RECURSIVE LEAST SQUARES ESTIMATION BY UPDATING CHOLESKY FACTORIZATION

Keijo Inkilä
Helsinki University of Technology
Finland
Commission III

Abstract

Two different types of methods for updating the Cholesky factorization are presented. The methods are compared in terms of computational efficiency and numerical stability with each other and the Kalman filter.

0. Introduction

As far as the number of papers is concerned the Kalman filter approach seems to have the dominating role in recursive least squares estimation also in photogrammetry. As. e.g. Gruen /3/ has pointed out there are however many estimation situations, where this approach is not the most appropriate one. The present paper is concerned with alternatives to the Kalman filter. Rather than updating the inverse, the Cholesky factor of the coefficient matrix of the normal equations is updated in these methods. This provides the basis for efficient parameter and accuracy estimation.

The paper attempts to give the fundamentals of the subject. Consequently, the main part of it is devoted to underlying mathematics and the special considerations for various applications are totally omitted. In preparing the paper the references /1/ and /4/ have served as primary source volumes.

1. Definition of the problem

Frequently, after the least squares solution of the system

$$v = Ax - 1, \text{ where } A \text{ is mxn design matric } (m>n), \quad 1 \text{ is m-vector (observations)} \qquad (1)$$

data is added (or deleted), which implies that the solution of the augmented system

$$\begin{bmatrix} v \\ v_z \end{bmatrix} = \begin{bmatrix} A \\ a^T \end{bmatrix} x - \begin{bmatrix} 1 \\ z \end{bmatrix} \quad \text{or} \quad \bar{v} = \bar{A}x - \bar{1} \qquad (2)$$

where a is a n-vector and z is a scalar, is required.

The batch solution by updating the normal equations, i.e., by forming and solving the system

$$(A^TA + aa^T)x = A^T1 + az \quad \text{or} \quad \bar{A}^T\bar{A} = \bar{A}^T\bar{1}$$

would approximately require the following number of multiplications

Updating of normal equations	$n^2/2$
Cholesky decomposition $\bar{R}^T\bar{R} = \bar{R}^T\bar{A}$	$n^3/6$
Solution of triangular systems $\bar{R}^Ty = \bar{A}^T\bar{1}$ and $\bar{R}_x = y$	n^2

Thus, the computation of the Cholesky factor takes the bulk of the computational work. This suggest that essential savings might be obtained, if the Cholesky factor were directly modified for new observational information. This is, indeed, the case and the primary objective of the present paper is to introduce efficient recursive solutions to the problem (2), which are based on updating the Cholesky factor.

2. Agee-Turner updating method

Suppose that we are given the Cholesky factor R satisfying

$$R^T R = A^T A \tag{3}$$

If we denote by $\bar{R}$ the updated Cholesky factor satisfying

$$\bar{R}^T \bar{R} = \bar{A}^T \bar{A} \tag{4}$$

then

$$\bar{R}^T \bar{R}^T = \bar{A}^T \bar{A} = \begin{bmatrix} A^T a \end{bmatrix} \begin{bmatrix} A \\ a^T \end{bmatrix} = A^T A + aa^T = R^T R + aa^T = R^T R + caa^T \tag{5}$$

where c=1 (The significance of c will become evident soon)

By introducing the partitioning

$$\begin{bmatrix} \bar{r}_{11} & 0 \\ \bar{r}_{12}^T & \bar{R}_{22}^T \end{bmatrix} \begin{bmatrix} \bar{r}_{11} & \bar{r}_{12} \\ 0 & \bar{R}_{22} \end{bmatrix} = \begin{bmatrix} r_{11} & 0 \\ r_{12}^T & \bar{R}_{22}^T \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ 0 & R_{22} \end{bmatrix} + c \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \begin{bmatrix} a_1 & a_2^T \end{bmatrix}$$

and performing the multiplications we have

$$\bar{r}_{11}^2 = r_{11}^2 + ca_1^2 \tag{6}$$

$$\bar{r}_{11}\bar{r}_{12} = r_{11}r_{12} + ca_1a_2^T \tag{7}$$

$$\bar{r}_{12}^T\bar{r}_{12} + \bar{R}_{22}^T\bar{R}_{22} = r_{12}^T r_{12} + R_{22}^T R_{22} + ca_2a_2^T \tag{8}$$

