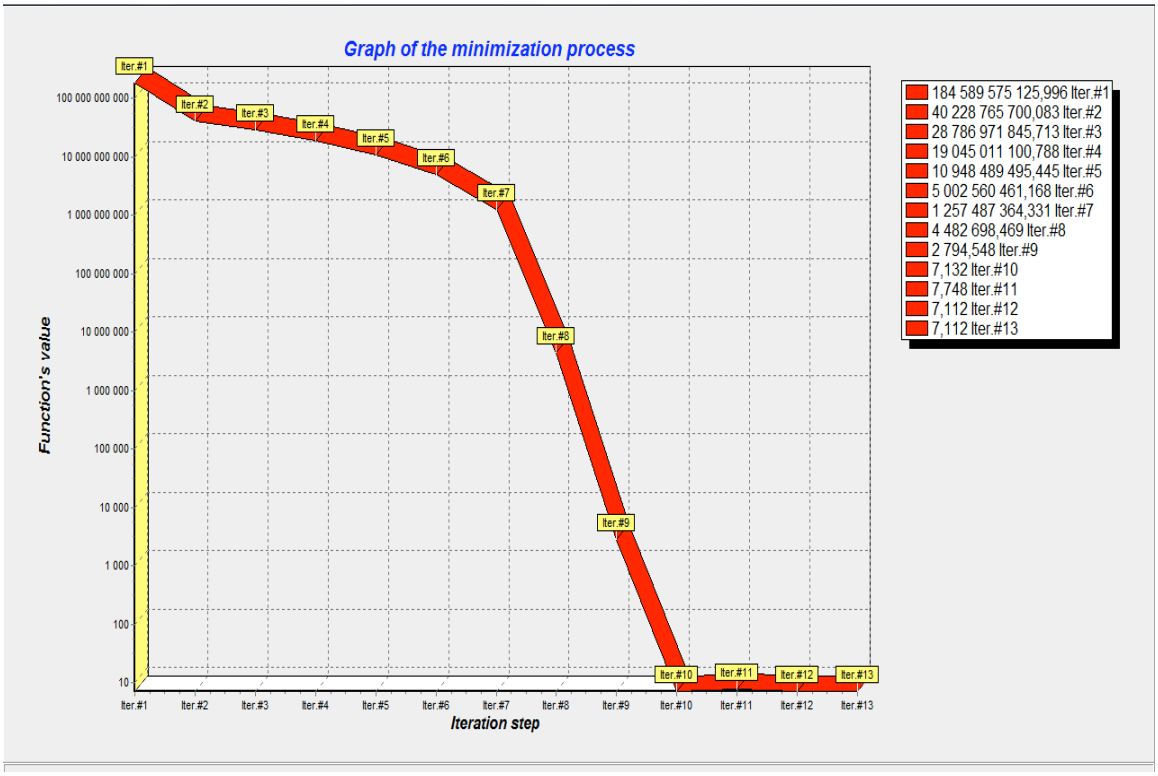

Figure 8 SVD Algorithm Function Minimization Process – Example 4-Slow

Example 4 in Fig. 8 converges faster than example 3, but it also exhibits a degenerate region up to about iteration 7. What is interesting in this example is the fact that at iteration 5 the subspace dimension reduces from five to four and then increases again. This demonstrates the ability of the method to change the dimensionality non-monotonously. This was the only example of the 14 examples investigated in which there was a non-monotonic increase in the behavior of the subspace dimension. Further investigation into this behavior is needed.

In the US there are some orbit determination programs that partially do this in an ad hoc manner. For objects in low Earth orbit (LEO) that are tracked by radar systems, the orbit determination programs have an option to first correct the in-track error, then the orbit plane, and, finally the full state. The Russian Space Surveillance Center does something similar in that their 1st step is a 1-dimensional step.

CONCLUSIONS

A new method of least squares orbit determination utilizing singular value decomposition (SVD) has been presented. Application to high altitude (near 24-hour) satellites with angles only measurements and a trivial a priori state estimate of assuming a semi-major axis equal to that of a geosynchronous satellite, and all other elements zero, showed excellent performance and an improvement over the standard least squares approach. A corrected/modified method of least squares for dealing with situations when the Hessian becomes degenerate was presented. It was not needed in any of the examples presented. The corrected method is essentially a safety net, which should not be needed, if the parameters are for the SVD method are properly chosen.

Future work will focus on determining the length of a step in a given direction, i.e., finding the minimum in a given direction and the application of the method to other orbits. Also, more research will be done on when to transfer from the SVD method to the modified/corrected method and back.

REFERENCES

1. D. A. Vallado, *Fundamentals of Astrodynamics and Applications*, McGraw Hill, 1997.
2. Z. Khutorovsky, S. Kamensky, N. Sbytov and K.T. Alfriend, “Orbit Determination of LEO Satellites For a Single pass Through a Radar: Comparison of Methods,”20th ISSFD, Annapolis, MD, 24-27 September 2007, Annapolis MD.
3. Z. N. Khutorovsky, “Monitoring of Geostationary Earth Orbit Satellites in Russian Space Surveillance Center,” *AIAA JGCD*, Vol. 25, No. 5, Sept.-oct. 2002, pp. 953-961.
4. P.E. Gill, W. Murray and R.A. Pitfield, “The Implementation of Two Revised Quasi-Newton Algorithms for Unconstrained Optimization*”, Rep. NAC 11*, Nat. Physical Lab., Teddington, Middlesex, England. (1972),