A New Recursive Least-Squares Algorithm based on Matrix Pseudo Inverses (ICCAS 2003)

Zhonghua Quan*, Soohee Han, and Wookhyun Kwon

*School of Electrical Engineering & Computer Science, Seoul National University, Seoul , 151-742, Korea
(Tel: +82-2-880-7314; http://cisl.snu.ac.kr; Email:jonghwa@cisl.snu.ac.kr)

Abstract: In this paper, a new Recursive Least-Squares(RLS) algorithm based on matrix pseudo-inverses is presented. The aim is to use the proposed new RLS algorithm for not only the over-determined but also the under-determined estimation problem. Compared with previous results, e.g., Jie Zhou *et al.*, the derivation of the proposed recursive form is much easier, and the recursion form is also much simpler. Furthermore, it is shown by simulations that the proposed RLS algorithm is more efficient and numerically stable than the existing algorithms.

Keywords: Estimation, Recursive least-squares estimation, Pseudo-inverse

1. INTRODUCTION

Estimation problem deals with the determination of some unknown variables or parameters that cannot be measured from known variables that can be measured. For the last thirty years, the estimation problem has been widely investigated in control and signal processing areas. The least-squares(LS) approach, tracing back to Gauss and Legendre, is certainly the most widely used in system identification and in many other estimation problems.

Consider the generic linear model:

$$y_k = H_k\theta + v_k, \qquad k = 1, 2, ... \tag{1}$$

where $\theta \in \Re^r$ is the unknown random vector to be estimated, which is independent of v_k, a white noise with zero mean and covariance denoted by R. Then, the estimation $\hat{\theta}$ can be computed by using the following batch formula of least-squares:

$$\hat{\theta}_k = (\bar{H}_k^T\bar{W}_k\bar{H}_k)^{-1}\bar{H}_k^T\bar{W}_k\bar{Y}_k \tag{2}$$

where

$$\bar{H}_k = \begin{bmatrix} H_1 \\ . \\ . \\ . \\ H_k \end{bmatrix} = \begin{bmatrix} \bar{H}_{k-1} \\ H_k \end{bmatrix}, \bar{Y}_k = \begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_k \end{bmatrix} = \begin{bmatrix} \bar{Y}_{k-1} \\ y_k \end{bmatrix},$$

$$\bar{W}_k = diag\{W_1, ..., W_k\} = \begin{bmatrix} \bar{W}_{k-1} & \\ & W_k \end{bmatrix} \tag{3}$$

and the weighting matrices $W_i,\ i = 1, 2, ..., k$, are all positive definite.

In adaptive controllers measurements are obtained periodically in real time. It is desirable to make the calculations recursively to save computation time. The parameter estimate $\hat{\theta}_k$ may be obtained as a function of the previous estimate $\hat{\theta}_{k-1}$ and of the new measurements.

In 1950, a standard RLS based on matrix inverse lemma was proposed by Plackett, in which $\hat{\theta}_k(\forall k \geq k_0)$ in Eq.(2) can be written recursively as[1]:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_{k+1}(y_{k+1} - H_{k+1}\hat{\theta}k) \tag{4}$$

$$K_{k+1} = P_kH_{k+1}^T/(1 + H_{k+1}P_kH_{k+1}^T) \tag{5}$$

$$\begin{aligned} P_{k+1} &\overset{def}{=} (\bar{H}_{k+1}^T\bar{H}_{k+1})^{-1} \\ &= (P_k{}^{-1} + H_{k+1}^TH_{k+1})^{-1} \\ &= (I - P_kH_{k+1}^TH_{k+1}/(1 + H_{k+1}P_kH_{k+1}^T))P_k \\ &= (I - K_{k+1}H_{k+1})P_k \end{aligned} \tag{6}$$

The equations Eqs.(4)∼(6) can be initialized by setting $\hat{\theta}_0 = 0$ and P_0 equal to a diagonal matrix of very large numbers in practice. The standard RLS algorithm is widely used in parameter estimation and adaptive control because of its conceptual simplicity, ease of implementation. This recursive least-squares solution greatly promotes the applications of LS method in many fields where real time processing is required. The two significant advantages of the recursive solution in Eqs.(4)∼(6) are (i) free of the matrix inverse operation and reducing the computational complexity for $\hat{\theta}_k$ in Eq.(2); (ii) very suitable to computerize the on-line processing since the number of algebraic operations and required memory locations at each iteration are fixed and not increasing as k is getting large and large as the batch LS solution in Eq.(2).

