2.8 Improving Recursively

2.8.1 Least-Squares

A common situation arises that one has a solution $\tilde{\mathbf{x}}, \tilde{\mathbf{n}}, \mathbf{P}$, and more information becomes available, often in the form of further noisy linear constraints. One way of using the new information is to combine the old and new equations into one larger system, and re-solve. This approach may well be the best one. Sometimes, however, perhaps because the earlier equations have been discarded, or for reasons of storage or both, one prefers to retain the information in the previous solution without re-solving the entire system. So-called recursive methods, in both least-squares and minimum variance estimation, provide the appropriate recipes.

Let the original equations be re-labeled so that we can distinguish them from those that come later, in the form,

{38001}
$$\mathbf{E}(1)\mathbf{x} + \mathbf{n}(1) = \mathbf{y}(1) \tag{2.418}$$

where the noise $\mathbf{n}(1)$ has zero mean and covariance matrix $\mathbf{R}_{nn}(1)$. Let the estimate of the solution to (2.418) from one of the estimators be written as $\tilde{\mathbf{x}}(1)$, with uncertainty $\mathbf{P}(1)$. As a specific example, suppose (2.418) is full-rank overdetermined, and was solved using row weighted

least-squares as,

{38002}
$$\tilde{\mathbf{x}}(1) = \left[\mathbf{E}(1)^T \mathbf{R}_{nn}(1)^{-1} \mathbf{E}(1)\right]^{-1} \mathbf{E}(1)^T \mathbf{R}_{nn}(1)^{-1} \mathbf{y}(1), \tag{2.419}$$

with corresponding $\mathbf{P}(1)$ (column weighting is redundant in the full-rank fully-determined case).

Some new observations, $\mathbf{y}(2)$, are obtained, with the error covariance of the new observations given by $\mathbf{R}_{nn}(2)$, so that the problem for the unknown $\mathbf{x}$ is

{38003}
$$\begin{Bmatrix} \mathbf{E}(1) \\ \mathbf{E}(2) \end{Bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}(1) \\ \mathbf{n}(2) \end{bmatrix} = \begin{bmatrix} \mathbf{y}(1) \\ \mathbf{y}(2) \end{bmatrix} \tag{2.420}$$

where $\mathbf{x}$ is the same unknown. We assume $\langle \mathbf{n}(2) \rangle = \mathbf{0}$ and

{3.7.4}
$$\langle \mathbf{n}(1)\mathbf{n}(2)^T \rangle = \mathbf{0}, \tag{2.421}$$

that is, *no correlation of the old and new measurement errors.* A solution to (2.420) should give a better estimate of $\mathbf{x}$ than (2.418) alone, because more observations are available. It is sensible to row weight the concatenated set to

{38004}
$$\begin{bmatrix} \mathbf{R}_{nn}(1)^{-T/2} \mathbf{E}(1) \\ \mathbf{R}_{nn}(2)^{-T/2} \mathbf{E}(2) \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{R}_{nn}(1)^{-T/2} \mathbf{n}(1) \\ \mathbf{R}_{nn}(2)^{-T/2} \mathbf{n}(2) \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{nn}(1)^{-T/2} \mathbf{y}(1) \\ \mathbf{R}_{nn}(2)^{-T/2} \mathbf{y}(2) \end{bmatrix}. \tag{2.422}$$

"Recursive weighted least-squares" seeks the solution to (2.422) without inverting the new, larger matrix, by taking advantage of the existing knowledge of $\mathbf{x}(1)$, $\mathbf{P}(1)$ —however they might actually have been obtained. The objective function corresponding to finding the minimum weighted error norm in (2.422) is,

{recurs11}
$$\begin{aligned} J = &(\mathbf{y}(1) - \mathbf{E}(1)\mathbf{x})^T \mathbf{R}_{nn}(1)^{-1} (\mathbf{y}(1) - \mathbf{E}(1)\mathbf{x}) \\ &+ (\mathbf{y}(2) - \mathbf{E}(2)\mathbf{x})^T \mathbf{R}_{nn}(2)^{-1} (\mathbf{y}(2) - \mathbf{E}(2)\mathbf{x}). \end{aligned} \tag{2.423}$$

Taking the derivatives with respect to $\mathbf{x}$, the normal equations produce a new solution,

