Then Eq. (22.32) can be written *in extenso* as

$$\dot{\Phi}_{\mathbf{rr}} = \Phi_{\mathbf{vr}}, \tag{22.39}$$
$$\dot{\Phi}_{\mathbf{rv}} = \Phi_{\mathbf{vv}}, \tag{22.40}$$
$$\dot{\Phi}_{\mathbf{vr}} = \frac{\partial \mathbf{a}}{\partial \mathbf{r}}\Phi_{\mathbf{rr}} + \frac{\partial \mathbf{a}}{\partial \mathbf{v}}\Phi_{\mathbf{vr}}, \tag{22.41}$$
$$\dot{\Phi}_{\mathbf{vv}} = \frac{\partial \mathbf{a}}{\partial \mathbf{r}}\Phi_{\mathbf{rv}} + \frac{\partial \mathbf{a}}{\partial \mathbf{v}}\Phi_{\mathbf{vv}}, \tag{22.42}$$

while Eq. (22.36) separates into

$$\dot{\Phi}_{\mathbf{rp}} = \Phi_{\mathbf{vp}}, \tag{22.43}$$
$$\dot{\Phi}_{\mathbf{vp}} = \frac{\partial \mathbf{a}}{\partial \mathbf{r}}\Phi_{\mathbf{rp}} + \frac{\partial \mathbf{a}}{\partial \mathbf{v}}\Phi_{\mathbf{vp}} + \frac{\partial \mathbf{a}}{\partial \mathbf{P}}. \tag{22.44}$$

The initial condition for the numerical integration of the STM are given blockwise as

$$\Phi_{\mathbf{rr}}(t_0, t_0) = I, \tag{22.45}$$
$$\Phi_{\mathbf{rv}}(t_0, t_0) = 0, \tag{22.46}$$
$$\Phi_{\mathbf{vr}}(t_0, t_0) = 0, \tag{22.47}$$
$$\Phi_{\mathbf{vv}}(t_0, t_0) = I, \tag{22.48}$$
$$\Phi_{\mathbf{rp}}(t_0, t_0) = 0, \tag{22.49}$$
$$\Phi_{\mathbf{vp}}(t_0, t_0) = 0. \tag{22.50}$$

The *variational equations*, as the differential equations for determining the STM $\Phi(t, t_0)$ are known, exhibit the peculiar property according to which they are column-decoupled. This is clear first of all in terms of the block-column structure in which they have been so far written. In fact, the blocks referred to are those defined by the position components (index **r**), the velocity components (index **v**) and the dynamical parameter components (index **p**) and it may be seen upon inspection that equations (22.39) and (22.41) are only coupled between them, as are equations (22.40) and (22.42) as well as equations (22.43) and (22.44). The same property, as may be easily verified, is also valid at the level of each single column, i.e., for each single position, velocity and dynamic parameter component. Each column of the STM being decoupled from the other columns, its integration can be carried out independently. However, the fact that the dynamical equations must also be concurrently integrated makes the decoupled integration more of a burden than an advantage.

The state transition matrix has the following useful properties. For any instant of time t_i, t_j and t_k we have

$$\begin{aligned} \Phi(t_k, t_k) &= I\,, \\ \Phi(t_i, t_k) &= \Phi(t_i, t_j)\Phi(t_j, t_k)\,, \\ \Phi(t_i, t_k) &= \Phi^{-1}(t_k, t_i)\,. \end{aligned} \tag{22.51}$$

22.2 Weighted Least Squares (WLS) Estimate

If the following definitions are used

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_l \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_l \end{bmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_l \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{w}_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{w}_l \end{bmatrix}, \tag{22.52}$$

and if *the subscript and the tilde on* $\tilde{\mathbf{x}}_0$ *are dropped* for shortness, then the system of observation equations (??) can be expressed even more compactly as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}\,, \qquad \mathbf{W}, \tag{22.53}$$

where $\mathbf{y}$ is an $m \times 1$ vector, $\mathbf{x}$ is an $n \times 1$ vector, $\boldsymbol{\varepsilon}$ is an $m \times 1$ vector, $\mathbf{H}$ is an $m \times n$ mapping matrix, where $m = p \times l$ is the total number of observations. If p or l is sufficiently large, the condition $m > n$ that ensures an overdetermined system of observations is satisfied. $\mathbf{W}$ is the block diagonal measurement weight matrix.

