(Preprint) AAS 19-XXX

A FRAMEWORK FOR SCALING IN FILTERING AND LINEAR COVARIANCE ANALYSIS

Christopher D'Souza,[*] Renato Zanetti,[†] and David Woffinden[‡]

Scaling is used extensively for numerical optimization and trajectory optimization. Its use in the estimation community is almost nonexistent. This paper creates the framework for practical scaling in space navigation, in general, and linear covariance analysis, in particular.

INTRODUCTION

It has been known since Apollo that changing the units of elements of the state-space can have a favorable affect on the accuracy of state estimates. Because of word-length limitations in the Apollo era, the units of position was chosen to be 'Earth-radii'. However, it was believed that with 32-bit and 64-bit processors, the need for such techniques had vanished. In fact, it was believed that the need for square-root filters had gone the way of slide-rules. However, recent analysis on the EM-1 and EM-2 trajectory have uncovered a need to revisit scaling, at the very least for Linear Covariance analysis. It was observed that even with a square-root implementation in Lincov, that numeric issues arose which introduced unexpected and initially unexplained behavior in the covariance matrices. In light of this, we began an investigation into whether these numeric issues could be avoided. We took a page from the optimization community where scaling is routinely practiced to avoid numeric issues; in the optimization community the Hessian is chosen as the focus and source of the scaling parameters.[1,2] We thus began developing the mathematics of scaling as applied to navigation practice.

In Section 2, the motivation behind scaling is presented. Section 3 contains the mathematics of scaling as applied to Kalman Filtering. Section 4 extends these mathematics to Linear Covariance Analysis. Section 5 treats scaling in batch filtering. Section 6 addresses the question of how/what scaling to choose. Section 7 contains a discussion of scaling philosophies. Section 8 contains an example from the EM-1 trajectory and the results of scaling are compared with those when scaling is not performed. Finally, Section 9 contains a few concluding comments.

THE MOTIVATION

One major source of numeric issues in Kalman filtering is ill-conditioning of the covariance matrix, which is required to be at least positive semidefinite. This is particularly acute in LinCov where the covariance matrices often exhibit negative eigenvalues. One measure of the ill-conditioning of

[*] Aerospace Engineer, EG6, NASA JSC, 2101 NASA Parkway, Houston, TX 77058.
[†] Assistant Professor, Department of Aerospace Engineering and Engineering Mechanics, University of Texas, Austin, TX 7XXX
[‡] Aerospace Engineer, EG6, NASA JSC, 2101 NASA Parkway, Houston, TX 77058.

the covariance matrix is the condition number. The condition number of a matrix is defined as the ratio of the largest to the smallest singular value. For a square, symmetric matrix $\mathbf{P}$, this is equivalent to the ratio of largest to the smallest eigenvalue, or

$$\text{cond}(\mathbf{P}) \triangleq \frac{|\lambda_{max}(\mathbf{P})|}{|\lambda_{min}(\mathbf{P})|} \tag{1}$$

A well conditioned covariance matrix has a small condition number and the absolute minimum a condition number could be is 1.

Normally, in LinCov, the covariance matrices have condition numbers approaching 10^{14}, and this results in a loss of precision (in the benign case), or worse, in negative eigenvalues. Even the use of square root factorization methods ($\mathbf{P} = \mathbf{S}\mathbf{S}^T$) does not solve this problem since

$$\text{cond}(\mathbf{S}) = \sqrt{\text{cond}(\mathbf{P})} \tag{2}$$

and if the condition number of the (full) covariance matrix is, say, on the order of 10^{14}, the condition number of the (square-root) factorized covariance matrix is merely 10^7, an improvement but worrisome nonetheless.

The UDU factorization doesn't improve this situation at all because the $\mathbf{D}$ matrix has the same condition number as the original. The situation with the UDU factorization is somewhat ameliorated because the $\mathbf{U}$ matrix has a lower condition number; however the $\mathbf{D}$ matrix still retains a high condition number.

This has exposed the need for a scaling methodology that improves the numeric issues associated with the condition number of the covariance matrix by transforming the problem into one which has better numerics.

THE MATHEMATICS OF SCALING IN KALMAN FILTERING

Given a $n \times 1$ state $\mathbf{x}$, the scaled state, $\mathbf{x}_s$ (which is also of dimension n) is defined as

$$\mathbf{x}_s = \mathbf{M}_s \mathbf{x} \iff \mathbf{x} = \mathbf{M}_s^{-1} \mathbf{x}_s \iff \frac{\partial \mathbf{x}_s}{\partial \mathbf{x}} = \mathbf{M}_s \tag{3}$$

where the non-singular $n \times n$ matrix $\mathbf{M}_s$ is the *scaling matrix*. The variation of $\mathbf{x}$ thus becomes

$$\delta\mathbf{x} = \mathbf{M}_s^{-1} \delta\mathbf{x}_s \tag{4}$$

and the covariance $\mathbf{P}_s$ is*

$$\mathbf{P} = \mathbf{M}_s^{-1} \mathbf{P}_s \mathbf{M}_s^{-T} \tag{5}$$

or

$$\mathbf{P}_s = \mathbf{M}_s \mathbf{P} \mathbf{M}_s^T \tag{6}$$

*That this is the case can be seen as follows:

$$\mathbf{P} = E[\delta\mathbf{x}\delta\mathbf{x}^T] = E[\mathbf{M}_s^{-1} \delta\mathbf{x}_s\delta\mathbf{x}_s^T\mathbf{M}_s^{-T}] = \mathbf{M}_s^{-1}E[\delta\mathbf{x}_s\delta\mathbf{x}_s^T]\mathbf{M}_s^{-T} = \mathbf{M}_s^{-1} \mathbf{P}_s \mathbf{M}_s^{-T}$$

Given an affine nonlinear system of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \mathbf{B}(t)\mathbf{w} \tag{7}$$

the perturbed system can be expressed as

$$\delta\dot{\mathbf{x}} = \mathbf{A}(t)\delta\mathbf{x} + \mathbf{B}(t)\mathbf{w} \tag{8}$$

where $\mathbf{A}(t)$ and $\delta\mathbf{x}$ are

$$A(t) = \left.\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}} \tag{9}$$

$$\delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}} \tag{10}$$

In terms of $\delta\mathbf{x}_s$ we find that Eq. (8) is

$$\delta\dot{\mathbf{x}}_s = \mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1}\delta\mathbf{x}_s + \mathbf{M}_s\mathbf{B}(t)\mathbf{w} \tag{11}$$

so that we get

$$\delta\dot{\mathbf{x}}_s = \mathbf{A}_s(t)\delta\mathbf{x}_s + \mathbf{B}_s(t)\mathbf{w} \tag{12}$$

where $\mathbf{A}_s(t)$ and $\mathbf{B}_s(t)$ are

$$\mathbf{A}_s(t) = \mathbf{M}_s\,\mathbf{A}(t)\,\mathbf{M}_s^{-1} \tag{13}$$

$$\mathbf{B}_s(t) = \mathbf{M}_s\,\mathbf{B}(t) \tag{14}$$

Given $\mathbf{\Phi}(t, t_0)$ which satisfies

$$\dot{\mathbf{\Phi}}(t, t_0) = \mathbf{A}(t)\mathbf{\Phi}(t, t_0), \qquad \mathbf{\Phi}(t, t_0) = I \tag{15}$$

we can surmise that there is a $\mathbf{\Phi}_s(t, t_0)$ which satisfies

$$\dot{\mathbf{\Phi}}_s(t, t_0) = \mathbf{A}_s(t)\mathbf{\Phi}_s(t, t_0), \qquad \mathbf{\Phi}_s(t, t_0) = I \tag{16}$$

so that

$$\mathbf{\Phi}_s(t, t_0) = \mathbf{M}_s\mathbf{\Phi}(t, t_0)\mathbf{M}_s^{-1} \tag{17}$$