Although the RLS solution in Eqs.(4)∼(6) has the above advantages, the existence of the inverse of P_k^{-1} in Eq.(6) is a necessary condition. Thus, the data matrix $\bar{H}_k$ must have full column rank. However, in the numerical computations, even if $\bar{H}_k$ has full column rank, it is possible for $\bar{H}_k^T\bar{H}_k$ to be singular, i.e., the RLS is numerically unstable. For example, we consider the following matrix:

$$H = \begin{pmatrix} 1 & 1 & 1 \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \tag{7}$$

When ϵ is a constant close to the machine precision, $1 + \epsilon^2 \approx 1$. Thus,

$$\bar{H}_k^T\bar{H}_k = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{8}$$

This paper is supported by BK21.

becomes singular to working precision, and we can not compute via matrix inverse as done in Eq.(2) or Eqs.(4)∼(6).

This results in poor robustness of the least-squares algorithm especially when implemented in computers with finite accuracy. In order to achieve more robust numerical performance, matrix pseudo-inverses(Moore-Penrose generalized inverses) were used in the derivation of RLS algorithm. Due to the computation of matrix pseudo-inverses involves an increasing number of variables with a corresponding increase in the matrix order, a recursive form of pseudo-inverses were introduced in the previous work. In [2]-[5], a recursive form of pseudo-inverse is presented, which turns out to be computationally demanding. Jie Zhou *et al.* proposed a variant of the Greville formula as well as the recursive form of RLS in [6]-[7]. This variant reduces the storage requirements at each recursion by almost half and more convenient for deriving recursive solutions for optimization problems involving pseudo-inverses. However, this approach as well as Greville formula may also result in the divergence of $\hat{\theta}$ in numerical computations because of the finite machine precision. So, a new recursive least-squares approach is proposed which can overcome the divergence problem with inheritance of the advantages of the existing algorithm.

This paper is organized as follows. In section 2, previous results are introduced with the analysis of the shortcomings and the new recursive least-squares algorithm is proposed in section3. The simulation results are given in section 4 and conclusions are stated in section 5.

2. Previous Results

Consider the system denoted in Eq.(1), then the estimate $\hat{\theta}$ is the minimum-norm solution of the optimization problem

$$min\|\bar{H}_k\theta - \bar{Y}_k\|_{\bar{W}_k} \tag{9}$$

where $\bar{H}_k$, $\bar{Y}_k$ and the weighting matrices $\bar{W}_i,\ i=1,...,k$, are denoted in Eq.(3).

Theorem 1: Consider a system of linear equations $Ax = b, A\in\Re^{m\times n}$, $rankA = r$. The vector $x^* = A^+b$ minimizes $\|Ax-b\|^2$. Furthermore, among all vectors that minimizes $\|Ax-b\|^2$, the vector $x^* = A^+b$ is the unique vector with minimum norm.

Let $\bar{H}_k^+$ denotes the Pseudo-inverse(Moore-Penrose inverse) of $\bar{H}_k$, then the unweighted LSE of θ is

$$\hat{\theta}_k = \bar{H}_k^+\bar{Y}_k. \tag{10}$$

To give the recursive form of Eq.(10), we introduce Greville's result on representation of the generalized inverse of a partitioned matrix.

Lemma 1:

$$\overline{H}_k^+ = [M_k\overline{H}_{k-1}^+ \qquad G_k] \tag{11}$$

where $k\geq 1$, G_k and M_k are determined as:

$$\begin{aligned}
Q_{k-1} &= I - \overline{H}_{k-1}^+\overline{H}_{k-1}, & A_k &= H_kQ_{k-1},\\
C_k &= I - A_kA_k^+, & P_{k-1} &= \overline{H}_{k-1}^+\overline{H}_{k-1}^{+T},\\
B_k &= H_kP_{k-1}, & D_k &= I + C_kB_kH_k^TC_k,\\
G_k &= A_k^+ + (I - A_k^+H_k)B_k^TD_k^{-1}C_k, &&\\
M_k &= I - G_kH_k &&
\end{aligned} \tag{12}$$