The weighted least squares solution selects the estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ as that value that minimizes the sum of the squares of the calculated observation residuals. That is, $\hat{\mathbf{x}}$ is selected to minimize the *performance index*

$$J(\mathbf{x}) = \frac{1}{2}\boldsymbol{\varepsilon}^T\mathbf{W}\boldsymbol{\varepsilon} = \sum_{i=1}^{l} \frac{1}{2}\boldsymbol{\varepsilon}_i^T\mathbf{w}_i\boldsymbol{\varepsilon}_i \ . \tag{22.54}$$

Using equation (22.53), $J(\mathbf{x})$ can be expressed as

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T\mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) \, . \tag{22.55}$$

A necessary condition for a minimum of $J(\mathbf{x})$ is that its first derivative with respect to $\mathbf{x}$ vanishes

$$\frac{\partial J}{\partial \mathbf{x}} = -(\mathbf{y} - \mathbf{H}\mathbf{x})^T\mathbf{W}\mathbf{H} = \mathbf{0}, \tag{22.56}$$

which can also be written as

$$\mathbf{H}^T\mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) = \mathbf{0}. \tag{22.57}$$

This expression can be rearranged to obtain the *normal equations* in the least squares formulation as

$$(\mathbf{H}^T\mathbf{W}\mathbf{H})\mathbf{x} = \mathbf{H}^T\mathbf{W}\mathbf{y} \, . \tag{22.58}$$

If the *normal matrix* $\mathbf{H}^T\mathbf{W}\mathbf{H}$ is positive definite, it will have an inverse and the solution to (22.58) is

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{y} \, . \tag{22.59}$$

The value of $\hat{\mathbf{x}}$ given by equation (22.59) is the *weighted least squares estimate* and is the estimate that minimizes the sum of squares of the weighted observation errors. Note that equation (22.59) can be expressed as

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{H}^T\mathbf{W}\mathbf{y} \, , \tag{22.60}$$

where

$$\mathbf{P} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1} \tag{22.61}$$

is the symmetric $n \times n$ *solution covariance matrix* $\mathbf{P}$. Furthermore, if it exists, it must be positive definite, since it is computed as the inverse of the positive definite matrix, $\mathbf{H}^T\mathbf{W}\mathbf{H}$. Parameter observability is related to the rank of this matrix. If all parameters in $\mathbf{x}_k$ are observable (i.e., they can be uniquely determined from the observation set $\mathbf{y}$), then $\mathbf{P}$ will be full rank and will have an inverse. The number of independent observations must be greater than or equal to the number of parameters being estimated if $\mathbf{P}$ is to be invertible. Furthermore, $\mathbf{P}$ is related to the accuracy of the estimate, $\hat{\mathbf{x}}$. In general, the larger the magnitude of the elements of the matrix, $\mathbf{P}_k$, the less accurate the estimate.

If an a priori value, $\bar{\mathbf{x}}$, is available for $\mathbf{x}$ and an associated weighting matrix, $\bar{\mathbf{W}}$, is given, the weighted least square estimate for $\mathbf{x}$ can be obtained by choosing for $\hat{\mathbf{x}}$ the value of $\mathbf{x}$, which minimizes the performance index

$$J(\mathbf{x}_k) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T\mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x}) + \frac{1}{2}(\bar{\mathbf{x}} - \mathbf{x})^T\bar{\mathbf{W}}(\bar{\mathbf{x}} - \mathbf{x}) \, . \tag{22.62}$$

This results in

$$\hat{\mathbf{x}}_k = (\mathbf{H}^T\mathbf{W}\mathbf{H} + \bar{\mathbf{W}})^{-1}\left(\mathbf{H}^T\mathbf{W}\mathbf{y} + \bar{\mathbf{W}}\bar{\mathbf{x}}\right) \, , \tag{22.63}$$

where $\bar{\mathbf{x}}$ represents an a priori estimate of $\mathbf{x}$ and $\bar{\mathbf{W}}$ represents a weighting matrix for the a priori estimate of $\mathbf{x}$.