To see that this is the case, begin with the Taylor series approximation to $\mathbf{\Phi}(t, t_0)$ as

$$\mathbf{\Phi}(t, t_0) = \mathbf{I} + \mathbf{A}(t)(t - t_0) + \frac{1}{2!}\mathbf{A}^2(t)(t - t_0)^2 + \frac{1}{3!}\mathbf{A}^3(t)(t - t_0)^3 + \cdots \tag{18}$$

and premultiply the above by $\mathbf{M}_s$ and postmultiply by $\mathbf{M}_s^{-1}$ to get

$$\begin{aligned}
\mathbf{M}_s\mathbf{\Phi}(t, t_0)\mathbf{M}_s^{-1} &= \mathbf{I} + \mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0) + \frac{1}{2!}\mathbf{M}_s\mathbf{A}(t) \cdot \mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0)^2 \\
&\quad + \frac{1}{3!}\mathbf{M}_s\mathbf{A}(t) \cdot \mathbf{A}(t) \cdot \mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0)^3 + \cdots \\
&= \mathbf{I} + \mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0) + \frac{1}{2!}\mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1}\mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0)^2 \\
&\quad + \frac{1}{3!}\mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1}\mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1}\mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1} \cdot (t - t_0)^3 + \cdots \\
&= \mathbf{I} + \mathbf{A}_s(t)(t - t_0) + \frac{1}{2!}\mathbf{A}_s^2(t)(t - t_0)^2 + \frac{1}{3!}\mathbf{A}_s^3(t)(t - t_0)^3 + \cdots \\
&= \mathbf{\Phi}_s(t, t_0)
\end{aligned} \tag{19}$$

Recall that the covariance propagation equations are

$$\dot{\mathbf{P}}(t) = \mathbf{A}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}^T(t) + \mathbf{B}(t)\mathbf{Q}\,\mathbf{B}^T(t) \tag{20}$$

whose solution is

$$\mathbf{P}(t) = \mathbf{\Phi}(t,t_0)\mathbf{P}(t_0)\mathbf{\Phi}^T(t,t_0) + \overline{\mathbf{Q}}(t) \tag{21}$$

where $\overline{\mathbf{Q}}(t)$ is

$$\overline{\mathbf{Q}}(t) = \int_{t_0}^{t} \mathbf{\Phi}_(\tau,t)\mathbf{B}(\tau)\mathbf{Q}\,\mathbf{B}^T(\tau)\mathbf{\Phi}^T(\tau,t)\,d\tau \tag{22}$$

With this in hand, differentiate Eq. (6) to get

$$\begin{aligned}
\dot{\mathbf{P}}_s(t) &= \mathbf{M}_s\,\dot{\mathbf{P}}(t)\,\mathbf{M}_s^T \\
&= \mathbf{M}_s\mathbf{A}(t)\mathbf{P}(t)\mathbf{M}_s^T + \mathbf{M}_s\mathbf{P}(t)\mathbf{A}^T(t)\mathbf{M}_s^T + \mathbf{M}_s\mathbf{B}(t)\mathbf{Q}\,\mathbf{B}^T(t)\mathbf{M}_s^T \\
&= \mathbf{M}_s\mathbf{A}(t)\mathbf{M}_s^{-1}\mathbf{M}_s\mathbf{P}(t)\mathbf{M}_s^T + \mathbf{M}_s\mathbf{P}(t)\mathbf{M}_s^T\mathbf{M}_s^{-T}\mathbf{A}^T(t)\mathbf{M}_s^T \\
&\quad + \mathbf{M}_s\mathbf{B}(t)\mathbf{Q}\,\mathbf{B}^T(t)\mathbf{M}_s^T \\
&= \mathbf{A}_s(t)\mathbf{P}_s(t) + \mathbf{P}_s(t)\mathbf{A}_s^T(t) + \mathbf{B}_s(t)\mathbf{Q}\,\mathbf{B}_s^T(t)
\end{aligned} \tag{23}$$

The solution of the above equation is

$$\mathbf{P}_s(t) = \mathbf{\Phi}_s(t,t_0)\mathbf{P}_s(t_0)\mathbf{\Phi}_s^T(t,t_0) + \overline{\mathbf{Q}}_s(t) \tag{24}$$

where $\overline{\mathbf{Q}}_s(t)$ is

$$\overline{\mathbf{Q}}_s(t) = \int_{t_0}^{t} \mathbf{\Phi}_s(\tau,t)\mathbf{B}_s(\tau)\mathbf{Q}\,\mathbf{B}_s^T(\tau)\mathbf{\Phi}_s^T(\tau,t)\,d\tau \tag{25}$$

which can be further expressed as

$$\begin{aligned}
\overline{\mathbf{Q}}_s(t) &= \int_{t_0}^{t} \mathbf{M}_s\mathbf{\Phi}(\tau,t)\mathbf{M}_s^{-1}\mathbf{M}_s\mathbf{B}(\tau)\mathbf{Q}\,\mathbf{B}^T(\tau)\mathbf{M}_s^{-T}\mathbf{M}_s^T\mathbf{\Phi}^T(\tau,t)\mathbf{M}_s^T\,d\tau \\
&= \mathbf{M}_s\left(\int_{t_0}^{t} \mathbf{\Phi}(\tau,t)\mathbf{B}(\tau)\mathbf{Q}\,\mathbf{B}^T(\tau)\mathbf{\Phi}^T(\tau,t)\,d\tau\right)\mathbf{M}_s^T \\
&= \mathbf{M}_s\overline{\mathbf{Q}}(t)\mathbf{M}_s^T
\end{aligned} \tag{26}$$

which is consistent with Eq. (6). Thus, the *scaled covariance propagation (time update)* equation takes the form

$$\mathbf{P}_{k_s}^- = \mathbf{\Phi}_s(t_k,t_{k-1})\mathbf{P}_{k-1_s}^+\mathbf{\Phi}_s^T(t_k,t_{k-1}) + \overline{\mathbf{Q}}_{k_s} \tag{27}$$

where

$$\overline{\mathbf{Q}}_{k_s} = \overline{\mathbf{Q}}_s(t_k) \tag{28}$$

For the case of square-root filters (and covariance matrices), the *scaled **square-root** covariance propagation (time update)* equation takes the form

$$\mathbf{S}_{k_s}^- = \mathtt{qr}\left(\left[\ \mathbf{\Phi}_s(t_k,t_{k-1})\mathbf{S}_{k-1_s}^+ \ \vdots\ \mathbf{M}_s\overline{\mathbf{Q}}_k^{1/2}\ \right]\right) \tag{29}$$