From (6) and (7) we can solve

$$\bar{r}_{11} = (r_{11}^2 + ca_1^2)^{\frac{1}{2}} \tag{9}$$

and

$$\bar{r}_{12} = (r_{11}r_{12} + ca_1a_2^T)/\bar{r}_{11} \tag{10}$$

Thus, the formulas for computing the first row of $\bar{R}$ have been derived. To continue it is crucial to observe that we can write

$$\bar{R}_{22}^T\bar{R}_{22} = R_{22}^T R_{22} + r_{12}^T r_{12} + c a_2 a_2^T - \bar{r}_{12}^T \bar{r}_{12}$$

$$= R_{22}^T R_{22} + r_{22}^T r_{12} + c a_2 a_2^T - (r_{11} r_{12} + c a_1 a_2^T)^T (r_{11} r_{12} + c a_1 a_2^T)/\bar{r}_{11}^2$$

$$= R_{22}^T R_{22} + r_{22}^T r_{12} + c a_2 a_2^T - (r_{11}^2 r_{12}^T r_{12} + 2 c r_{11} a_1 a_2 r_{12} + c^2 a_1^2 a_2 a_2^T)/\bar{r}_{11}^2$$

$$= R_{22}^T R_{22} + \left(\frac{\bar{r}_{11}^2 - r_{11}^2}{\bar{r}_{11}^2}\right) r_{12}^T r_{12} - \left(\frac{2 c a_1 r_{11}}{\bar{r}_{11}^2}\right) a_2 r_{12} + \left(\frac{c \bar{r}_{11}^2 - c^2 a_1}{\bar{r}_{11}^2}\right) a_2 a_2^T$$

$$= R_{22}^T R_{22} + \left(\frac{c a_1^2}{\bar{r}_{11}^2}\right) r_{12}^T r_{12} - \left(\frac{2 c a_1 r_{11}}{\bar{r}_{11}^2}\right) a_2 r_{12} + \left(\frac{c r_{11}^2}{\bar{r}_{11}^2}\right) a_2 a_2^T$$

$$= R_{22}^T R_{22} + c(r_{11}/\bar{r}_{11})^2 \left[a_2 - (a_1/r_{11}) r_{12}^T\right]\left[a_2 - (a_1/r_{11}) r_{12}^T\right]^T$$

By defining

$$\bar{c} = c(r_{11}/\bar{r}_{11})^2 \tag{11}$$

and

$$\bar{a}_2 = a_2 - (a_1/r_{11}) r_{12}^T \tag{12}$$

we finally have

$$R_{22}^T R_{22} = R_{22}^T R_{22} + \bar{c} \bar{a}_2 \bar{a}_2^T \tag{13}$$

which is of the same form as (5) but the dimension is reduced by one. This result implies that the updated Cholesky factor can be computed recursively in n-1 steps. The main ingredients of the algorithm are the formulas (9)-(12).

By examining the formulas it is immediately seen that the data deletion case is obtained by simply putting $c = -1$ in the beginning.

3. Updating by using orthogonal transformations

3.1 Solution of the least squares problem by using the orthogonal decomposition

A theorem of linear algebra states that for any real mxn-matrix A with m>n and rank (A)=n there exist an orthogonal mxm-matrix Q such that

$$QA = \begin{bmatrix} R \\ 0 \end{bmatrix} \quad \text{or} \quad A = Q^T \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{14}$$

where R is an nxn upper triangular matrix with rank (R)=n. Because

$$A^TA = \begin{bmatrix} R^T 0^T \end{bmatrix} Q\, Q^T \begin{bmatrix} R \\ 0 \end{bmatrix} = R^TR$$

and $R^TR = R^TD^TDR = R'^TR'$, where D is a diagonal with diagonal elements +1 or -1, we observe that R of the QR decomposition and the Cholesky decomposition differ, if any, by signs of the rows.

