Brief Paper

A directional forgetting algorithm based on the decomposition of the information matrix

Liyu Cao*, Howard Schwartz

Department of Systems and Computer Engineering, Carleton University, 1125 Colonel By Drive, Ottawa, Canada K1S 5B6

Received 8 December 1998; revised 13 August 1999; received in final form 5 April 2000

Abstract

A novel algorithm for directional forgetting is proposed based on a matrix decomposition method, which is developed in this paper. This algorithm performs exponential forgetting according to the direction of the data vector, thus preventing the problem known as estimator windup which is a characteristic of the standard exponential forgetting algorithm. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Recursive algorithms; Least-squares method; Parameter estimation; Exponential forgetting; Directional forgetting

1. Introduction

The recursive least-squares (RLS) algorithm with a forgetting factor has been widely used as a basic method for on-line parameter estimation of linear time-varying systems. This algorithm can be described by

$$\hat{y}(t) = \hat{y}(t-1) + P(t)\phi(t)[y(t) - \phi^T(t)\hat{y}(t-1)],$$  (1)

$$R(t) = \mu R(t-1) + \phi(t)\phi^T(t),$$  (2)

where $\phi(t)$ is the data vector measured at $t$, $R(t)$ is called the information matrix, and $P(t) = R^{-1}(t)$ is the covariance matrix. The scalar $\mu \in (0,1)$ is the forgetting factor. By using the forgetting factor, the old information is discounted according to an exponential function. Therefore, this algorithm is also called the exponential forgetting (EF) RLS algorithm.

The EF algorithm behaves well when the data sequence $\{\phi(t)\}$ is persistently exciting and therefore contains sufficient information about the system dynamics. When $\{\phi(t)\}$ is not persistently exciting, then it does not carry sufficient information. The old data is discounted continuously but only a part of the old data can be replaced by $\phi(t)$. As a consequence, some eigenvalues of $R(t)$ will tend to zero and the algorithm gain $P(t)\phi(t)$ will tend to be unbounded. This phenomenon is known as estimator windup (Åström & Wittenmark, 1995). Estimator windup is undesirable because it means the algorithm becomes very sensitive to noise and thus the estimation may be completely unreliable.

The reason for estimator windup in the EF method is that the old data is forgotten regardless of whether the forgotten data can be replaced by the new data. A more suitable forgetting philosophy should be to forget old data only when it can be replaced by new data. This philosophy led to the so-called directional forgetting algorithm (DF) suggested by Kulhavý and Kárný (1984), Hägglund (1985) and Kulhavý (1987). In the DF algorithm, the data is considered to have directions, and the old data is forgotten only in some specified direction. However, how to implement the directional forgetting remains a question. Kulhavý and Kárný (1984), and furthermore Kulhavý (1987) proposed a directional forgetting algorithm based on the Bayesian estimation approach. Although Kulhavý’s algorithm can prevent estimator windup, a problem recognized in this algorithm is that some of the eigenvalues of $R(t)$ may become unbounded (Bittanti, Bolzern & Campi, 1990), which means that the algorithm may lose its tracking capability in some directions.

*This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor A. Vicino under the direction of Editor T. Söderström.

* Corresponding author. Tel.: +1-613-520-2600 ext 8281; fax: +1-613-520-5727.

E-mail address: cao@sce.carleton.ca (L. Cao).

0005-1098/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PH: S0005-1098(00)00093-5
In this paper, a new directional forgetting algorithm is proposed based on a matrix decomposition method developed by the authors (refer to Lemma 1 in this paper). The motivation for this method can be explained as follows. As has been pointed out above, estimator windup in the EF algorithm is due to the fact that the forgetting is applied uniformly to all elements of $R(t-1)$ (refer to (2)). In other words, the problem is in the fact that the forgetting is performed through the scalar $\mu$. This fact naturally motivates one to replace $\mu$ by a matrix, that is, to replace (2) by the following update equation:

$$R(t) = F(t) R(t - 1) + \varphi(t) \varphi^T(t), \quad (3)$$

where $F(t)$ is called the forgetting matrix. $F(t)$ should be able to discount $R(t - 1)$ according to the direction of $\varphi(t)$ in some sense. Obviously, with a suitable $F(t)$, it is possible to adjust the various eigenvalues of $R(t - 1)$ with different scaling, and therefore, to overcome the drawback of the EF method. Eq. (3) can be seen as a general representation of various forgetting strategies, including the EF algorithm. Eq. (3) can be rewritten as

$$R(t) = \tilde{R}(t - 1) + \varphi(t) \varphi^T(t), \quad (4)$$

$$\tilde{R}(t - 1) = F(t) R(t - 1). \quad (5)$$

The matrix $\tilde{R}(t - 1)$ is called the modified information matrix, and is obtained after the forgetting operation on $R(t - 1)$. The matrix $F(t)$ should be selected in such a way that $\tilde{R}(t - 1)$ is positive definite and $\tilde{R}(t - 1) \leq R(t - 1)$. These conditions ensure that $R(t)$ remains positive definite for all $t$, and furthermore the old data is forgotten (discounted) after the modification.

There are many possible ways to select the forgetting matrix $F(t)$, depending on different forgetting strategies.

Kreisselmeier (1990) proposed the concept of the stabilized RLS algorithm, and gave two choices of $F(t)$ to implement the stabilized RLS algorithm, which require to calculate the inverse of some matrix. In this paper, we propose a method to determine the forgetting matrix based on the orthogonal decomposition of $R(t)$ along the exciting direction, which leads to a new directional forgetting algorithm. It is shown that through the matrix decomposition, the directional forgetting can be implemented in an effective way. In this approach, only a specified part of the information matrix is forgotten at each update. Even when excitation is poor, windup does not occur because nothing is forgotten in the part of the information matrix, which is orthogonal to the excited space in some sense.

2. Decomposition of information matrix

As stated in the previous section, forgetting should not be applied uniformly to the information matrix. For this propose, before forgetting, $R(t - 1)$ is divided into two parts

$$R(t - 1) = R_1(t - 1) + R_2(t - 1) \quad (6)$$

and then forgetting is applied only to one part, say $R_2(t - 1)$. It remains how to determine $R_1(t - 1)$ and $R_2(t - 1)$. The following approach is motivated by the results in Aström and Wittenmark (1995). We extend their work by proposing a complete decomposition theory.

The matrix $R_1(t - 1)$ is required to satisfy the following equation:

$$R_1(t - 1) \varphi(t) = 0, \quad \varphi(t) \neq 0, \quad (7)$$

which means that $\varphi(t)$ is in the kernel space of $R_1(t - 1)$. In the other words, the image space of $R_1(t - 1)$ is orthogonal to $\varphi(t)$. Therefore, by letting $R_1(t - 1)$ satisfy (7), we establish an orthogonal relationship between $R_1(t - 1)$ and $\varphi(t)$.

From (7), one can find $R_2(t - 1)$ satisfying

$$R_2(t - 1) \varphi(t) = R(t - 1) \varphi(t). \quad (8)$$

With (7) and (8), $R_1(t - 1)$ and $R_2(t - 1)$ cannot be determined uniquely. One needs some restriction on the rank of $R_1(t - 1)$ and $R_2(t - 1)$. The new data is coming in the form $\varphi(t) \varphi^T(t)$ whose rank is 1, therefore it is reasonable to require that the forgotten part, $R_2(t - 1)$ has rank 1. With this rank restriction on $R_2(t - 1)$, one can get a unique $R_2(t - 1)$ that satisfies (8). This is demonstrated by the following lemma.