The (nonlinear) measurement equation, which is affine in the measurement noise, can be expressed as

$$\mathbf{z}(t) \quad = \quad \mathbf{h}(\mathbf{x}, t) + \boldsymbol{\nu}(t) \tag{30}$$

whose linearized counterpart is

$$\delta\mathbf{z}(t) \quad = \quad \mathbf{H}(t)\delta\mathbf{x} + \boldsymbol{\nu}(t) \tag{31}$$

Proceeding as before we find

$$\delta\mathbf{z}(t) \quad = \quad \mathbf{H}(t)\,\mathbf{M}_s^{-1}\delta\mathbf{x}_s + \boldsymbol{\nu}(t) \tag{32}$$

so that the linearized measurement equation becomes

$$\delta\mathbf{z}(t) \quad = \quad \mathbf{H}_s(t)\delta\mathbf{x}_s + \boldsymbol{\nu}(t) \tag{33}$$

where

$$\mathbf{H}_s(t) \quad \triangleq \quad \mathbf{H}(t)\mathbf{M}_s^{-1} \tag{34}$$

The Kalman Gain at time t_k is defined as

$$\mathbf{K}_k \quad \triangleq \quad \mathbf{P}_k^- \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{35}$$

which can be expressed in terms of the scaled quantities as

$$\mathbf{K}_k \quad = \quad \mathbf{M}_s^{-1}\mathbf{P}_{s_k}^- \mathbf{H}_{s_k}^T \left(\mathbf{H}_{s_k} \mathbf{P}_{s_k}^- \mathbf{H}_{s_k}^T + \mathbf{R}_k\right)^{-1} \tag{36}$$

so that upon defining $\mathbf{K}_{k_s}$ as

$$\mathbf{K}_{k_s} \quad \triangleq \quad \mathbf{M}_s\mathbf{K}_k \tag{37}$$

$$\mathbf{K}_{k_s} \quad = \quad \mathbf{P}_{k_s}^- \mathbf{H}_{k_s}^T \left(\mathbf{H}_{k_s} \mathbf{P}_{k_s}^- \mathbf{H}_{k_s}^T + \mathbf{R}_k\right)^{-1} \tag{38}$$

and recalling that covariance update is

$$\mathbf{P}_k^+ \quad = \quad \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^- \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \tag{39}$$

Upon substituting for $\mathbf{P}_k$ from Eq. (5), for $\mathbf{H}_k$ from Eq. (34) and $\mathbf{K}_k$ from Eq. (37) and simplifying, the scaled covariance update is

$$\mathbf{P}_{k_s}^+ \quad = \quad \left(\mathbf{I} - \mathbf{K}_{k_s}\mathbf{H}_{k_s}\right)\mathbf{P}_{k_s}^- \left(\mathbf{I} - \mathbf{K}_{k_s}\mathbf{H}_{k_s}\right)^T + \mathbf{K}_{k_s}\mathbf{R}_k\mathbf{K}_{k_s}^T \tag{40}$$

Finally, for the scaled state update we begin with the Kalman state update as

$$\delta\mathbf{x}_k^+ \quad = \quad \delta\mathbf{x}_k^- + \mathbf{K}_k\left(\delta\mathbf{z}_k - \mathbf{H}_k\delta\mathbf{x}_k^-\right) \tag{41}$$

Substituting for $\delta\mathbf{x}$ from Eq. (4), for $\mathbf{H}_k$ from Eq. (34) and $\mathbf{K}_k$ from Eq. (37), we get

$$\delta\mathbf{x}_{k_s}^+ \quad = \quad \delta\mathbf{x}_{k_s}^- + \mathbf{K}_{k_s}\left(\delta\mathbf{z}_k - \mathbf{H}_{k_s}\delta\mathbf{x}_{k_s}^-\right) \tag{42}$$

Scaling and the Square Root Filter

We define the square root of the covariance matrix as

$$\mathbf{P} \triangleq \mathbf{S}\,\mathbf{S}^T \quad \text{and} \quad \mathbf{P}_s \triangleq \mathbf{S}_s\,\mathbf{S}_s^T \tag{43}$$

so that the relationship between the scaled and unscaled (square-root) covariance as

$$\mathbf{S} = \mathbf{M}_s^{-1}\,\mathbf{S}_s \tag{44}$$

or

$$\mathbf{S}_s = \mathbf{M}_s\,\mathbf{S} \tag{45}$$

As was stated in Section 2, the condition number of $\mathbf{S}$ is

$$\mathrm{cond}(\mathbf{S}) = \sqrt{\mathrm{cond}(\mathbf{P})} \tag{46}$$

and therefore

$$\mathrm{cond}(\mathbf{S}_s) = \sqrt{\mathrm{cond}(\mathbf{P}_s)} \tag{47}$$

Thus, whereas the very act of formulating the navigation filter in terms of the square root reduces the condition number by the square root of the condition number of the (full) covariance, scaling the square root of the covariance matrix *further reduces* the condition number of the square root of the covariance matrix.

To demonstrate this, given a covariance matrix (obtained from a LinCov analysis EM-1 radiometric-based onboard covariance at MET = 170 hours) whose condition number

$$\mathrm{cond}(\mathbf{P}) = 2.009 \times 10^{20} \tag{48}$$

the condition number of the square root of the covariance matrix is

$$\mathrm{cond}(\mathbf{S}) = 1.417 \times 10^{10} \tag{49}$$

If ‘powers-of-10’ scaling is performed,

$$\mathrm{cond}(\mathbf{P}_s) = 1.327 \times 10^{6} \tag{50}$$

and the condition number of the scaled square root covariance is

$$\mathrm{cond}(\mathbf{S}_s) = 1.152 \times 10^{3} \tag{51}$$

Scaling and the UDU^T Filter

From Eq. (6) we recall that

$$\mathbf{P}_s = \mathbf{M}_s\,\mathbf{P}\,\mathbf{M}_s^T$$

and substituting for $\mathbf{P} = \mathbf{U}\,\mathbf{D}\,\mathbf{U}^T$ we get

$$\mathbf{P}_s \;=\; \mathbf{U}_s\,\mathbf{D}_s\,\mathbf{U}_s^T \;=\; \mathbf{M}_s\,\mathbf{U}\,\mathbf{D}\,\mathbf{U}^T\,\mathbf{M}_s^T \tag{52}$$