The computation of the QR decomposition is discussed in the next chapter. In this chapter it is presented how least squares problems are solved by using it. We start by examining the solution of the system (1) given the QR decomposition of the design matrix A. By defining

$$Ql = \begin{bmatrix} c \\ d \end{bmatrix} \begin{matrix} \} \; n \\ \} \; m-n \end{matrix} \tag{15}$$

and recalling that the orthogonal transformation preserves the Euclidean length of a vector we have

$$S(x) = v^Tv = \|v\|^2 = \|Ax - l\|^2 = \|Q(Ax - l)\|^2 = \|QAx - Ql\|^2$$
$$= \left\| \begin{bmatrix} R \\ 0 \end{bmatrix} x - \begin{bmatrix} c \\ d \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} Rx - c \\ d \end{bmatrix} \right\|^2 = \|Rx - c\|^2 + \|d\|^2 \tag{16}$$

This implies that

(1) the least squares solution, $\hat{x}$, is obtained by solving the upper triangular system

$$Rx = c \tag{17}$$

(2) the residual sum of squares is

$$S(\hat{x}) = \hat{v}^T\hat{v} = \|d\|^2 = d^Td \tag{18}$$

(3) the residual vector $\hat{v}$ can, alternatively, be computed by using the formula

$$\hat{v} = Q^T \begin{bmatrix} 0 \\ d \end{bmatrix} \tag{19}$$

Note that , if desired, (18) and (19) can be computed before solving x.

The solution of the least squares problem using the orthogonal decomposition is numerically more accurate than the solution via normal equations, but it requires, on the other hand, about twice as many arithmetic operations. This is obviously the main reason why QR decomposition is infrequently applied to the solution of the batch least squares problem. We, however, are interested here in the solution of the recursive least squares problem.

Let us first consider the data addition case, i.e., the solution of the system (2). By exploiting (16) we first have

$$S(x) = \|v\|^2 = \left\| \begin{bmatrix} A \\ a^T \end{bmatrix} x - \begin{bmatrix} 1 \\ z \end{bmatrix} \right\|^2 = \|Ax - 1\|^2 + \|a^T x - z\|^2$$

$$= \|Rx - c\|^2 + \|d\|^2 + \|a^T x - z\|^2 = \left\| \begin{bmatrix} R \\ a^T \end{bmatrix} x - \begin{bmatrix} c \\ z \end{bmatrix} \right\|^2 + \|d\|^2 \qquad (20)$$

Referring to (14) we now assume that

$$Q \begin{bmatrix} R \\ a^T \end{bmatrix} = \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix} , \qquad (21)$$

where $\bar{R}$ is an nxn upper triangular matrix and define

$$Q \begin{bmatrix} c \\ z \end{bmatrix} = \begin{bmatrix} \bar{c} \\ e \end{bmatrix} \qquad (22)$$

Then we can reduce (20) to the form

$$S(x) = \left\| \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix} x - \begin{bmatrix} \bar{c} \\ e \end{bmatrix} \right\|^2 + \|d\|^2 = \|\bar{R}x - \bar{c}\|^2 + \|d\|^2 + \|e\|^2 \qquad (23)$$

The updated least squares estimate thus satisfies the equation

$$\bar{R}x = \bar{c} \qquad (24)$$

and the corresponding residual sum of squares is

$$\|d\|^2 + \|e\|^2 = d^T d + e^T e \qquad (25)$$

By introducing the imaginary unit $i^2 = -1$ very similar derivation for the data (row) deletion case results.

$$S(x) = \|Ax - 1\|^2 - \|a^T x - z\|^2 = \|Rx - c\|^2 + \|d\|^2 - \|a^T x - z\|^2$$

$$= \left\| \begin{bmatrix} Rx - c \\ i(a^T x - z) \end{bmatrix} \right\|^2 + \|d\|^2 = \left\| \begin{bmatrix} R \\ ia^T \end{bmatrix} x - \begin{bmatrix} c \\ iz \end{bmatrix} \right\|^2 + \|d\|^2 = \left\| \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix} x - \begin{bmatrix} \bar{c} \\ ie \end{bmatrix} \right\|^2 + \|d\|^2$$

$$= \left\| \begin{bmatrix} \bar{R}x - \bar{c} \\ ie \end{bmatrix} \right\|^2 + \|d\|^2 = \|\bar{R}x - \bar{c}\|^2 + \|d\|^2 - \|e\|^2 \qquad (26)$$

It should be emphasized that the algorithms for the two cases differs only by signs of some quantities from each other.