Lemma 1. Assume that $R(t - 1)$ is positive definite for $t \geq 1$. Furthermore, assume that $R_2(t - 1)$ is nonnegative definite and has rank 1. Then matrix equation (8) has a unique solution which is given by

$$R_2(t - 1) = \frac{1}{\varphi^T(t) R(t - 1) \varphi(t)} [R(t - 1) \varphi(t)] [R(t - 1) \varphi(t)]^T. \quad (9)$$

Proof. See the appendix.

Let

$$\alpha(t) = \frac{1}{\varphi^T(t) R(t - 1) \varphi(t)}. \quad (10)$$

Then $R_1(t - 1)$ is obtained from (6)

$$R_1(t - 1) = R(t - 1) - \alpha(t) [R(t - 1) \varphi(t)] [R(t - 1) \varphi(t)]^T. \quad (11)$$
As shown in Cao and Schwartz (1999) $R_1(t)$ is positive semidefinite and its rank is $n - 1$ providing that $R(t - 1)$ is positive definite.

In order to ensure that $\phi(t)$ is well-defined, it is obvious from (10) that $\phi(t)$ must be a nonzero vector. In fact, to ensure the algorithm is well-behaved, a dead zone is needed for $\phi(t)$. The decomposition is not performed unless $\phi(t)$ is outside the dead zone. Therefore, when $\phi(t)$ is within the dead zone, we let

$$\phi(t) = 0 \text{ if } |\phi(t)| \leq \varepsilon,$$

(12)

where $\varepsilon$ can be determined based on the noise level in data. In the following, it is assumed that $|\phi(t)| > \varepsilon$ unless the opposite situation is mentioned.

Applying exponential forgetting only to $R_2(t - 1)$, the update equation for the information matrix becomes

$$R(t) = R_1(t - 1) + \mu R_2(t - 1) + \phi(t)\phi^T(t).$$

(13)

One can see that the new information matrix, $R(t)$ consists of three parts: $\phi(t)\phi^T(t)$ represents the new data, $R_1(t - 1)$ represents the old data that is orthogonal to $\phi(t)$, and $R_2(t - 1)$ represents the old data that should be forgotten.

Comparing Eq. (13) with Eq. (4) one can see that the modified information matrix is given by

$$\tilde{R}(t - 1) = (I - M(t))R(t - 1),$$

(14)

where $M(t)$ is defined by

$$M(t) = (1 - \mu)\tilde{\phi}(t)R(t - 1)\tilde{\phi}(t)^T.$$  

(15)

Therefore, the forgetting matrix is given by

$$F(t) = I - M(t).$$

(16)

As shown in Cao and Schwartz (1999), one of the eigenvalues of $F(t)$ is $\mu$, and all of the rest eigenvalues are 1.

It is desirable to preserve the positive-definite property of the information matrix after making a modification on it. The proposed forgetting strategy has this ideal property. This is shown by the following lemma.

**Lemma 2.** If $R(t)$ is positive definite and $0 < \mu \leq 1$, then $\tilde{R}(t)$ is also positive definite.

**Proof.** Refer to Cao and Schwartz (1999).

Noting that $M(t)R(t - 1)$ is nonnegative definite, then from (14) we get

$$\tilde{R}(t - 1) \leq R(t - 1).$$

(17)

Inequality (17) indicates that the old information is definitely forgotten (discounted) at each update.

As a brief summary, the update equations for $R(t)$ are given as

$$R(t) = [I - M(t)]R(t - 1) + \phi(t)\phi^T(t),$$

(18)

$$M(t) = (1 - \mu) R(t - 1)\phi(t)\phi^T(t) / \phi^T(t)R(t - 1)\phi(t), \quad |\phi(t)| > \varepsilon,$$

(19)

$$M(t) = 0, \quad |\phi(t)| \leq \varepsilon.$$  

(20)

To start the algorithm, it is required that $R(0) > 0$. To complete the algorithm, the inverse of $R(t)$ is needed, which is obtained by applying the matrix inversion Lemma to (4) as follows:

$$P(t) = \tilde{P}(t - 1) - \tilde{P}(t - 1)\phi(t)\phi^T(t)\tilde{P}(t - 1) / [1 + \phi^T(t)\tilde{P}(t - 1)\phi(t)],$$