Now, inserting $\mathbf{M}_s^{-1}\,\mathbf{M}_s$ (and it's transpose) in the middle of the above equation we get

$$\mathbf{U}_s\,\mathbf{D}_s\,\mathbf{U}_s^T \;=\; \mathbf{M}_s\,\mathbf{U}\,\mathbf{M}_s^{-1}\,\mathbf{M}_s\,\mathbf{D}\,\mathbf{M}_s^T\,\mathbf{M}_s^{-T}\,\mathbf{U}^T\,\mathbf{M}_s^T \tag{53}$$

so that we find that

$$\mathbf{U}_s \;\triangleq\; \mathbf{M}_s\,\mathbf{U}\,\mathbf{M}_s^{-1} \tag{54}$$

$$\mathbf{D}_s \;\triangleq\; \mathbf{M}_s\,\mathbf{D}\,\mathbf{M}_s^T \tag{55}$$

If $\mathbf{M}_s$ is diagonal, we retain the structure of the UDU^T factorization, particularly the upper triangular nature of the $\mathbf{U}$ matrix (with 1's on the main diagonal) and the diagonal structure of the $\mathbf{D}$ matrix. This is the heart of scaling in the UDU^T filter.

One method of scaling that emerges directly from the UDU^T factorization paradigm is to use the $\mathbf{D}$ matrix itself to determine the scaling factors so that the entries of the diagonal $\mathbf{M}_s$ matrix are

$$\mathbf{M}_{s_{ii}} \;=\; \frac{1}{\sqrt{D_{ii}}} \tag{56}$$

If this is done the condition number of $\mathbf{D}_s$ is precisely 1. The condition number of the $\mathbf{U}_s$ matrix is similarly reduced.

In order to demonstrate the efficacy of scaling in the UDU^T filter, as in the square root filter, given a covariance matrix (obtained from a LinCov analysis EM-1 radiometric-based onboard covariance at MET = 170 hours) whose condition number is

$$\mathrm{cond}(\mathbf{P}) \;=\; 2.009 \times 10^{20} \tag{57}$$

the condition number of the $\mathbf{U}$ and $\mathbf{D}$ factors of the covariance matrix is

$$\mathrm{cond}(\mathbf{U}) \;=\; 7.697 \times 10^{16} \tag{58}$$

$$\mathrm{cond}(\mathbf{D}) \;=\; 1.992 \times 10^{16} \tag{59}$$

With the same (diagonal, 'powers-of-ten') scaling as in the previous section, the scaled $\mathbf{U}$ and $\mathbf{D}$ factors of the covariance matrix are

$$\mathrm{cond}(\mathbf{U}_s) \;=\; 8.761 \times 10^{1} \tag{60}$$

$$\mathrm{cond}(\mathbf{D}_s) \;=\; 2.919 \times 10^{4} \tag{61}$$

At the very least, this demonstrates that scaling significantly reduces the condition number of the $\mathbf{U}$ and $\mathbf{D}$ factors of the covariance matrix.

THE APPLICATION TO LINEAR COVARIANCE ANALYSIS

With the above in hand, we can use the same machinery as is already available in the LinCov framework to recast the covariance equations in scaled form without affecting the computation of the partials and process noise matrix. In the current LinCov framework, we compute $\mathbf{\Phi}(t, t_0)$, $\overline{\mathbf{Q}}(t)$, and $\mathbf{H}(t)$. We now recast the scaled analogs ($\mathbf{\Phi}_s(t, t_0)$, $\overline{\mathbf{Q}}_s(t)$, and $\mathbf{H}_s(t)$), in terms of the aforementioned quantities ($\mathbf{\Phi}(t, t_0)$, $\overline{\mathbf{Q}}(t)$, and $\mathbf{H}(t)$) as

$$\mathbf{\Phi}_s(t, t_0) = \mathbf{M}_s \mathbf{\Phi}(t, t_0) \mathbf{M}_s^{-1} \tag{62}$$

$$\overline{\mathbf{Q}}_s(t) = \mathbf{M}_s \overline{\mathbf{Q}}(t) \mathbf{M}_s^T \tag{63}$$

$$\mathbf{H}_s(t) = \mathbf{H}(t)\, \mathbf{M}_s^{-1} \tag{64}$$

with

$$\mathbf{P}_s(t_0) = \mathbf{M}_s \mathbf{P}(t_0) \mathbf{M}_s^T \tag{65}$$

and

$$\mathbf{P}(t) = \mathbf{M}_s^{-1} \mathbf{P}_s(t) \mathbf{M}_s^{-T} \tag{66}$$

The augmented and onboard covariance measurement updates (dropping the k subscript) require careful consideration. Begin with the onboard measurement Kalman Gain matrix

$$\mathbf{K}_{onb} = \mathbf{P}_{onb}^- \mathbf{H}_{onb}^T \left(\mathbf{H}_{onb} \mathbf{P}_{onb}^- \mathbf{H}_{onb}^T + \mathbf{R}_{onb} \right)^{-1} \tag{67}$$

so that the scaled onboard measurement Kalman Gain matrix (in terms of the prior development) is

$$\mathbf{K}_{onb_s} = \mathbf{P}_{onb_s}^- \mathbf{H}_{onb_s}^T \left(\mathbf{H}_{onb_s} \mathbf{P}_{onb_s}^- \mathbf{H}_{onb_s}^T + \mathbf{R}_{onb} \right)^{-1} \tag{68}$$

where $\mathbf{K}_{onb_s}$, $\mathbf{P}_{onb_s}$, and $\mathbf{H}_{onb_s}$ are defined in terms of the onboard scaling matrix, $\mathbf{M}_{onb_s}$, as

$$\mathbf{P}_{onb_s} \triangleq \mathbf{M}_{onb_s} \mathbf{P}_{onb} \mathbf{M}_{onb_s}^T \tag{69}$$

$$\mathbf{K}_{onb_s} \triangleq \mathbf{M}_{onb_s} \mathbf{K}_{onb} \tag{70}$$

$$\mathbf{H}_{onb_s} \triangleq \mathbf{H}_{onb} \mathbf{M}_{onb_s}^{-1} \tag{71}$$

With this in hand, the *scaled onboard covariance matrix measurement update* is

$$\mathbf{P}_{onb_s}^+ = \left(\mathbf{I} - \mathbf{K}_{onb_s} \mathbf{H}_{onb_s} \right) \mathbf{P}_{onb_s}^- \left(\mathbf{I} - \mathbf{K}_{onb_s} \mathbf{H}_{onb_s} \right)^T + \mathbf{K}_{onb_s} \mathbf{R}_{onb} \mathbf{K}_{onb_s}^T \tag{72}$$

and defining the square-root of the covariance matrix as

$$\mathbf{P}_{onb_s}^- \triangleq \mathbf{S}_{onb_s}^- \left(\mathbf{S}_{onb_s}^- \right)^T \tag{73}$$

$$\mathbf{P}_{onb_s}^+ \triangleq \mathbf{S}_{onb_s}^+ \left(\mathbf{S}_{onb_s}^+ \right)^T \tag{74}$$

the *scaled onboard **square-root** covariance matrix measurement update* is

$$\mathbf{S}_{onb_s}^+ = \mathtt{qr}\left(\left[\begin{array}{ccc} \left(\mathbf{I} - \mathbf{K}_{onb_s} \mathbf{H}_{onb_s} \right) \mathbf{S}_{onb_s}^- & \vdots & \mathbf{K}_{onb_s} \mathbf{R}_{onb}^{1/2} \end{array} \right] \right) \tag{75}$$