Summing up the results of the present chapter we observe that the sequential processing of this type essentially involves the reduction of the certain augmented matrix to triangular form by orthogonal transformation, that is, the key operation can be expressed as

$$Q\begin{bmatrix} R & c \\ a^T & z \end{bmatrix} = \begin{bmatrix} \bar{R} & \bar{c} \\ 0 & e \end{bmatrix} . \qquad (28)$$

It remains now to construct the required orthogonal matrix. In the construction we explicitly take into consideration the special structure of the matrix to be reduced.

3.2 Construction of the Householder and Givens transformations

Denote
$$T = \begin{bmatrix} t_1 & t_2 & \dots & t_{n+1} \end{bmatrix} = \begin{bmatrix} R & c \\ a^T & z \end{bmatrix} \qquad (29)$$

and note that the columns t_j are of the special form

$$t_j^T = \begin{bmatrix} t_j(1), & t_j(2), & \dots & t_j(j), & 0, & \dots, 0, & t_j(n+1) \end{bmatrix}^T \qquad (30)$$

If we define

$$s = \sigma(t_j(j)^2 + t_j(n+1)^2)^{\frac{1}{2}}, \text{ where } \sigma = \begin{cases} +1, & \text{if } t_j(j) < 0 \quad *) \\ -1, & \text{if } t_j(j) > 0 \end{cases}$$

$$u_j(i) = 0 \quad , \ i = 1, \dots, j-1$$

$$u_j(j) = t_j(j) - s$$

$$u_j(i) = t_j(i), \ i = j+1, \dots, n+1 \qquad (31)$$

$$Q_j = I_{n+1} - \frac{2}{u_j^T u_j} u_j u_j^T \qquad \text{(Orthogonal for any } u \neq 0 \text{ !)}$$

then by observing that

$$Qt = \left(I - \frac{2}{u^T u} uu^T \right) t = t - 2\frac{u^T t}{u^T u} u = t - \gamma u \qquad (32)$$

it is easily verified that

(i) $$Q_j t_k = \begin{bmatrix} t_j(1), & t_j(2), \dots, t_j(j-1), & s, & 0, \dots, 0 \end{bmatrix}^T$$

(:. Last element is zeroed)

(ii)
$$Q_j t_k = \begin{bmatrix} t_j(1), t_j(2), \dots, t_j(j-1), t_j(j) - \gamma u_j(j), 0, \dots, 0, t_j(n+1) - \gamma t_j(n+1) \end{bmatrix}^T, \text{if } k>j$$

(:. Only the j^{th} and $(n+1)^{th}$ element are changed. The structure of the vector t_k does not alter)

(iii) $Q_j(Q_k t_k) = Q_k t_k$, if $k<j$

(∴ The columns already "zoeroed" remain unchanged in subsequent transformations)

Based on properties (i) - (iii) and the observation that

$$Q_j T = Q_j\left[t_1,\ t_2,\ldots,t_{n+1}\right] = \left[Q_j t_1,\ Q_j t_2,\ldots,\ Q_j t_{n+1}\right]$$

it is obvious that the (n+1)x(n+1) matrix T can be reduced to upper triangular form by n successive transformations, that is,

$$Q\begin{bmatrix} R & c \\ a^T & z \end{bmatrix} = \begin{bmatrix} \bar{R} & \bar{c} \\ 0 & e \end{bmatrix} \tag{33}$$