$$P_k = M_kP_{k-1}M_k^T + G_kG_k^T, \qquad Q_k = M_kQ_{k-1} \tag{13}$$

with the initial values $P_0 = 0,\ \ Q_0 = I$.
Thus the equation Eq.(10) can be rewritten as:

$$\begin{aligned}
\hat{\theta}_k &= M_k\hat{\theta}_{k-1} + G_ky_k\\
&= (I - G_kH_k)\hat{\theta}_{k-1} + G_ky_k\\
&= \hat{\theta}_{k-1} + G_k(y_k - H_k\hat{\theta}_{k-1}))
\end{aligned} \tag{14}$$

where

$$G_k = \begin{cases} A_k^+ & \text{if } A_k \neq 0\\ B_k^T(I + B_kH_k^T)^{-1} & \text{if } A_k = 0\end{cases} \tag{15}$$

The above formula is the transpose of the Greville formula in its original form. While the two versions are equivalent, this version fits our formation of the problem better. It can be noted that the form of $\hat{\theta}$ in Eq.(14) is similar with that in Eq.(4).
Using this recursive formula, one can recursively compute the pseudo-inverse $\overline{H}_k^+$ of a high order matrix $\overline{H}_k$ from vector H_1^+ so as to be free of the computation for a high order matrix pseudo-inverse. It is, however, not in a form handy for deriving the recursive versions of the optimal solutions involving the above matrix pseudo-inverse. In light of this, Jie Zhou *et al.* proposed the following variant of Greville's formula, which can be viewed as an improvement of the Greville formula.
Lemma 2:

$$\begin{aligned}
\overline{H}_1^+ &= G_1,\\
\overline{H}_{k+1}^+ &= (I - G_{k+1}H_{k+1} \quad G_{k+1})\begin{pmatrix}\overline{H}_k^+ & 0\\ 0 & 1\end{pmatrix}\\
k &\geq 1
\end{aligned} \tag{16}$$

where G_{k+1} is defined by the following:
$(i)When\ \ A_k = 0$,

$$\begin{aligned}
G_{k+1} &= P_kH_{k+1}^T/(1 + H_{k+1}P_kH_{k+1}^T)\\
P_{k+1} &= (I - G_{k+1}H_{k+1})P_k\\
Q_{k+1} &= Q_k
\end{aligned} \tag{17}$$

$(ii)When\ \ A_k \neq 0$,

$$\begin{aligned}
G_{k+1} &= Q_kH_{k+1}^T/(H_{k+1}Q_kH_{k+1}^T)\\
P_{k+1} &= (I - G_{k+1}H_{k+1})P_k(I - G_{k+1}H_{k+1}^T) + G_{k+1}G_{k+1}^T\\
Q_{k+1} &= (I - G_{k+1}H_{k+1})Q_k
\end{aligned} \tag{18}$$

and the initial values are

$$P_0 = 0, \quad Q_0 = I.$$

The cases with $A_k = 0$ and $A_k \neq 0$ in *lemma* 2 can be combined to the following:

$$\begin{aligned} G_{k+1} &= A_k^+ + (1 - H_k A_k^+) P_k H_{k+1}^T / (1 + H_{k+1} P_k H_{k+1}^T) \\ P_{k+1} &= (I - G_{k+1} H_{k+1}) P_k (I - G_{k+1} H_{k+1})^T + G_{k+1} G_{k+1}^T \\ Q_{k+1} &= (I - G_{k+1} H_{k+1}) Q_k \end{aligned} \quad (19)$$

and the initial values are

$$P_0 = 0, \quad Q_0 = I.$$

Not only do the formulas in *lemma* 2 reduce the required memory locations of the Greville formula at each recursion by almost half, but they are also very convenient to derive some recursive algorithms for the optimal solutions involving matrix pseudo-inverses. This formula as well as the Greville formula, however, will result in the divergence in the numerical computations under finite machine precision because the pseudo-inverse is ill-conditioned with respect to rank-changing perturbations. For example,

$$\left[\begin{pmatrix} 1 & -1 \\ 2 & -2 \end{pmatrix} + \varepsilon \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \right]^+ = \begin{cases} \dfrac{1}{\varepsilon^2} \begin{pmatrix} -1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{pmatrix} + \dfrac{1}{\varepsilon} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} & \text{if } \varepsilon \neq 0 \\ \dfrac{1}{10} \begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix} & \text{if } \varepsilon = 0 \end{cases} \quad (20)$$