The augmented covariance matrix measurement update is

$$\begin{aligned}\mathbf{P}_{aug}^{+} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}_{onb}\mathbf{H}_{env} & (\mathbf{I}-\mathbf{K}_{onb}\mathbf{H}_{nav}) \end{bmatrix} \mathbf{P}_{aug}^{-} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}_{onb}\mathbf{H}_{env} & (\mathbf{I}-\mathbf{K}_{onb}\mathbf{H}_{nav}) \end{bmatrix}^{T} \\ &\quad + \begin{bmatrix} \mathbf{0} \\ \mathbf{K}_{onb} \end{bmatrix} \mathbf{R}_{onb} \begin{bmatrix} \mathbf{0} \\ \mathbf{K}_{onb} \end{bmatrix}^{T} \end{aligned} \tag{76}$$

and the scaled augmented covariance matrix measurement update is

$$\begin{aligned}\mathbf{P}_{aug_s}^{+} &= \mathbf{M}_{aug} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{env} & \left(\mathbf{I}-\mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{nav}\right) \end{bmatrix} \mathbf{M}_{aug}^{-1}\mathbf{P}_{aug_s}^{-}\mathbf{M}_{aug}^{-T} \\ &\quad \times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{env} & \left(\mathbf{I}-\mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{nav}\right) \end{bmatrix}^{T} \mathbf{M}_{aug}^{T} \\ &\quad + \mathbf{M}_{aug} \begin{bmatrix} \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s} \end{bmatrix} \mathbf{R}_{onb} \begin{bmatrix} \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s} \end{bmatrix}^{T} \mathbf{M}_{aug}^{T} \end{aligned} \tag{77}$$

where we note that the unscaled $\mathbf{H}_{env}$ and $\mathbf{H}_{nav}$ appear in the above equation.

Defining $\mathcal{A}_{aug}$ and $\mathcal{B}_{aug}$ as

$$\mathcal{A}_{aug} \triangleq \mathbf{M}_{aug} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{env} & \left(\mathbf{I}-\mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s}\mathbf{H}_{nav}\right) \end{bmatrix} \mathbf{M}_{aug}^{-1} \tag{78}$$

$$\mathcal{B}_{aug} \triangleq \mathbf{M}_{aug} \begin{bmatrix} \mathbf{0} \\ \mathbf{M}_{onb_s}^{-1}\mathbf{K}_{onb_s} \end{bmatrix} \tag{79}$$

the *scaled augmented covariance matrix measurement update* becomes

$$\mathbf{P}_{aug_s}^{+} = \mathcal{A}_{aug}\mathbf{P}_{aug_s}^{-}\mathcal{A}_{aug}^{T} + \mathcal{B}_{aug}\mathbf{R}_{onb}\mathcal{B}_{aug}^{T} \tag{80}$$

and defining the square-root of the augmented covariance matrix as

$$\mathbf{P}_{aug_s}^{-} \triangleq \mathbf{S}_{aug_s}^{-}\left(\mathbf{S}_{aug_s}^{-}\right)^{T} \tag{81}$$

$$\mathbf{P}_{aug_s}^{+} \triangleq \mathbf{S}_{aug_s}^{+}\left(\mathbf{S}_{aug_s}^{+}\right)^{T} \tag{82}$$

the *scaled augmented* ***square-root*** *covariance matrix measurement update* becomes

$$\mathbf{S}_{aug_s}^{+} = \mathtt{qr}\left(\left[\ \mathcal{A}_{aug}\mathbf{S}_{aug_s}^{-} \ \vdots\ \mathcal{B}_{aug}\mathbf{R}_{onb}^{1/2}\ \right]\right) \tag{83}$$

THE APPLICATION TO BATCH FILTERING

A similar process can be carried out for the case of batch filtering. To this end, begin with the normal equation (with *a priori* information at an epoch t_0) after processing m measurements is

$$\hat{\mathbf{x}}_0 = \left(\overline{\mathbf{P}}_0^{-1} + \sum_{i=1}^{m}\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\right)^{-1}\left(\sum_{j=1}^{m}\mathbf{H}_j^T\mathbf{R}_j^{-1} + \overline{\mathbf{P}}_0^{-1}\bar{\mathbf{x}}_0\right) \tag{84}$$

As before we recall the relationship between the scaled and unscaled quantities as

$$\begin{aligned}
\mathbf{x}_s &= \mathbf{M}_s\,\mathbf{x} \\
\mathbf{P}_s &= \mathbf{M}_s\,\mathbf{P}\,\mathbf{M}_s^T \\
\mathbf{H}_s &= \mathbf{H}\,\mathbf{M}_s^{-1}
\end{aligned}$$

so the normal equation becomes

$$\hat{\mathbf{x}}_{s_0} = \mathbf{M}_s\left[\mathbf{M}_s^T\overline{\mathbf{P}}_{s_0}^{-1}\mathbf{M}_s + \sum_{i=1}^{m}\mathbf{M}_s^T\mathbf{H}_{s_i}^T\mathbf{R}_i^{-1}\mathbf{H}_{s_i}\mathbf{M}_s\right]^{-1}\left(\sum_{j=1}^{m}\mathbf{M}_s^T\mathbf{H}_{s_j}^T\mathbf{R}_j^{-1} + \mathbf{M}_s^T\overline{\mathbf{P}}_{s_0}^{-1}\mathbf{M}_s\mathbf{M}_s^{-1}\bar{\mathbf{x}}_{s_0}\right) \quad (85)$$

which, after factoring the $\mathbf{M}_s^T$ and $\mathbf{M}_s$ out of the square bracket and cancelling, leads to

$$\hat{\mathbf{x}}_{s_0} = \left(\overline{\mathbf{P}}_{s_0}^{-1} + \sum_{i=1}^{m}\mathbf{H}_{s_i}^T\mathbf{R}_i^{-1}\mathbf{H}_{s_i}\right)^{-1}\left(\sum_{j=1}^{m}\mathbf{H}_{s_j}^T\mathbf{R}_j^{-1} + \overline{\mathbf{P}}_{s_0}^{-1}\bar{\mathbf{x}}_{s_0}\right) \quad (86)$$

CHOICE OF THE SCALING MATRIX, M_S

So far we haven't said anything about how to choose the scaling matrix, $\mathbf{M}_s$. It turns out that there is no limit to the choice of the scaling matrix, so long as it is is invertible. Whereas in Apollo, they scaled by the so-called 'canonical' units (Earth radii for distance and earth rotation rate for time), there is no need to restrict ourselves to this. Wouldn't it better to let the problem itself dictate the scaling parameters instead of imposing it *a priori*? To this end, each component of, say, position would be allowed to have a different scaling if the problem dictates it. These have been recently called 'designer units'.[2] In fact, there is no requirement that the scaling matrix needs to be diagonal – it can be a full matrix if needed. We ought to allow the covariance matrix to dictate the scaling parameters. But how to choose a proper scaling matrix?