where $Q = Q_n Q_{n-1}\ldots Q_2 Q_1$

The transformation defined in (31) is called the elementary Householder transformation. An alternative is the Givens transformation, whose construction is now briefly discussed.
Let t_j be the same as in (30). If we define

$$c = \frac{t_j(j)}{(t_j(j)^2 + t_j(n+1)^2)^{\frac{1}{2}}}, \qquad s = \frac{t_j(n+1)}{(t_j(j)^2 + t_j(n+1)^2)^{\frac{1}{2}}}$$

$$Q_j = \begin{bmatrix} I_{j-1} & 0 & 0 & 0 \\ 0 & c & 0 & s \\ 0 & 0 & I_{n-j-1} & 0 \\ 0 & -s & 0 & c \end{bmatrix} \tag{34}$$

then it can be verified by simple calculations that the properties (i) - (iii) also hold for the Givens transformation matrix defined in (34). Hence, the (n+1)x(n+1)- matrix T can alternatively be reduced to upper triangular form by n successive Givens transformations Q_j (j= 1,..., n)

From the computational point of view it is of crucial importance to observe that whether the reduction is based on the Householder or Givens transformation the explicit computation of the transformation matrix is not needed for the solution of the least squares problem. Rather we exploit directly the given properties of the matrix-vector-product of these transformations in the construction of algorithms.

4. Discussion of the presented methods

Conserning all the updating methods presented above the introduction of weights is straightforward: the vector a_i and the scalar z_i are simply multiplied by $\sqrt{p_i}$ before processing, where p_i is the weight of the observation z_i. In contrast, the treatment of a group of correlated observations is much more involved: corresponding design matrix and observation vector must then be multiplied by the Cholesky factor of the weight matrix /2/.

The computational efficiency of the methods depends of course to some extent on algorithmic details, but the numbers of arithmetic operations given below are anyway typical. The numbers refer to the updating of the nxn Cholesky factor only.

Method	Additions	Multiplications	Divisions	Square roots
Agee-Turner	n^2	$2n^2$	2n	n
Givens	n^2	$2n^2$	2n	n
Householder	$2n^2$	$5n^2/2$	n	n

Recalling that the computation of the Cholesky factorization requires e.g. about $n^3/6$ multiplications we find that an order of magnitude fewer operations are required when the updating algorithms are used. On the other hand, there are only minor differences between the three algorithms.

With reference to the Householder algorithm additional computational savings are attained, if more than one observation equation is processed at the same time. For p observations we have the following table.

Processing	Additions	Multiplications	Divisions	Square roots
one by one	$2pn^2$	$5pn^2/2$	pn	pn
simultaneously	$(p+1)n^2$	$(p+3/2)n^2$	n	n

The examination of (31) gives easily the modifications required for processing several observations simultaneously.

Because the solution of two triangular systems also requires about n^2 multiplications we observe that the total number of multiplications for the new estimate is of the order of n^2 only, that is, of the same order as with the Kalman filter.

With regard to numerical accuracy, consideration should be paid to the following aspects / 1 , 4 /:

- Orthogonal transformations are well-known for their favourable numerical properties. Moreover, the recursive algorithms being based on them operate directly on design matrix A, the condition number of which is the square root of the condition number of A^TA. These algorithms should, therefore, be preferred in poorly conditioned problems.
- The deletion of data can be an inherently unstable operation. This is seen immediately e.g. from (9), where if the numbers are close in magnitude digit cancellation occurs in subtraction.
- Compared with the Kalman filter the methods presented above are expected to give significantly more reliable solutions.

Finally, it is easy to show /2/, that Q of (28) transforms $(R^T)^{-1}$ to $(\bar{R}^T)^{-1}$, i.e. the inverse $(R^T)^{-1}$ can be updated simultaneously with R. By means of $\bar{R}^{-1}$ we can further compute the inverse of $\bar{A}^T\bar{A}$, for $(\bar{A}^T\bar{A})^{-1} = \bar{R}^{-1}(\bar{R}^T)^{-1}$.

5. Addition or removal of parameters

The change in the number of observation equations frequently results in the change in the number of parameters too. We discuss in this chapter, how Cholesky factor is modified in such a case.

