KYBERNETIKA — VOLUME 33 (1997), NUMBER 4, PAGES 387-398

THE DAMPED MODIFIED ITERATED KALMAN FILTER FOR NONLINEAR DISCRETE TIME SYSTEMS[1]

MYOUNGHO OH AND U JIN CHOI

The modified iterated Kalman filter, which will be called MIKF for brevity, is derived from the modified Newton method to approximate a maximum likelihood estimate. The MIKF is also obtained by an iteration scheme for the extended Kalman filter equations. A convergence analysis of the MIKF is given. By the damping method, we can reduce the total CPU time needed to estimate the state variables or may even obtain a convergent scheme when the MIKF diverges. A numerical example shows the effective convergence behavior of the damped MIKF.

1. INTRODUCTION

The extended Kalman filter (EKF) is a well-known method for estimation of unknown state variables in nonlinear discrete time systems. Although the EKF gives efficient estimation properties, it converges slowly or even diverges in particular problem. To achieve better convergence performance, a higher-order estimation technique is used despite of computational complexity. The comparative convergence performance of the EKF and the higher-order methods, the iterated and second-order Kalman filters, can be found in [1] and [4].

The damped modified iterated Kalman filter is introduced to make a compromise the accuracy and complexity between the extended and iterated Kalman filters. The derivation of the MIKF is based on the modified Newton method for approximating a maximum likelihood estimate, and it's convergence analysis is given. The modified Newton method was first considered by Kantorovich and the error bounds for this scheme were provided by Smooke [9]. The MIKF is also obtained by an iteration scheme for the EKF formulae. When a single iteration is performed, the modified iterated Kalman filter reduces to the EKF.

By the damping method for convergence acceleration, we can reduce the total CPU time needed to estimate the state variables or may even obtain a convergent scheme when the MIKF diverges. The damped MIKF shows better convergence behavior than the EKF and requires less calculations than the iterated Kalman filter

[1]This studies were supported in part by the Wharangdae Research Institute Program, Korea Military Academy.

(IKF) to attain a given error tolerance. The decision as to which of the three types of methods should be used depends on several factors, including convergence and computational efficiency. The bistatic ranging example is included in this point to illustrate the difference between the EKF, the IKF and the damped MIKF algorithm.

2. PROBLEM DESCRIPTION

Consider the following nonlinear stochastic system described by the state-space model:

$$\begin{aligned} x_{k+1} &= f_k(x_k) + w_k, \quad k = 0, 1, \ldots & (1) \\ y_k &= h_k(x_k) + v_k, \quad k = 0, 1, \ldots & (2) \end{aligned}$$

where k denotes time, x_k is the $n \times 1$ state vector, y_k is the $m \times 1$ observation vector, f_k is the nonlinear vector valued function, h_k is the nonlinear measurement vector, and w_k and v_k are the mutually independent zero-mean guassian white noise vectors with variance matrices

$$\begin{aligned} Q_k &= E[w_k w_k^T] & (3) \\ R_k &= E[v_k v_k^T] & (4) \end{aligned}$$

in which $E[\cdot]$ represents the expectation and superscript T denotes the transpose of a matrix or vector. Let $\hat{x}_k$ be the $n \times 1$ estimate vector.

Let h_k be third-order differentiable. It is assumed that $\hat{x}_k, y_k, P_k$ and R_k are known where $\hat{x}_k$ and y_k are independent gaussian vectors with [1]

$$\begin{aligned} \hat{x}_k &\sim N(x_k, P_k) & (5) \\ y_k &\sim N(h_k(x_k), R_k). & (6) \end{aligned}$$

The purpose of this paper is to obtain an improved estimate and corresponding estimation error covariance of the state vector by using the damped MIKF.

3. THE MODIFIED NEWTON METHOD

The problems of finding the simultaneous solution of n nonlinear equations and optimization problems for nonlinear multivariate functions are very closely related to each other. Specifically, the problem of finding the minimizing point of a nonlinear function $F(x)$ of n real variables is equivalent to the problem of solving the system of n nonlinear equations in n unknown variables

$$\nabla F(x) = 0 \tag{7}$$

where $\nabla F(x)$ denotes the n component gradient column vector of first partial derivatives of the function $F(x)$ [10].

The modified Newton method seeks the solution of (7) using the iterate procedure

$$x^{i+1} = x^i - J^{-1}(x^0)\nabla F(x^i), \quad i = 0, 1, 2, \ldots \tag{8}$$

where x^{i+1} is the $(i+1)$th solution iterate, $J(x^0)$ is the $n \times n$ nonsingular matrix of the second partial derivatives of F evaluated at x^0 and the initial estimate x^0 is given. Let $\tilde{J}$ be the numerical approximation to the analytic Hessian matrix J. Then the modified Newton method can be rewritten as

$$x^{i+1} = x^i - \tilde{J}^{-1}(x^0)\nabla F(x^i), \quad i = 0, 1, 2, \dots. \tag{9}$$

For a convergence analysis of this method we refer to [5]. Let η be an $n \times 1$ free variable vector. Let $F(\eta)$ be a function of η and observations $Y(1), Y(2), \dots, Y(N)$, which will be derived from a likelihood function. The true value of η and the solution of the equation $\nabla F(\eta) = 0$ will be denoted by x and $\hat{x}^+$, respectively.