(21)

where $\tilde{P}(t - 1)$ is the modified covariance matrix and is defined by the following equation:

$$\tilde{P}(t - 1) = R^{-1}(t - 1) = P(t - 1)F^{-1}(t).$$

(22)

Applying the matrix inverse lemma to $F(t)$ and using Eq. (15), one can get the following equation for $|\phi(t)| > \varepsilon$

$$\tilde{P}(t - 1) = P(t - 1) + 1 - \mu \mu \phi^T(t) / \phi^T(t)P^{-1}(t - 1)\phi(t).$$

(23)

For $|\phi(t)| \leq \varepsilon$, $\tilde{P}(t - 1)$ is defined by

$$\tilde{P}(t - 1) = P(t - 1).$$

(24)

Finally, we will give a brief comparison between the proposed algorithm and the original version of the directional forgetting algorithms suggested by Hägglund (1985) and Kulhavý (1987). Although their starting point is different, their update equations have basically the same structure and can be represented by the following equation (Bittanti et al., 1990):

$$R(t) = R(t - 1) + \beta(t)\phi(t)\phi^T(t),$$

(25)

where $\beta(t)$ is a scalar and defined by

$$\beta(t) = \begin{cases} \mu(1 - \mu)/r(t) & \text{if } r(t) > 0, \\ 1 & \text{if } r(t) = 0, \end{cases}$$

(26)

where $\mu < 1$ and $r(t)$ is defined by

$$r(t) = \phi^T(t)R^{-1}(t - 1)\phi(t).$$

(27)

For convenience, we will call this the HK forgetting algorithm. Comparing (25) with (4), one can see that the update equations are quite different. Eq. (4) means that old information in $R(t - 1)$ is forgotten through the matrix $F(t)$ at each update. On the other hand, the meaning of Eq. (25) is not explicit because $\beta(t)$ may be negative (Kulhavý, 1987). If we rewrite Eq. (25) as

$$R(t) = R(t - 1) - (1 - \beta(t))\phi(t)\phi^T(t) + \phi(t)\phi^T(t),$$

(28)
then one can see that since \( \beta(t) \leq 1 \), information is subtracted from \( R(t-1) \) in the direction of \( \varphi(t) \) at each update. While in the proposed algorithm, information is subtracted in the direction of \( R(t-1)\varphi(t) \) (refer to (18)). More important, as indicated by Bittanti et al. (1990), the HK algorithm does not guarantee that \( R(t) \) is upper bounded. While as will be shown in the next section, in the proposed algorithm \( R(t) \) is upper bounded.

### 3. Boundedness of the information matrix

It has been recognized for a long time that one of the most desirable properties for a recursive parameter estimation algorithm is that the information matrix is bounded from above and below (Salgado, Goodwin & Middleton, 1988). It has been proven in Parkum, Poulsen and Holst (1992) that this property ensures that (i) estimation errors are bounded, (ii) normalized prediction errors are square summable and (iii) incremental changes in parameter estimates tend to zero. In the following, it is shown that the proposed algorithm has these desirable properties.

The following theorem shows that the information matrix is bounded from below.

**Theorem 1.** Consider the update Eqs. (18) and (19). If \( R(0) \) is positive definite, then for bounded \( \{\varphi(t)\} \) we have

\[
R(t) > \beta I \quad \text{for all } t,
\]

where \( \beta > 0 \).

To prove this theorem, some preliminary lemmas are needed. The first one is an extension of Lemma 1 to multi-dimensional subspace.