So it can be concluded that the less computations of pseudo-inverses, the more the algorithm is numerically stable. Although there is the case with $A_k = 0$ theoretically in the above recursive form, however, A_k is not precisely identical with zero due to the finite machine precision. Thus A_k^+ will be diverge because of the ε which is a constant close to the machine precision. As a result of the above reason, the $\bar{H}^+$ will diverge, and so is $\hat{\theta}_k$. To overcome this drawback, a new recursive least-squares algorithm based on pseudo-inverse is proposed by the following in the current paper .

3. Description of the new RLS algorithm

The well known recursive least-squares provides an optimal parameter estimate $\hat{\theta}(k)$ as follows:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + P_{k+1} H_{k+1}^T W_{k+1} [y_{k+1} - H_{k+1} \hat{\theta}_k] \quad (21)$$

$$P_{k+1}^{-1} = P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1} \quad (22)$$

From Eq.(21) and Eq.(22),

$$P_{k+1}^{-1} \hat{\theta}_{k+1} = P_{k+1}^{-1} \hat{\theta}_k + H_{k+1}^T W_{k+1} [y_{k+1} - H_{k+1} \hat{\theta}_k] \quad (23)$$

$$P_{k+1}^{-1} = P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1} \quad (24)$$

Substituting Eq.(24) into the right side of Eq.(23), Eq.(23) can be rewritten as:

$$\begin{aligned} P_{k+1}^{-1} \hat{\theta}_{k+1} = (P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1}) \hat{\theta}_k \\ + H_{k+1}^T W_{k+1} [y_{k+1} - H_{k+1} \hat{\theta}_k] \end{aligned} \quad (25)$$

$$P_{k+1}^{-1} \hat{\theta}_{k+1} = P_k^{-1} \hat{\theta}_k + H_{k+1}^T W_{k+1} y_{k+1} \quad (26)$$

To define the initial values $\hat{\theta}_0$ and $\hat{P}_0$ clearly and avoid singularity problem, a new recursive least-squares algorithm is proposed. Let define

$$\hat{\eta}_k = P_k^{-1} \hat{\theta}_k \quad (27)$$

Then the new recursive least-squares algorithm can be described as the following:

$$\hat{\theta}_k = [P_k^{-1}]^+ \hat{\eta}_k \quad (28)$$

$$P_{k+1}^{-1} = P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1} \quad (29)$$

$$\hat{\eta}_{k+1} = \hat{\eta}_k + H_{k+1}^T W_{k+1} y_{k+1} \quad (30)$$

where $\hat{\eta}_0 = 0$ and $P_0^{-1} = 0$.
Due to the use of pseudo-inverse in Eq.(28), it need not guarantee the nonsingulariry of the matrix P_k^{-1}. If the matrix P_k^{-1} is invertible, then Eq.(28) is identical with the following equation:

$$\hat{\theta}_k = P_k \hat{\eta}_k \quad (31)$$

When $\bar{H}_k$ is not full column rank, then from the following normal equation

$$(\bar{H}_k^T \bar{H}_k) \hat{\theta}_k = \bar{H}_k^T \bar{Y}_k \quad (32)$$

We can obtain

$$\hat{\theta}_k = (\bar{H}_k^T \bar{H}_k)^+ \bar{H}_k^T \bar{Y}_k \quad (33)$$

instead of Eq.(2). Because the equation Eq.(31) obtained from Eq.(33), so it should be shown that $\bar{H}^+$ and $(\bar{H}^T \bar{H})^+ \bar{H}^T$ are equivalent , from which it can be shown that the proposed algorithm is equivalent with the existing recursive algorithm.
Theorem 2: $\bar{H}^+$ can be written as

$$\bar{H}^+ = (\bar{H}^T \bar{H})^+ \bar{H}^T \quad (34)$$

Proof: If $(\bar{H}^T \bar{H})^+ \bar{H}^T$ satisfy the following equation:

$$\bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H} = \bar{H} \quad (35)$$

$$(\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T = (\bar{H}^T \bar{H})^+ \bar{H}^T \quad (36)$$

$$((\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H})^T = (\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H} \quad (37)$$

$$(\bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T)^T = \bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T \quad (38)$$