{recurs10}
$$\begin{aligned} \tilde{\mathbf{x}}(2) = &\left\{\mathbf{E}(1)^T \mathbf{R}_{nn}(1)^{-1} \mathbf{E}(1) + \mathbf{E}(2)^T \mathbf{R}_{nn}(2)^{-1} \mathbf{E}(2)\right\}^{-1} \\ &\left\{\mathbf{E}(1)^T \mathbf{R}_{nn}(1)^{-1} \mathbf{y}(1) + \mathbf{E}(2)^T \mathbf{R}_{nn}(2)^{-1} \mathbf{y}(2)\right\}. \end{aligned} \tag{2.424}$$

This is the result from the brute-force re-solution. But one can manipulate (2.424) into[57] (see Appendix 3),

$$\begin{aligned}
\tilde{\mathbf{x}}(2) =& \tilde{\mathbf{x}}(1) + \\
& \mathbf{P}(1)\mathbf{E}(2)^T \left[\mathbf{E}(2)\mathbf{P}(1)\mathbf{E}(2)^T + \mathbf{R}_{nn}(2)\right]^{-1} [\mathbf{y}(2) - \mathbf{E}(2)\tilde{\mathbf{x}}(1)] \\
=& \tilde{\mathbf{x}}(1) + \mathbf{K}(2)[\mathbf{y}(2) - \mathbf{E}(2)\tilde{\mathbf{x}}(1)], \qquad (2.425) \\
& \mathbf{P}(2) = \mathbf{P}(1) - \mathbf{K}(2)\mathbf{E}(2)\mathbf{P}(1), \qquad (2.426)
\end{aligned}$$

---

where

{recurs4}
$$\mathbf{K}(2) = \mathbf{P}(1)\mathbf{E}(2)^T \left[\mathbf{E}(2)\mathbf{P}(1)\mathbf{E}(2)^T + \mathbf{R}_{nn}(2)\right]^{-1}. \qquad (2.427)$$

(Compare Eq. (2.425) with (2.358).)

An alternate form, for $\mathbf{P}(2)$, found from the matrix inversion lemma, is

{recurs5}
$$\mathbf{P}(2) = \left[\mathbf{P}(1)^{-1} + \mathbf{E}(2)^T \mathbf{R}_{nn}(2)^{-1} \mathbf{E}(2)\right]^{-1}. \qquad (2.428)$$

A similar alternative for $\tilde{\mathbf{x}}(2)$, involving different dimensions of the matrices to be inverted, is also available from the matrix inversion lemma, but is generally less useful. (In some large problems, however, matrix inversion can prove less onerous than matrix multiplication.)

The solution (2.425) is just the least-squares solution to the full set, but rearranged after a bit of algebra. *The original data,* $\mathbf{y}(1)$*, and coefficient matrix,* $\mathbf{E}(1)$*, have disappeared, to be replaced by the first solution* $\tilde{\mathbf{x}}(1)$*, and its uncertainty,* $\mathbf{P}(1)$*. That is to say, one need not retain the original data and* $\mathbf{E}(1)$ *for the new solution to be computed.* Furthermore, because the new solution depends only upon $\tilde{\mathbf{x}}(1)$, and $\mathbf{P}(1)$, the particular methodology originally employed for obtaining them is irrelevant: they might even have been obtained from an educated guess, or from some long previous calculation of arbitrary complexity. If the initial set of equations (2.418) is actually underdetermined, and should it have been solved using the SVD, one must be careful that $\mathbf{P}(1)$ includes the estimated error owing to the missing nullspace. Otherwise, these elements would be assigned zero error variance, and the new data could never affect them. Similarly, the dimensionality and rank of $\mathbf{E}(2)$ is arbitrary, as long as the matrix inverse exists.

Example

Suppose we have a single measurement of a scalar, x*, so that ,* $x + n(1) = y(1), \langle n(1) \rangle = 0, \left\langle n(1)^2 \right\rangle = R(1)$*. Then an estimate of* x *is* $\tilde{x}(1) = y(1)$*, with uncertainty* $P(1) = R(1)$*. A second measurement then becomes available,* $x + n(2) = y(2), \langle n(2) \rangle = 0, \left\langle n(2)^2 \right\rangle = R(2)$*. By Eq. (2.425), an improved solution is*

$$\tilde{x}(2) = y(1) + R(1)/(R(1) + R(2))(y(1) - y(2)),$$

with uncertainty by Eq. (2.428),

$$P\left(2\right)=1/\left(1/R\left(1\right)+1/R(2)\right)=R(1)R(2)/(R\left(1\right)+R\left(2\right)).$$