**Lemma 3.** Given an \( n \times n \) positive-definite matrix \( B \) and an \( n \times m \) matrix \( V \) with rank \( m, m < n \). Then \( B \) can be decomposed into

\[
B = A + C,
\]

where both \( A \) and \( C \) are nonnegative definite, and they satisfy the following equations:

\[
AV = BV, \quad CV = 0.
\]

**Brief proof.** One can see that the following matrices:

\[
A = BV(V^T BV)^{-1} V^T B, \quad C = B - A
\]

are the required ones. From the construction of \( A \), one can see that \( A \) is nonnegative definite. To show \( C \) is also nonnegative definite, consider the matrix \( D \) defined by

\[
D = AB^{-1} = BV(V^T BV)^{-1} V^T.
\]

We have \( D^2 = D \). Therefore, \( D \) is an idempotent matrix and its eigenvalues are all equal to zero or one. Then according to Theorem 7.7.3 in Horn and Johnson (1985), the matrix \( C = B - A \) is positive semidefinite.

The second preliminary lemma determines the ranks of \( A \) and \( C \).

**Lemma 4.** The rank of \( A \) given by (33) is \( m \), and the rank of \( C \) given by (34) is \( n - m \).

**Proof.** Refer to Lemma 2.4 of Chu, Funderlic and Golub (1998).

Now we are ready to prove Theorem 1.

**Proof of Theorem 1.** First, we assume that \( \{\varphi(t)\} \) is persistently exciting, that is, there exists an integer \( s \geq n \) and a positive number \( a \) such that

\[
\sum_{i=1}^{t+s} \varphi(i)\varphi^T(i) > aI,
\]

where \( n \) is the dimension of \( \varphi(t) \).

In this case, we use the approach of Johnstone, Johnson, Bitmead and Anderson (1982) to prove that the information matrix is bounded from below. For this purpose, consider the following matrix:

\[
D(t) = (I - M(t))R(t-1) - \mu R(t-1).
\]

It can be shown that the spectral radius of \( \mu R(t-1) \)

\[
[I - M(t)]^{-1} = 1. \text{ Then according to Theorem 7.7.3 in Horn and Johnson (1985), the matrix}
\]

\( D(t) \) is positive semidefinite. Using this fact, one can find that

\[
R(t) = (I - M(t))R(t-1) + \varphi(t)\varphi^T(t)
\]

\[
\geq \mu R(t-1) + \varphi(t)\varphi^T(t).
\]

Therefore, we have the following inequality:

\[
R(t+s) \geq \mu R(t+s-1)
\]

\[
\geq \mu^2 R(t+s-2) \geq \cdots \geq \mu^s R(t),
\]

where \( s \) is defined in (36). Furthermore, we can get

\[
R(t+s) + \mu^{-1} R(t+s) + \cdots + \mu^{-s} R(t+s)
\]

\[
\geq R(t+s) + R(t+s-1) + \cdots + R(t) > aI.
\]
From the above inequality, we can get for $t \geq s$,
\[ R(t) > \frac{a(1 - 1/\mu)}{1 - 1/\mu^t}. \quad (41) \]

Next, consider the situation where $|\phi(t)|$ is not persistently exciting. Then there exists an excited subspace and an unexcited subspace in $\mathbb{R}^n$. Let the excited subspace be $\phi$ and have dimension $m$, and the unexcited subspace be $\phi^\perp$, which is the orthogonal complement of $\phi$.

The updated Eq. (18) can be rewritten as
\[ R(t) = R(t - 1) - (1 - \mu)\varsigma(t)R(t - 1)\phi(t)\phi^T(t)R(t - 1) + \phi(t)\phi^T(t). \quad (42) \]