22.3 Computational algorithm for the batch processor

Assume that we wish to estimate the state deviation vector $\mathbf{x}_0$ at a reference time, t_0. Given a set of initial conditions $\mathbf{X}^*(t_0)$, an a priori estimate $\bar{\mathbf{x}}_0$ and the associated error covariance matrix, $\bar{\mathbf{P}}_0$, the computational algorithm for the batch processor generally uses the normal equation form for $\hat{\mathbf{x}}_0$. Writing equation (22.63) in normal equations form for a batch of observations and recognizing that $\mathbf{W} = \mathbf{R}^{-1}$ and $\bar{\mathbf{W}} = \bar{\mathbf{P}}_0^{-1}$ yields

$$(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \bar{\mathbf{P}}_0^{-1})\hat{\mathbf{x}}_0 = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} + \bar{\mathbf{P}}_0^{-1}\bar{\mathbf{x}}_0\,. \tag{22.64}$$

Here t_0 is an arbitrary epoch and all quantities in equation (22.64) are assumed to have been mapped to this epoch using the appropriate state transition matrices as illustrated in equations (??) and (??).

If $\mathbf{R}$ is a block diagonal matrix, i.e., if the observations are uncorrelated in time although correlations between the observations at any given time may exist, these matrices simply may be accumulated as follows:

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \sum_{i=1}^{l}[\tilde{\mathbf{H}}_i\boldsymbol{\Phi}(t_i,t_0)]^T\mathbf{R}_i^{-1}\tilde{\mathbf{H}}_i\boldsymbol{\Phi}(t_i,t_0)\,, \tag{22.65}$$

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} = \sum_{i=1}^{l}[\tilde{\mathbf{H}}_i\boldsymbol{\Phi}(t_i,t_0)]^T\mathbf{R}_i^{-1}\mathbf{y}_i\,. \tag{22.66}$$

In general $\mathbf{X}^*(t_0)$ would be chosen so that $\bar{\mathbf{x}}_0 = \mathbf{0}$, and $\bar{\mathbf{P}}_0$ would reflect the relative accuracy of the elements of the initial conditions vector $\mathbf{X}^*(t_0)$. In theory $\bar{\mathbf{x}}_0$ and $\bar{\mathbf{P}}_0$ represent information and should be treated as data that are merged with the observation data, as indicated by equation (22.64). Consequently, the value of $\mathbf{X}_0^* + \bar{\mathbf{x}}_0$ should be held constant for the beginning of each iteration. Since the initial condition vector $\mathbf{X}_0^*$ is augmented by the value of $\hat{\mathbf{x}}_0$ after each iteration, that is, $(\mathbf{X}_0^*)_n = (\mathbf{X}_0^*)_{n-1} + (\hat{\mathbf{x}}_0)_{n-1}$, holding $\mathbf{X}_0^* + \bar{\mathbf{x}}_0$ constant results in the following expression for $(\bar{\mathbf{x}}_0)_n$

$$(\bar{\mathbf{x}}_0)_n = (\bar{\mathbf{x}}_0)_{n-1} - (\hat{\mathbf{x}}_0)_{n-1}\,. \tag{22.67}$$