Three choices rise to the forefront for consideration: 'Powers of 10' scaling, eigenvalue decomposition scaling, and Cholesky scaling. Each will be considered in turn.

'Powers of 10' Scaling

This rather straightforward scaling constructs a diagonal matrix whose elements are scaled by the power of the diagonal elements of the covariance matrix as

$$\begin{aligned}
m_{s_{ii}} &= 10^{-\mathrm{floor}(\log_{10}(\sqrt{P_{ii}})} && (87) \\
m_{s_{ii}}^{-1} &= 10^{\mathrm{floor}(\log_{10}(\sqrt{P_{ii}})} && (88)
\end{aligned}$$

In the above equation, $m_{s_{ii}}^{-1}$ populates the (diagonal) elements of the diagonal matrix $\mathbf{M}_s^{-1}$. Whereas this type of scaling is simple, it does not take into account the correlations of the covariance matrix. As well, while it does reduce the condition number of the transformed covariance matrix, it does not, in general, reduce it to anywhere near 1. However, it does provide a great deal of insight/intuition in scaling the covariance matrix by only scaling by powers of 10 (essentially changing units from, say, meters to kilometers).

Eigenvalue Decomposition Scaling

This type of scaling is an eigenvalue decomposition of the covariance matrix and takes into account the correlations of the covariance matrix. Thus, if the eigenvalues and eigenvectors of $\mathbf{P}$ are placed in $\mathbf{D}$ and $\mathbf{V}$ ($\mathbf{P} = \mathbf{V}\mathbf{D}\mathbf{V}^T$, with $\mathbf{V}\mathbf{V}^T = \mathbf{I}$), respectively, with $\mathbf{D}$ being a diagonal matrix whose entries are the eigenvalues (usually sorted from smallest to largest), the scaling matrix in this paradigm is

$$\mathbf{M}_s = \mathbf{D}^{-1/2}\mathbf{V}^T \tag{89}$$

and

$$\mathbf{M}_s^{-1} = \mathbf{V}\sqrt{\mathbf{D}} \tag{90}$$

This has the benefit of reducing the condition number of the scaled covariance matrix to 1 at the cost of a loss of intuition as to the elements of the full scaling matrix. That the condition number of the transformed matrix is 1 is seen because

$$\mathbf{P}_s \ = \ \mathbf{M}_s\,\mathbf{P}\,\mathbf{M}_s^T \ = \ \mathbf{D}^{-1/2}\,\mathbf{V}^T\left(\mathbf{V}\,\mathbf{D}\,\mathbf{V}^T\right)\mathbf{V}\,\mathbf{D}^{-1/2} \ = \ \mathbf{D}^{-1/2}\,\mathbf{D}\,\mathbf{D}^{-1/2} \ = \ \mathbf{I} \tag{91}$$

and the condition number of the identity matrix is 1. As well, the transformed covariance matrix at the time of the conditioning is the identity matrix, a rather intriguing result and something we were after. So this is an ‘optimal’ scaling matrix if we are seeking the optimal condition number of the covariance matrix.

Cholesky Scaling

This scaling is a Cholesky decomposition of the covariance matrix. Recall that the Cholesky decomposition of the covariance matrix is

$$\mathbf{P} \ = \ \mathbf{S}\,\mathbf{S}^T \tag{92}$$

where $\mathbf{S}$ is the square-root of the covariance matrix, so that were we to choose the scaling matrix as

$$\mathbf{M}_s \ = \ \mathbf{S}^{-1} \tag{93}$$

the scaled covariance matrix becomes

$$\mathbf{P}_s \ = \ \mathbf{M}_s\,\mathbf{P}\,\mathbf{M}_s^T \ = \ \mathbf{S}^{-1}\mathbf{S}\,\mathbf{S}^T\mathbf{S}^{-T} \ = \ \mathbf{I} \tag{94}$$

whose condition number is 1. Generally this scaling matrix is a triangular (upper or lower) matrix, but it doesn’t have to be, for there are an infinite number of square root factorizations of a square matrix.

SCALING PHILOSOPHIES

A myriad of scaling methodologies can be brought to bear on conditioning the covariance matrices. A short list includes: initial scaling, continuous scaling, scaling at the time of measurements, and occasional scaling.

Initial scaling is just as it sounds: scale the onboard and the augmented covariance matrices once at the initial time. Continuous scaling denotes computing and applying the scaling matrices at every time step. Alternatively, the scaling can be calculated and applied only at the time of measurements. Finally, the scaling matrices can be computed and applied at discrete/isolated epochs. These epochs may include (but are not limited to) times of the maneuvers. Any or all of these methodologies can be utilized.

We note that the goal is to have a well-conditioned covariance matrix (*i.e.* close to 1). Limited experience to date has demonstrated that keeping the condition number of the covariance matrices to be under 10^5 is the sweet spot. Heroic efforts to reduce the condition number below this are not worth the sweat.

As well, our limited experience with scaling a system with a large state-space indicates that eigenvalue scaling, while it may be intuitive to our sensibilities, actually does not work well because of the condition number of the matrices at hand – ironically the very issue we were trying to remedy. Our experience has shown that if the condition number of the covariance matrix is large, the computation of the eigenvalues and eigenvectors could themselves introduce errors to the scaling matrices. In light of this, it is not recommended that eigenvalue scaling be utilized, as much as one may want to use it.

THE EFFECT OF SCALING ON EM-1 LINEAR COVARIANCE ANALYSIS

To demonstrate the efficacy of this methodology, an EM-1 Trajectory (with optical navigation) was analyzed. In this case, scaling was performed for $\mathbf{P}_{aug}$ and $\mathbf{P}_{onboard}$ at the initial time and after that it was performed only on $\mathbf{P}_{onboard}$. To this end, periodic scaling was performed whenever a measurement was taken and whenever a maneuver was performed. The scaling chosen was the 'Powers of 10' paradigm.

Figure 1 shows the effect of scaling on $\mathbf{P}_{onb}$; the left panel contains the condition number of the *unscaled* onboard covariance matrix while the right panel contains the condition number of the *scaled* onboard covariance matrix. Note that the condition number for the scaled $\mathbf{P}_{aug}$ is 5 orders of magnitude less than for the unscaled $\mathbf{P}_{aug}$. Nevertheless, the magnitude of the condition number for either is still troublesome. Likewise Figure 2 contains the condition number for the Unscaled (left) and Scaled $\mathbf{P}_{onboard}$. Here we note that the condition number for the scaled $\mathbf{P}_{onboard}$ is 10 orders of magnitude less than that of the unscaled $\mathbf{P}_{onboard}$.