By applying Lemma 3 to $R(t-1)$ for $t = 1, 2, \ldots$ and using (42) recursively, we can get the following equations:
\[ R(t) = R_\phi(t) + R_p(t), \quad (43) \]
\[ R_\phi(t) = R_\phi(t - 1) - R_\phi(t - 2) = \cdots = R_\phi(0), \quad (44) \]
\[ R_p(t) = R_p(t) - (1 - \mu)\varsigma(t)R_p(t - 1)\phi(t)\phi^T(t) \]
\[ + \phi(t)\phi^T(t), \quad (45) \]
where $R_\phi(0)$ satisfies
\[ R_\phi(0)\phi(t) = 0 \quad \text{for all } \phi(t) \in \phi \quad (46) \]
and is positive semidefinite with rank $n - m$ according to Lemma 4. Define the following matrix:
\[ U = [U_1 \ U_2], \quad (47) \]
where $U_1$ is an $n \times m$ matrix whose columns are the orthonormal basis of the excited subspace $\phi$, and $U_2$ is an $n \times (n - m)$ matrix whose columns are the orthonormal basis of the unexcited subspace $\phi^\perp$. $U$ is an orthogonal matrix. One can get
\[ U^T\tilde{R}_\phi(0)U = \begin{bmatrix} 0 & 0 \\ 0 & U_2^T\tilde{R}_\phi(0)U_2 \end{bmatrix}, \quad (48) \]
\[ \psi(t)\phi^T(t)U = \begin{bmatrix} \psi(t)\phi^T(t) \\ 0 \\ 0 \end{bmatrix}, \quad (49) \]
\[ \psi(t) = U_1^T\phi(t) \]
is an $m \times 1$ column vector. It can be shown that $\psi(t)$ is persistently exciting.

Therefore, we have
\[ S(t) = U^TR(t)U \geq \begin{bmatrix} 0 & 0 \\ 0 & U_2^T\tilde{R}_\phi(0)U_2 \end{bmatrix}, \quad (50) \]
\[ + \begin{bmatrix} \mu U_1^T\tilde{R}_p(t - 1)U_1 + \psi(t)\psi^T(t) & 0 \\ 0 & 0 \end{bmatrix}. \quad (51) \]

The columns of $U_2$ belong to $\phi^\perp$, therefore $R_\phi(0)U_2 \neq 0$ and $U_2^T\tilde{R}_\phi(0)U_2$ is positive definite. Inequality (56) proves that as $t \to \infty$ all eigenvalues of $S(t)$ and hence $R(t)$ are bounded from below by a positive number.

The following theorem shows that under the sufficient and bounded excitation, the information matrix is also bounded from above.

**Theorem 2.** Assume $\varsigma < |\phi(t)| \leq c$. Then there exists a constant $\gamma > 0$ such that
\[ R(t) \leq \gamma I \quad \text{for all } t. \quad (57) \]

**Proof.** Rewrite updated Eq. (42) as
\[ R(t) = R(t) - (1 - \mu)S_t + \sum_{i=1}^{t} \phi(i)\phi^T(i), \quad (58) \]
where $S_t$ is defined by
\[ S_t = \sum_{i=1}^{t} \varsigma(i)R(i - 1)\phi(i)\phi^T(i)R(i - 1). \quad (59) \]
The trace of $S_t$ is given by
\[
\text{trace}[S_t] = \sum_{i=1}^{t} \varphi(i)R^2(i-1)\varphi(i) \tag{60}
\]
On the other hand, one can get
\[
\text{trace}[R(t)] = \text{trace}[R(0)] + \sum_{i=1}^{t} \varphi(i)R(i-1)\varphi(i) \tag{61}
\]
From these equations, one can get the trace of $R(t)$ as
\[
\text{trace}[R(t)] = \text{trace}[R(0)] + \sum_{i=1}^{t} \varphi(i)R(i-1)\varphi(i), \tag{62}
\]
where $R(i-1)$ is a symmetric matrix and is defined by
\[
R(i-1) = [(\varphi(i))^2 I - (1 - \mu)R(i-1)]R(i-1) \tag{63}
\]
Let the $j$th eigenvalue of $R(i-1)$ be $\rho_{i-1,j}$, and $j$th eigenvalue of $R(i-1)$ be $\lambda_{i-1,j}$. Then it can be shown that
\[
\rho_{i-1,j} = [\varphi(i)]^2 - (1 - \mu)\lambda_{i-1,j} = \lambda_{i-1,j}. \tag{64}
\]
From Eq. (64) it can be shown that it is impossible that the trace of $R(t)$ becomes unbounded. To make a conflict with this, assume that one of the eigenvalues of $R(t)$ is unbounded. Let this eigenvalue be $\lambda_{s,t}$, where $1 \leq s \leq n$. Then $\lambda_{s,t} \rightarrow \infty$ as $t \rightarrow \infty$. Thus, from (64) it can be seen that $\rho_{i,s}$ is unbounded, and when $i$ is sufficiently large, $\rho_{i-1,s}$ will become negative. Furthermore, when $i$ is sufficiently large the second term in (62) is almost dominated by the following ratio:
\[
\frac{\rho_{i-1,s}}{\lambda_{i-1,s}} = [\varphi(i)]^2 - (1 - \mu)\lambda_{i-1,s}. \tag{65}
\]
Therefore, the second term in (62) will also become negative and unbounded. As a consequence, the trace of $R(t)$ will tend to be negative, which is in conflict with the fact $R(t) > 0$. Then we conclude that the trace and hence the eigenvalues of $R(t)$ are bounded from above.