If $R\left(1\right)=R\left(2\right)=R$*, we have* $\tilde{x}\left(2\right)=\left(y\left(1\right)+y\left(2\right)\right)/2, P\left(2\right)=R/2$*. If there are* M *succesive measurements all with the same error variance,* R*, one finds the last estimate is,*

$$\begin{aligned}\tilde{x}\left(M\right) &= \tilde{x}\left(M-1\right)+R/\left(M-1\right)\left(R/\left(M-1\right)+R\right)^{-1}y\left(M\right)\\ &= \tilde{x}\left(M-1\right)+\frac{1}{M}y\left(M\right)\\ &= \frac{1}{M}\left(y\left(1\right)+y\left(2\right)+...+y\left(M\right)\right),\end{aligned}$$

with uncertainty,

$$P\left(M\right)=\frac{1}{\left(\left(M-1\right)/R+1/R\right)}=\frac{R}{M},$$

the conventional average and its variance. Note that each new measurement is given a weight $1/M$ *relative to the average,* $\tilde{x}\left(M-1\right)$*, already computed from the previous* $M-1$ *data points.*

The structure of the improved solution (2.425) is also interesting and suggestive. It is made up of two terms: the previous estimate plus a term proportional to the difference between the new observations $\mathbf{y}(2)$, and *a prediction of what those observations should have been* were the first estimate the wholly correct one and the new observations perfect. It thus has the form of a "predictor-corrector." The difference between the prediction and the forecast can be called the "prediction error," but recall there is observational noise in $\mathbf{y}(2)$. The new estimate is a weighted average of this difference and the prior estimate, with the weighting depending upon the details of the uncertainty of prior estimate and new data. The behavior of the updated estimate is worth understanding in various limits. For example, suppose the initial uncertainty estimate is diagonal, $\mathbf{P}(1)=\Delta^2\mathbf{I}$. Then,

$$\mathbf{K}(2)=\mathbf{E}(2)^T\left[\mathbf{E}(2)\mathbf{E}(2)^T+\mathbf{R}_{nn}(2)/\Delta^2\right]^{-1}. \tag{2.429}$$ {38010}

If the observations are extremely accurate, the norm of $\mathbf{R}_{nn}(2)/\Delta^2$ is small, and if the second set of observations is full rank underdetermined,

$$\mathbf{K}(2)\longrightarrow\mathbf{E}(2)^T(\mathbf{E}(2)\mathbf{E}(2)^T)^{-1}$$

and

$$\begin{aligned}\tilde{\mathbf{x}}(2) &= \tilde{\mathbf{x}}(1)+\mathbf{E}(2)^T(\mathbf{E}(2)\mathbf{E}(2)^T)^{-1}[\mathbf{y}(2)-\mathbf{E}(2)\tilde{\mathbf{x}}(1)]\\ &= [\mathbf{I}-\mathbf{E}(2)^T(\mathbf{E}(2)\mathbf{E}(2)^T)^{-1}\mathbf{E}(2)]\tilde{\mathbf{x}}(1)+\mathbf{E}(2)^T(\mathbf{E}(2)\mathbf{E}(2)^T)^{-1}\mathbf{y}(2).\end{aligned} \tag{2.430}$$

Now, $[\mathbf{I}-\mathbf{E}(2)^T(\mathbf{E}(2)\mathbf{E}(2)^T)^{-1}\mathbf{E}(2)] = \mathbf{I}_N - \mathbf{V}\mathbf{V}^T = \mathbf{Q}_v\mathbf{Q}_v^T$, where $\mathbf{V}$ is the full-rank singular vector matrix for $\mathbf{E}(2)$,and it spans the nullspace of $\mathbf{E}(2)$ (see Eq. 2.290) The update thus replaces, in the the first estimate, all the structures given perfectly by the second set of observations, but retains those structures from the first estimate about which the new observations say nothing—a sensible result. At the opposite extreme, when the new observations are very noisy compared to the previous ones, $\left\|\mathbf{R}_{nn}/\Delta^2\right\| \to \infty$, $\|\mathbf{K}(2)\| \to 0$, and the first estimate is left unchanged.