Let us first consider the addition of a parameter. The design matrix is then of the form

$$\tilde{A} = \begin{bmatrix} A & b \end{bmatrix} \quad \text{, where b is the added colums} \tag{35}$$

If

$$\tilde{R} = \begin{bmatrix} \bar{R} & \bar{r} \\ 0 & \alpha \end{bmatrix} \text{, where } \bar{r} \text{ is a n-vector and } \alpha \text{ is a scalar} \tag{36}$$

is the corresponding Cholesky factor we have

$$\tilde{R}^T\tilde{R} = \tilde{A}^T\tilde{A}$$

or

$$\begin{bmatrix} \bar{R}^T & 0 \\ \bar{r}^T & \alpha \end{bmatrix} \begin{bmatrix} \bar{R} & \bar{r} \\ 0 & \alpha \end{bmatrix} = \begin{bmatrix} A^T \\ b^T \end{bmatrix} \begin{bmatrix} A & b \end{bmatrix} \tag{37}$$

or

$$\begin{bmatrix} \bar{R}^T\bar{R} & \bar{R}^T\bar{r} \\ \bar{r}^T\bar{R} & \bar{r}^T\bar{r}+\alpha^2 \end{bmatrix} = \begin{bmatrix} A^TA & A^Tb \\ b^TA & b^Tb \end{bmatrix}$$

Thus,

$$\begin{aligned} \bar{R}^T\bar{R} &= A^TA = R^TR \\ \bar{R}^T\bar{r} &= A^Tb \\ \alpha^2 &= b^Tb - \bar{r}^T\bar{r} \end{aligned} \tag{38}$$

i.e., the solution of a triangular system is essentially required (number of multiplications is proportional to n^2.

The derivation for the parameter removal case is rather similar.

Le us write the mxn design matrix $\tilde{A}$ in the partioned form

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 & \tilde{A}_3 \end{bmatrix} \tag{39}$$

where $\tilde{A}_2$ represents the k^{th} column of A to be removed. After removing we thus have mx(n-1) design matrix

$$\bar{A} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_3 \end{bmatrix} \tag{40}$$

If $\tilde{R}$ and $\bar{R}$ are the corresponding Cholesky factors, that is,

$$\tilde{R}^T\tilde{R} = \tilde{A}^T\tilde{A} \quad \text{and} \quad \bar{R}^T\bar{R} = \bar{A}^T\bar{A} \qquad \begin{matrix} (41) \\ (42) \end{matrix}$$

and they are partitioned corresponding to (39) and (40) as

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix} \quad \text{and} \quad \bar{R} = \begin{bmatrix} \bar{R}_{11} & \bar{R}_{12} \\ 0 & \bar{R}_{22} \end{bmatrix} \tag{43}$$

we find by expanding (41) and (42) and examining resulting expressions that

$$\begin{aligned} \bar{R}_{11} &= R_{11} \\ \bar{R}_{12} &= R_{13} \\ \bar{R}_{22}^T\bar{R}_{22} &= A_3^T A_3 - \bar{R}_{12}^T\bar{R}_{12} \end{aligned} \tag{44}$$

Thus, instead of the factorization of the (n-1)x(n-1)- matrix, the factorisation of the (n-k)x(n-k)- matrix is required. The reduction in computational effort of course depends crucially on how close k is to n.

The modifications on the right-hand side for adding or removing a parameter are obvious and are therefore omitted here.

REFERENCES

/1/ Bierman, G. I. : Factorization Methods for Discrete Sequential Estimation. Academic Press, New York, 1977.

/2/ Blais, J. A. R.: Recursive Least Squares Estimation Using Givens Transformations. ISP Comm. III Symposium, Helsinki, 1982.

/3/ Gruen, A.: An Optimum Algorithm for On-Line Triangulation. ISP Comm. III Symposium, Helsinki, 1982.

/4/ Lawson, C. L., Hanson, R. J.: Solving Least Squares Problems. Prentice-Hall, New Jersey, 1974.