We can say that $(\bar{H}^T \bar{H})^+ \bar{H}^T$ is the pseudo-inverse of $\bar{H}$. To check the first requirement (35), it should be shown that the following equations are satisfied:

$$\Leftrightarrow \bar{H} (\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H} = \bar{H} \quad (39)$$

$$\Leftrightarrow \bar{H} \{ (\bar{H}^T \bar{H})^+ \bar{H}^T \bar{H} - I \} = 0 \quad (40)$$

We try to prove Eq.(40). $\bar{H}^T\bar{H}\{(\bar{H}^T\bar{H})^+\bar{H}^T\bar{H}-I\}=0$ is satisfied from the definition of $(\bar{H}^T\bar{H})^+$ so that we can see that all columns of $\bar{H}\{(\bar{H}^T\bar{H})^+\bar{H}^T\bar{H}-I\}$ should be orthogonal with those of $\bar{H}$ and simultaneously consist of linear combinations of columns of $\bar{H}$. Therefore, $\bar{H}\{(\bar{H}^T\bar{H})^+\bar{H}^T\bar{H}-I\}$ is equal to a zero matrix.

The second and third requirements Eqs.(36)$\sim$ (37) are satisfied from the definition $(\bar{H}^T\bar{H})^+$. The final fourth requirement is an identical equation.

If $\bar{H}$ is of maximum rank, then $[\bar{H}^T\bar{H}]^{-1}$ exists and Eq.(33) is identical with Eq.(2) when the unweighted case of $\bar{W}=I$. This completes the proof. ■

From the above *Theorem* 2, we can conclude that the recursive algorithm based on the Greville's as well as Jie Zhou's formula and the proposed recursive algorithm are equivalent.

4. Simulation Results

Measurement data used in the simulation are generated by the following linear equation:

$$Y(k)=5.2*U(k)+2.7*V(k)-3.2*W(k)+e(k) \quad (41)$$

where $e(k)$ is zero mean white noise with standard deviation 0.5. The *Fig*.1 and *Fig*.2 show the simulation results used different set of data respectively. It can be seen from the two figures that the parameters can be estimated correctly by using the proposed recursive least-squares algorithm. Further, it is also can be noted that the proposed RLS algorithm is more stable than the existing algorithm based on recursive pseudo-inverses.

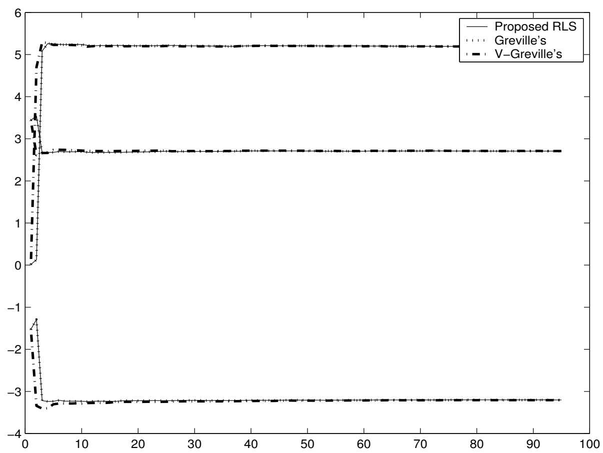

Fig. 1. Simulation Results(1)

Table 1. Comparison of Computation time

algorithm	Greville's	V-Greville's	Proposed RLS
elapsed time(s)	0.7433	0.2644	0.2399

The *Fig*.3 is also a result of simulation, in which the existing algorithm based on recursive pseudo-inverses resulted in the divergence but not the proposed algorithm. When A_k

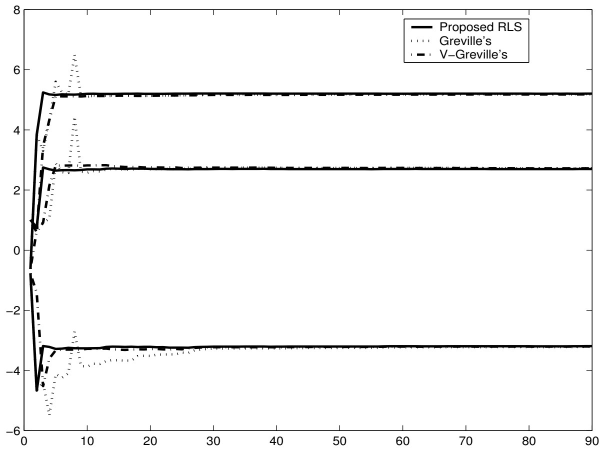

Fig. 2. Simulation Results(2)