Figure 3 shows the effect of scaling on $\mathbf{S}_{aug}$; the left panel contains the condition number of the *unscaled* onboard square-root covariance matrix, while the right panel contains the condition number of the *scaled* onboard square-root covariance matrix. Note that the condition number for the scaled $\mathbf{S}_{aug}$ is about 5 orders of magnitude less than for the unscaled $\mathbf{S}_{aug}$.

Figure 4 shows the effect of scaling on $\mathbf{S}_{onboard}$; the left panel contains the condition number of the *unscaled* onboard square-root covariance matrix, while the right panel contains the condition number of the *scaled* onboard square-root covariance matrix. Note that the condition number for the scaled $\mathbf{S}_{onboard}$ is about 7 orders of magnitude less than for the unscaled $\mathbf{S}_{onboard}$. In particular, since we have particularly focused on the condition number of $\mathbf{S}_{onboard}$ with respect to scaling, the condition number of the scaled $\mathbf{S}_{onboard}$ remains below 10!

We now investigate the effects of scaling on the EI parameters, particularly the four EI constraints. Figure 5 shows the comparison of the downrange versus flight path angle delivery statistics. There is a very noticeable difference between the two, indicating that scaling does significantly affect the performance of the solution. Figure 6 contains the comparison between effect on the entry interface velocity magnitude versus downrange position delivery ellipses the unscaled and scaled