4. Conclusion

In this paper, a new directional forgetting algorithm based on the decomposition of the information matrix has been developed. Theoretical studies have shown that this algorithm has desirable properties, such as to forget old data according to the information content in various directions, above and below bounded information matrix. Compared with various modifications to the EF algorithm, the new algorithm is simple in the sense that there are only two adjustable parameters that can be easily prespecified. These properties make this algorithm an attractive selection for on-line identification.

Appendix. Proof of Lemma 1

Let $B$ be a positive-definite matrix, and let $A$ be non-negative definite and of rank 1. Then $A$ can be written in the form
\[
A = xx^T, \tag{A.1}
\]
where $x$ is an eigenvector associated with the nonzero eigenvalue of $A$. If $A$ satisfies the following matrix equation
\[
Bv = Av, \tag{A.2}
\]
where $v$ is a nonzero vector, then (A.2) can be written as
\[
Bv = xx^Tv. \tag{A.3}
\]
By multiplying $v^T$ to the above equation from left, one can get
\[
v^Tx = x^Tv = \pm \sqrt{(v^TBv)}. \tag{A.4}
\]
Furthermore, from (A.3) one gets
\[
Bv = xx^Tv = (x^Tv)x = \pm \sqrt{(v^TBv)x}. \tag{A.5}
\]
The above equation gives two solutions to (A.3). However, there is only one solution to (A.3) which is given by
\[
A = xx^T = \frac{1}{v^TBv}Bv(Bv)^T. \tag{A.6}
\]
The above equation proves Lemma 1.

References


**Liyu Cao** received his bachelor and master degree from Tianjin University, Tianjin, China in 1982 and 1985, respectively, and Ph.D. degree from Tsinghua University, Beijing, China in 1989, all in electrical engineering. From October 1989 to October 1992, he was a lecturer of the Department of Electrical Engineering, Tsinghua University. From April 1993 to June 1995, he was with the Department of Electrical Engineering, the University of Tokyo, Tokyo, Japan as a visiting researcher. From July 1995 to March 1997 he worked at the Education and Research Center of Software Engineering, Tokyo Institute of Polytechnics, Astugi, Japan. From April 1997 to August 1997 he was an engineer of Toyo Electric Mfg. Co. Ltd., Abina, Japan. Since November 1997 he has been a research associate with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada. His research interests include control theory and applications, parameter estimation and nonlinear systems.

**Howard M. Schwartz** received the B.Eng. degree in Civil Engineering from McGill University, Montreal, Canada in 1981, the M.Sc. degree in Aeronautics and Astronautics and the Ph.D. degree in Mechanical Engineering from the Massachusetts Institute of Technology (MIT), Cambridge in 1982 and 1987, respectively. Since 1987 he has been a Professor in the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada. His research interests include system identification, estimation and tracking, adaptive and nonlinear control, robot control, vision systems and image processing.