The general case represents a weighted average of the previous estimate with elements found from the new data, with the weighting depending both upon the relative noise in each, and upon the structure of the observations relative to the structure of $\mathbf{x}$ as represented in $\mathbf{P}(1)$, $\mathbf{R}_{nn}(2)$, $\mathbf{E}(2)$. The matrix being inverted in (2.427) is the sum of the measurement error covariance
egammaerror} $\mathbf{R}_{nn}(2)$, and the error covariance of the "forecast" $\mathbf{E}(2)\tilde{\mathbf{x}}(1)$. To see this, let $\boldsymbol{\gamma}$ be the error component in $\tilde{\mathbf{x}}(1) = \mathbf{x}(1) + \boldsymbol{\gamma}$, which by definition has covariance $\langle\boldsymbol{\gamma}\boldsymbol{\gamma}^T\rangle = \mathbf{P}(1)$. Then the expected covariance of the error of prediction is $\langle\mathbf{E}(1)\boldsymbol{\gamma}\boldsymbol{\gamma}^T\mathbf{E}(1)^T\rangle = \mathbf{E}(1)\mathbf{P}(1)\mathbf{E}(1)^T$, which appears in $\mathbf{K}(2)$. Because of the assumptions (2.421), and $\left\langle\boldsymbol{\gamma}(1)\mathbf{x}(1)^T\right\rangle = \mathbf{0}$, it follows that,

{innov1} $$\langle\mathbf{y}(1)(\mathbf{y}(2)-\mathbf{E}(2)\tilde{\mathbf{x}}(1))\rangle = \mathbf{0}. \tag{2.431}$$

That is, the prediction error or "innovation," $\mathbf{y}(2)-\mathbf{E}(2)\tilde{\mathbf{x}}(1)$, is uncorrelated with the previous measurement.

It is useful to notice that Eq. (2.359), the solution to the least-squares problem subject to certain perfect constraints imposed by a Lagrange multiplier, can be recovered from the minimum variance solution (2.425) by putting $\mathbf{E}(2) = \mathbf{A}$, $\mathbf{y}(2) = \mathbf{b}$, $\mathbf{R}_{nn}(2) \to \mathbf{0}$. That is, this earlier solution can be conceived of as having been obtained by first solving the conventional least-squares problem, and then being modified by the *later* information that $\mathbf{A}\mathbf{x} = \mathbf{b}$, with very high accuracy.

The possibility of a recursion based on Eqs. 2.425, 2.426 (or 2.428) is obvious—all subscript 1 variables being replaced by subscript 2 variables, which in turn are replaced by subscript 3 variables, etc. The general form would be,

---

$$\tilde{\mathbf{x}}(n) = \tilde{\mathbf{x}}(n-1) + \mathbf{K}(n)[\mathbf{y}(n) - \mathbf{E}(n)\tilde{\mathbf{x}}(n-1)] \tag{2.432}$$

$$\mathbf{K}(n) = \mathbf{P}(n-1)\mathbf{E}(n)^T\left[\mathbf{E}(n)\mathbf{P}(n-1)\mathbf{E}(n)^T + \mathbf{R}_{nn}(n)\right]^{-1} \tag{2.433}$$

$$\mathbf{P}(n) = \mathbf{P}(n-1) - \mathbf{K}(n)\mathbf{E}(n)\mathbf{P}(n-1) \tag{2.434}$$

An alternative form for Eq. (2.434) is, from (2.428),

{recurs9}
$$\mathbf{P}(n) = \left[\mathbf{P}(n-1)^{-1} + \mathbf{E}(n)^T \mathbf{R}_{nn}(n)^{-1} \mathbf{E}(n)\right]^{-1}. \tag{2.435}$$

The computational load of the recursive solution needs to be addressed. A least-squares solution does *not* require one to calculate the uncertainty $\mathbf{P}$ (although the utility of $\tilde{\mathbf{x}}$ without such an estimate is unclear). But to use the recursive form, one must have $\mathbf{P}(n-1)$, otherwise the update step, Eq. (2.432) cannot be used. In very large problems, such as appear in oceanography and meteorology (Chapter 6), the computation of the uncertainty, from (2.434), or (2.435) can become prohibitive. In such a situation, one might simply store all the data, and do one large, single calculation—if this is feasible. Normally, it will involve less pure computation than will the recursive solution which must repeatedly update $\mathbf{P}(n)$.

The comparatively simple interpretation of the recursive, weighted least-squares problem will be used in Chapter 4 to derive the Kalman filter and suboptimal filters in a very simple form. It also becomes the key to understanding "assimilation" schemes such as "nudging," "forcing to climatology," and "robust diagnostic" methods.