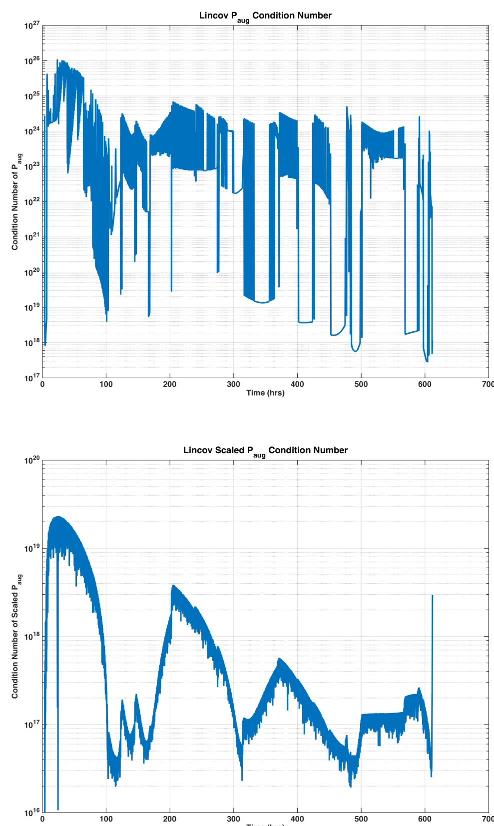

Figure 1. Condition Number for Unscaled and Scaled P_{aug}

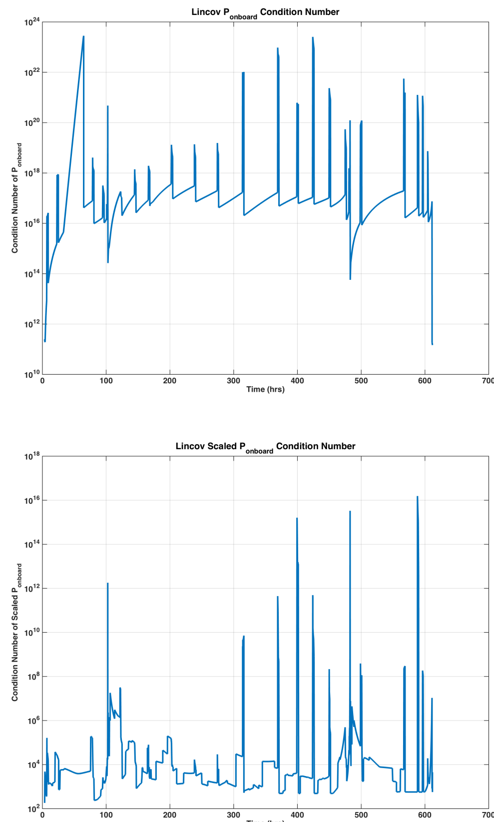

Figure 2. Condition Number for Unscaled and Scaled $\mathrm{P}_{onboard}$

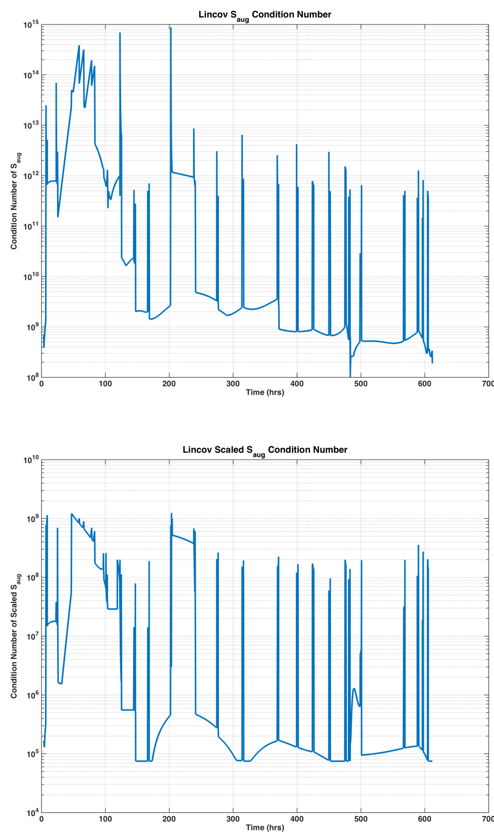

Figure 3. Condition Number for Unscaled and Scaled S_{aug}

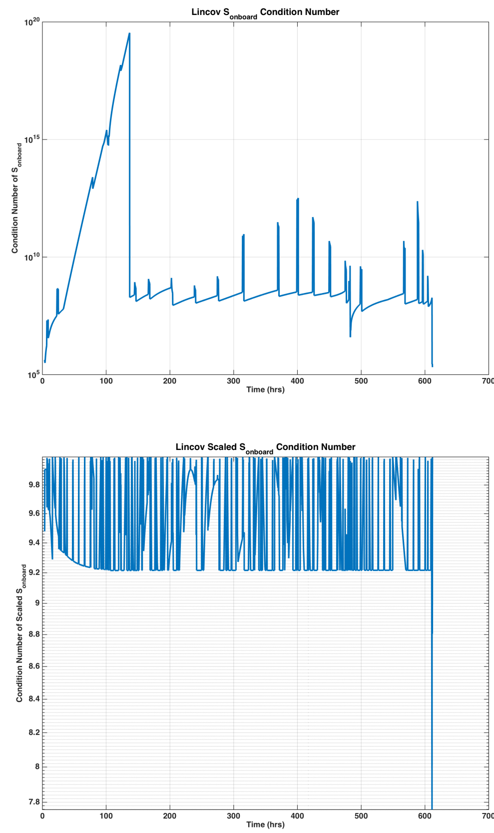

Figure 4. Condition Number for Unscaled and Scaled S$_{onboard}$

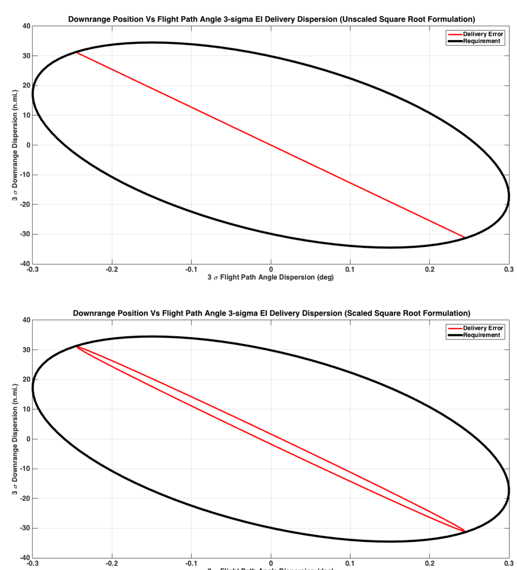

Figure 5. Comparison of Entry Interface Downrange Position Vs Flight Path Angle Delivery Ellipses for Unscaled and Scaled Covariance Matrix

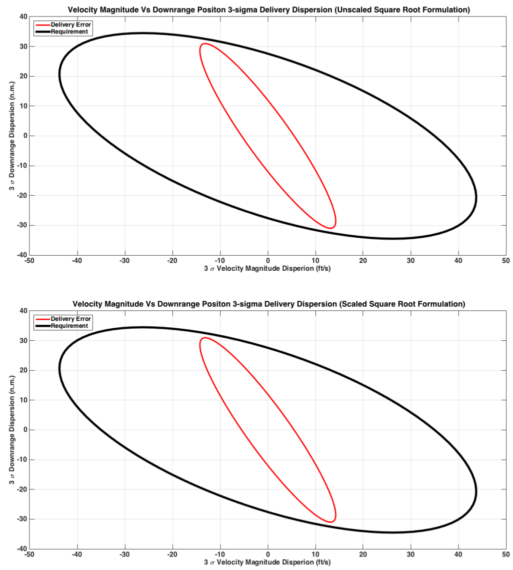

Figure 6. Comparison of Entry Interface Velocity Magnitude Vs Downrange Position Delivery Ellipses for Unscaled and Scaled Covariance Matrix

covariance formulations Figure 7 contains the comparison between effect on the entry interface velocity magnitude versus flight path angle delivery ellipses the unscaled and scaled covariance formulations Figure 8 contains the comparison between effect on the entry interface out-of-plane delivery ellipses the unscaled and scaled covariance formulations

We note that in Figures 6-8, the scaled and unscaled formulations show no discernible difference; that is most definitely not the case in Figure 5 where there is a noticeable difference. To further drive home the point, the Downrange Position Vs Flight Path Angle delivery constraint is the driving constraint and any improvement in the conditioning is significant. As well the delivery ellipse for the scaled formulation is 'ellipse-like' in contrast to its unscaled counterpart whose ellipse is rather linear. How do we know that the scaled formulation is the correct one? Simply because the condition numbers of the scaled $\mathbf{S}_{aug}$ and $\mathbf{S}_{onboard}$ are substantially smaller than their unscaled counterparts.

This also introduces a new metric which ought to be considered: the time-history of the condition number of the covariance matrix along the trajectory. Usually, navigators are just interested in ensuring that the covariance matrix stays positive definite. Now we may have a need to look at the condition number of the covariance matrix as well!!! This is most certainly the case for linear covariance analysis.

CONCLUSIONS

This memo has endeavored to present a reasonably complete development of scaling for Kalman filtering and linear covariance analysis (as well as for batch filtering). It is recommended that 'Powers of 10' or 'Cholesky' scaling (or a combination of the two) be used as a matter of course whenever the condition numbers rise above, say, 10^8 in order to preclude numerical errors in representing the covariance and hence the state update.

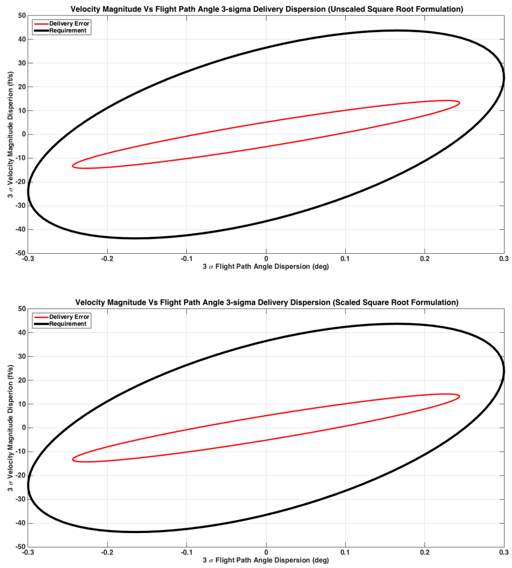

Figure 7. Comparison of Entry Interface Velocity Magnitude Vs Downrange Position Delivery Ellipses for Unscaled and Scaled Covariance Matrix

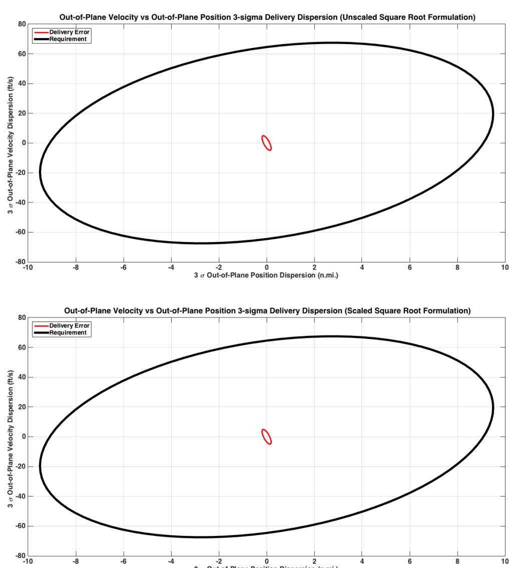

Figure 8. Comparison of Entry Out-of-Plane Position Vs Out-of-Plane Velocity Delivery Ellipses for Unscaled and Scaled Covariance Matrix

REFERENCES

[1] B. Conway, “A Survey of Methods Available for the Numerical Optimization of Continuous Dynamic Systems,” *Journal of Optimization Theory and Applications*, Vol. 152, No. 5, 2012, pp. 271–306.

[2] I. Ross, G. Q., M. Karpenko, and R. Proulx, “Scaling and Balancing for Faster Trajectory Optimization,” *AAS/AIAA Astrodynamics Specialists Conference*, Stevenson, WA, AAS 17-675, August 21–24, 2017.