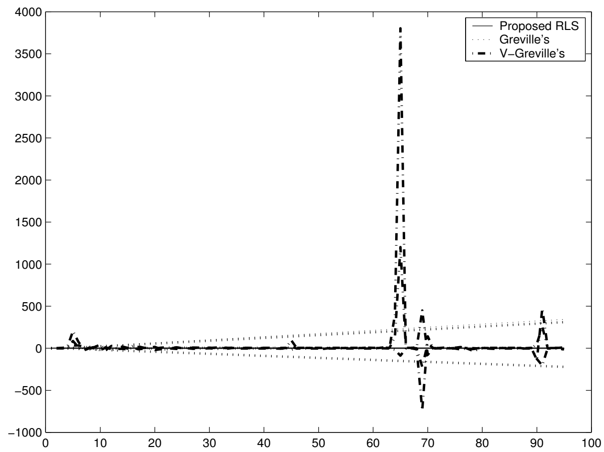

Fig. 3. Simulation Results(3)

approaches to zero, but not exactly identical with zero in numerical computations because of the finite machine precision. Thus, A_k^+ will diverge to infinity as well as $\bar{H}^+$, so do the $\hat{\theta}_k$. The *Fig*.3 supports the conclusion that the proposed RLS is more stable than other two algorithms mentioned in current paper.

The computation time spent in the simulation, in which a set of data consists of 1000 data, is listed in *Table* 1 for three different algorithms respectively. From the *Table* 1, it is concluded that the proposed algorithm is simpler with less computation time than the existing algorithm based on recursive pseudo-inverses.

5. Conclusion

In this paper, a new recursive least-squares algorithm based on pseudo-inverse is proposed. From the analysis and simulations it is shown that the proposed algorithm is simpler with numerical stability and accuracy than previous results of Greville and Jie Zhou *et al.*. By using the proposed algorithm, the initial values $\hat{\theta}_0$ and P_0 are defined clearly and the singularity problem also is avoided. Since the derivation of the recursive formulas of $\hat{\theta}_k$ proposed in this paper is based on matrix pseudo-inverses not on matrix inverses, it still can deal with not only the conventional RLS problem, but also the problem of rank deficient matrix $\bar{H}$.

References

[1] Jerry M. Mendel, *" Lessons in Estimation Theory for Signal Processing, Communications, and Control "*. NJ:Prentice-Hall, 1995

[2] Liu Danyang and Liu Xuanhuang, *"Technical Notes and Correspondence"*. IEEE Transactions on Automatic Control, Vol.39, NO.10, pp.2087-2091, Oct.1994.

[3] Randall E. Cline, *"Representations for the generalized inverse of a partitioned matrix"*. J.Soc.Indust.Appl.Math, Vol.12, NO.3, pp.588-600, Sep.1964.

[4] T.N.E.Greville, *"Some applications of the pseudoinverse of a matrix"*. SIAM Review, Vol.2, NO.1, pp.15-22, Jan.1960.

[5] T.N.E.Greville, *"The pseudoinverse of a rectangular or singular matrix and its application to the solution of systems of linear equations"*. SIAM Review, Vol.1, NO.1, pp.38-43, Jan.1959.

[6] Jie Zhou, X.Rong Li, Zhisheng You, *" Exactly Initialized Recursive Least Squares "*. Proceedings of the 40th IEEE Conference on Decision and Control, pp.3318-3323, 2001.

[7] Jie Zhou, Yunmin Zhu, X.Rong Li, Zhisheng You, *" Variants of the greville formula with applications to exact recursive least squares "*. SIAM J. Matrix ANAL. APPL., Vol.24, NO.1, pp.150-164, 2002.

[8] SooHee Han,*" Best linear unbiased estimation filters with FIR structures for state space models "*. Master Thesis, Seoul National University, Jan.2000.