2.8.2 Minimum Variance Recursive Estimates

The recursive least-squares result is identical to a recursive estimation procedure, if appropriate least-squares weight matrices were used. Suppose there exist two *independent* estimates of an unknown vector $\mathbf{x}$, denoted $\tilde{\mathbf{x}}_a$, $\tilde{\mathbf{x}}_b$ with estimated uncertainties $\mathbf{P}_a$, $\mathbf{P}_b$, respectively. They are either unbiased , or have the same bias, that is, $\langle\tilde{\mathbf{x}}_a\rangle = \langle\tilde{\mathbf{x}}_b\rangle = \mathbf{x}_B$ that is, have the same mean. How should the two be combined to give a third estimate $\tilde{\mathbf{x}}^+$ with minimum error variance? Try a linear combination,

$$\tilde{\mathbf{x}}^+ = \mathbf{L}_a\tilde{\mathbf{x}}_a + \mathbf{L}_b\tilde{\mathbf{x}}_b\,. \tag{2.436}$$ {38014}

If the new estimate is to be unbiased, or is to retain the prior bias (that is, the same mean), it follows that,

$$\langle\tilde{\mathbf{x}}^+\rangle = \mathbf{L}_a\langle\tilde{\mathbf{x}}_a\rangle + \mathbf{L}_b\langle\tilde{\mathbf{x}}_b\rangle \tag{2.437}$$ {38015}

or,

$$\mathbf{x}_B = \mathbf{L}_a\mathbf{x}_B + \mathbf{L}_b\mathbf{x}_B \tag{2.438}$$

or,

$$\mathbf{L}_b = \mathbf{I} - \mathbf{L}_a \tag{2.439}$$

Then the uncertainty is,

$$\begin{aligned}\mathbf{P}^+ &= \langle(\tilde{\mathbf{x}}^+ - \mathbf{x})(\tilde{\mathbf{x}}^+ - \mathbf{x})^T\rangle = \langle(\mathbf{L}_a\tilde{\mathbf{x}}_a + (\mathbf{I}-\mathbf{L}_a)\tilde{\mathbf{x}}_b)(\mathbf{L}_a\tilde{\mathbf{x}}_a + (\mathbf{I}-\mathbf{L}_a)\tilde{\mathbf{x}}_b)^T\rangle \\ &= \mathbf{L}_a\mathbf{P}_a\mathbf{L}_a^T + (\mathbf{I}-\mathbf{L}_a)\mathbf{P}_b(\mathbf{I}-\mathbf{L}_a)^T\end{aligned} \tag{2.440}$$

where the independence assumption has been used to set $\langle(\tilde{\mathbf{x}}_a-\mathbf{x})\,(\tilde{\mathbf{x}}_b-\mathbf{x})\rangle = \mathbf{0}$. $\mathbf{P}^+$ is positive definite; minimizing its diagonal elements with respect to $\mathbf{L}_a$ yields (after writing out the diagonal elements of the products),

$$\mathbf{L}_a = \mathbf{P}_b(\mathbf{P}_a+\mathbf{P}_b)^{-1}, \qquad \mathbf{L}_b = \mathbf{P}_a(\mathbf{P}_a+\mathbf{P}_b)^{-1}\,.$$

(Blithely differentiating and setting to zero produces the correct answer:

$$\frac{\partial\,(\mathrm{diag}\,\mathbf{P}^+)}{\partial \mathbf{L}_a} = \mathrm{diag}\left(\frac{\partial \mathbf{P}^+}{\partial \mathbf{L}_a}\right) = \mathrm{diag}\,[2\mathbf{P}_a\mathbf{L}_a - \mathbf{P}_b\,(\mathbf{I}-\mathbf{L}_a)] = 0,$$

or,$\mathbf{L}_a = \mathbf{P}_b(\mathbf{P}_a+\mathbf{P}_b)^{-1}$). The new combined estimate is,

{recurs15}
$$\tilde{\mathbf{x}}^+ = \mathbf{P}_b(\mathbf{P}_a+\mathbf{P}_b)^{-1}\tilde{\mathbf{x}}_a + \mathbf{P}_a(\mathbf{P}_a+\mathbf{P}_b)^{-1}\tilde{\mathbf{x}}_b\,. \tag{2.441}$$

This last expression can be rewritten by adding and subtracting $\tilde{\mathbf{x}}_a$ as,

$$\begin{aligned}\tilde{\mathbf{x}}^+ &= \tilde{\mathbf{x}}_a + \mathbf{P}_b(\mathbf{P}_a+\mathbf{P}_b)^{-1}\tilde{\mathbf{x}}_a \\ &\quad + \mathbf{P}_a(\mathbf{P}_a+\mathbf{P}_b)^{-1}\tilde{\mathbf{x}}_b - (\mathbf{P}_a+\mathbf{P}_b)(\mathbf{P}_a+\mathbf{P}_b)^{-1}\tilde{\mathbf{x}}_a \\ &= \tilde{\mathbf{x}}_a + \mathbf{P}_a(\mathbf{P}_a+\mathbf{P}_b)^{-1}(\tilde{\mathbf{x}}_b - \tilde{\mathbf{x}}_a)\,. \end{aligned} \tag{2.442}$$

Notice in particular, the re-appearance of a predictor-corrector form relative to $\tilde{\mathbf{x}}_a$.