Recall that the state transition matrix is obtained by integrating

$$\dot{\boldsymbol{\Phi}}(t,t_k) = \mathbf{A}(t)\boldsymbol{\Phi}(t,t_k)\,, \tag{22.68}$$

subject to the initial conditions $\boldsymbol{\Phi}(t_k,t_k) = \mathbf{I}$ along with the nonlinear equations, $\dot{\mathbf{X}}^* = \mathbf{F}(\mathbf{X}^*,t)$, which define the nominal trajectory, $\dot{\mathbf{X}}^*(t)$. The matrix $\mathbf{A}(t)$ is evaluated on the reference trajectory,

$$\mathbf{A}(t) = \frac{\partial\mathbf{F}(\mathbf{X}^*,t)}{\partial\mathbf{X}}\,, \tag{22.69}$$

where $\mathbf{F}(\mathbf{X}^*,t)$ is the time derivative of the state vector in the differential equations governing the time evolution of the system. The observation-state mapping matrix is given by

$$\tilde{\mathbf{H}}_i = \frac{\partial\mathbf{G}(\mathbf{X}_i^*,t_i)}{\partial\mathbf{X}}\,, \tag{22.70}$$

where $\mathbf{G}(\mathbf{X}_i^*,t_i)$ are the observation-state relationships evaluated on the nominal or reference trajectory.

Notice that the solution for $\hat{\mathbf{x}}_0$ involved inversion of the information matrix, $\boldsymbol{\Lambda}_0$, where

$$\boldsymbol{\Lambda}_0 = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \bar{\mathbf{P}}_0^{-1}\,. \tag{22.71}$$

Generally the normal equations would not be solved by a direct inversion of $\boldsymbol{\Lambda}_0$ but rather would be solved by an indirect but more accurate technique, such as the Cholesky decomposition. The sequence of operations required to implement the batch estimation process is outlined in Figure ??, where we assume that there are no observations at t_0. If observations exist at t_0, set $\boldsymbol{\Lambda} = \bar{\mathbf{P}}_0^{-1} + \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0$ and $\mathbf{N} = \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{y}_0$ in the initialization. As previously stated, the entire sequence of computations are repeated until the estimation process has converged. If there are observations at t_0, the state transition matrix for processing these observations is the identity matrix.

This procedure yields a minimum value of the performance index

$$J(\mathbf{x}) = (\hat{\mathbf{x}}_0 - \bar{\mathbf{x}}_0)^T \bar{\mathbf{P}}_0^{-1} + \sum_{i=1}^{l} \hat{\boldsymbol{\varepsilon}}_i^T \mathbf{R}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i \,, \tag{22.72}$$

where

$$\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_0 \,, \tag{22.73}$$

and $\hat{\boldsymbol{\varepsilon}}_i$ is the best estimate of the observation error.

In practice, $\bar{\mathbf{P}}_0$ is generally not a realistic representation of the accuracy of $\bar{\mathbf{x}}_0$ and it is used only to better condition to estimation error covariance matrix, $\mathbf{P}$. In this case, $\bar{\mathbf{x}}_0$ usually is set to zero for each iteration and $\bar{\mathbf{P}}_0$ is chosen to be a diagonal matrix with large diagonal values. Hence, the first term in equation (22.72) will be very small and the tracking data residuals will determine the value of $J(\mathbf{x})$. The rms (root mean square) of the observation residuals generally is computed and may be used as a measure of convergence; when the rms no larger changes the solution is assumed to be converged. The rms is computed from

$$\text{rms} = \left\{ \frac{\sum_{i=1}^{l} \hat{\boldsymbol{\varepsilon}}_i^T \mathbf{R}_i^{-1} \hat{\boldsymbol{\varepsilon}}_i}{m} \right\}^{\frac{1}{2}} \,, \tag{22.74}$$

where $\hat{\boldsymbol{\varepsilon}}_i$ is a p-vector and $m = l \times p$. Hence, m is the total number of observations. The equation (22.74) is referred to as the weighted rms. If the rms is computed without including the weighting matrix, $\mathbf{R}_i^{-1}$, it may be referred to as the unweighted rms or just the rms.