The uncertainty of the estimate (2.442) is easily evaluated as

{38019b}
$$\mathbf{P}^+ = \mathbf{P}_a - \mathbf{P}_a(\mathbf{P}_a+\mathbf{P}_b)^{-1}\mathbf{P}_a\,. \tag{2.443}$$

or, by straightforward application of the matrix inversion lemma, is,

{38019a}
$$\mathbf{P}^+ = (\mathbf{P}_a^{-1}+\mathbf{P}_b^{-1})^{-1}. \tag{2.444}$$

The uncertainty is again independent of the observations. Eqs. (2.442-2.444) are the general rules for combining two estimates with uncorrelated errors.

Now suppose that $\tilde{\mathbf{x}}_a$ and its uncertainty are known, but that instead of $\tilde{\mathbf{x}}_b$ there are measurements,

{38020}
$$\mathbf{E}(2)\mathbf{x} + \mathbf{n}(2) = \mathbf{y}(2), \tag{2.445}$$

with $\langle\mathbf{n}(2)\rangle = 0$, $\langle\mathbf{n}(2)\mathbf{n}(2)^T\rangle = \mathbf{R}_{nn}(2)$. From this second set of observations, we *estimate* the solution, using the minimum variance estimator (2.404, 2.406) with no use of the solution

variance; that is, let $\|\mathbf{R}_{xx}^{-1}\| \rightarrow 0$. The reason for suppressing $\mathbf{R}_{xx}$, which logically could come from $\mathbf{P}_a$, is to maintain the independence of the previous and the new estimates. Then,

{38021a}
$$\tilde{\mathbf{x}}_b = \left[\mathbf{E}\,(2)^T\,\mathbf{R}_{nn}\,(2)^{-1}\,\mathbf{E}\,(2)\right]^{-1}\mathbf{E}\,(2)^T\,\mathbf{R}_{nn}(2)^{-1}\mathbf{y}\,(2) \tag{2.446}$$

{38021b}
$$\mathbf{P}_b = \left[\mathbf{E}\,(2)^T\,\mathbf{R}_{nn}\,(2)^{-1}\,\mathbf{E}\,(2)\right]^{-1}\,. \tag{2.447}$$

Substituting (2.446), (2.447) into (2.442), (2.443), and using the matrix inversion lemma (see Appendix 3) gives

$$\tilde{\mathbf{x}}^+ = \tilde{\mathbf{x}}_a + \mathbf{P}_a\mathbf{E}(2)^T[\mathbf{E}(2)\mathbf{P}_a\mathbf{E}(2)^T + \mathbf{R}_{nn}(2)]^{-1}(\mathbf{y}\,(2) - \mathbf{E}(2)\tilde{\mathbf{x}}_a)\,, \tag{2.448}$$

$$\mathbf{P}^+ = \left(\mathbf{P}_a^{-1} + \mathbf{E}\,(2)^T\,\mathbf{R}_{nn}\,(2)^{-1}\,\mathbf{E}\,(2)\right)^{-1} \tag{2.449}$$

which is the same as (2.432), 2.435) and thus *a recursive minimum variance estimate coincides with a corresponding weighted least-squares recursion.* The new covariance may also be confirmed to be that in either of Eqs. (2.434) or (2.435). Notice that if $\tilde{\mathbf{x}}_a$ was itself estimated from an earlier set of observations, that those data have disappeared from the problem, with all the information derived from them contained in $\tilde{\mathbf{x}}_a$ and $\mathbf{P}_a$. Thus, again, earlier data can be wholly discarded after use. It does not matter where $\tilde{\mathbf{x}}_a$ originated, whether from over- or under-determined equations or a pure guess—as long as $\mathbf{P}_a$ is realistic. Similarly, expression (2.448) remains valid whatever the dimensionality or rank of $\mathbf{E}\,(2)$ as long as the inverse matrix exists. The general implementation of this sequence for a continuing data stream corresponds to Eqs. (2.432)-(2.435).