Theorem 3.1. Let $F(\eta)$ be third-order differentiable with respect to η in the open ball $U = U(x, \mu)$ and assume that $N^{-1}\partial^3 F(\eta)/\partial\eta_p\partial\eta_q\partial\eta_r$, p, q, $r = 1, 2, \dots, n$, is bounded in probability for all $\eta \in U$. Assume that $\tilde{J}(x^0)$ is nonsingular and the (p, q) component of $\tilde{J}(x^0)$ converges in probability to a constant C_{pq}. In addition, assume that x^0 and $\hat{x}^+$ are $\sqrt{N}$-consistent estimate of x, that is, $\sqrt{N}(x^0 - x)$ and $\sqrt{N}(\hat{x}^+ - x)$ are bounded in probability. Then $N^{\frac{1}{2}(i+1)}(x^i - \hat{x}^+)$, $i = 1, 2, \dots$, converges to 0 in probability.

Proof. Expand $\nabla F(x^i)$ about $\hat{x}^+$ as

$$\nabla F(x^i) = \nabla F(\hat{x}^+) + J(\hat{x}^+)(x^i - \hat{x}^+) + \frac{1}{2}\left\{\sum_{r=1}^{n}(x_r^i - \hat{x}_r^+)\partial J(\eta^*)/\partial\eta_r\right\}(x^i - \hat{x}^+) \tag{10}$$

where η^* lies between $\hat{x}_r^+$ and x_r^i. We also expand $J(\hat{x}^+)$ about x and find

$$J(\hat{x}^+) = J(x) + \sum_{r=1}^{n}(\hat{x}_r^+ - x_r)\partial J(\eta^{**})/\partial\eta_r \tag{11}$$

where η^{**} lies between x_r and $\hat{x}_r^+$.

Substituting (11) into (10) and the result into (9), we obtain

$$\begin{aligned}
x^{i+1} - \hat{x}^+ &= x^i - \hat{x}^+ - \tilde{J}^{-1}(x^0)\nabla F(x^i) \\
&= \left[I - \tilde{J}^{-1}(x^0)J(x)\right](x^i - \hat{x}^+) \\
&\quad -\tilde{J}^{-1}(x^0)\left\{\sum_{r=1}^{n}(\hat{x}_r^+ - x_r)\,\partial J(\eta^{**})/\partial\eta_r\right\}(x^i - \hat{x}^+) \\
&\quad -\frac{1}{2}\tilde{J}^{-1}(x^0)\left\{\sum_{r=1}^{n}(x_r^i - \hat{x}_r^+)\,\partial J(\eta^*)/\partial\eta_r\right\}(x^i - \hat{x}^+) \\
&= \tilde{J}^{-1}(x^0)\left\{\sum_{r=1}^{n}(x_r^0 - x_r)\,\partial J(\tilde{\eta})/\partial\eta_r\right\}(x^i - \hat{x}^+) \\
&\quad -\tilde{J}^{-1}(x^0)\left\{\sum_{r=1}^{n}(\hat{x}_r^+ - x_r)\,\partial J(\eta^{**})/\partial\eta_r\right\}(x^i - \hat{x}^+)
\end{aligned}$$

$$-\frac{1}{2}\tilde{J}^{-1}(x^0)\left\{\sum_{r=1}^{n}(x_r^i - \hat{x}_r^+)\ \partial J(\eta^*)/\partial\eta_r\right\}(x^i - \hat{x}^+) \quad (12)$$

where $\tilde{\eta}$ lies between x_r and x_r^0. To prove convergence, we will use mathematical induction. We assume that $N^{\frac{i}{2}}(x^{i-1}-\hat{x}^+)$, $i = 1, 2, \ldots$, converges to 0 in probability as $N \to \infty$. Letting $i = i - 1$ in (12) and multiplying both sides by $N^{\frac{1}{2}(i+1)}$, $i = 1, 2, \ldots$, we obtain

$$\begin{aligned} N^{\frac{1}{2}(i+1)}(x^i - \hat{x}^+) &= \tilde{J}^{-1}\left\{\sum_{r=1}^{n} N^{\frac{1}{2}}(x_r^0 - x_r)\ \partial J(\tilde{\eta})/\partial\eta_r\right\} N^{\frac{i}{2}}(x^{i-1} - \hat{x}^+) \\ &-\tilde{J}^{-1}\left\{\sum_{r=1}^{n} N^{\frac{1}{2}}(\hat{x}_r^+ - x_r)\ \partial J(\eta^{**})/\partial\eta_r\right\} N^{\frac{i}{2}}(x^{i-1} - \hat{x}^+) \\ &-\frac{1}{2}\tilde{J}^{-1}\left\{\sum_{r=1}^{n} N^{\frac{1}{2}}(x_r^{i-1} - \hat{x}_r^+)\ \partial J(\eta^*)/\partial\eta_r\right\} N^{\frac{i}{2}}(x^{i-1} - \hat{x}^+). \end{aligned} \quad (13)$$

Since $\tilde{J}^{-1}(x^0)\,\partial J(\tilde{\eta})/\partial\eta_r$, $\tilde{J}^{-1}(x^0)\,\partial J(\eta^{**})/\partial\eta_r$, and $\tilde{J}^{-1}(x^0)\,\partial J(\eta^*)/\partial\eta_r$ are asymptotically bounded and $N^{\frac{i}{2}}(x^{i-1} - \hat{x}^+)$ is bounded in probability, the whole terms on the right of (13) converge to 0 in probability. □

4. DERIVATION OF THE MIKF

For any $a \in R^n$, we define the Euclidean norm of a to be the real number $\|a\|_2 = \sqrt{a^T a}$. Let a nonlinear function $F(\eta)$ be given by

$$F(\eta) = \|u(\eta)\|_2^2 \quad (14)$$

where $u : R^n \to R^n$ has continuous third partial derivatives with respect to all of its variables. Then we find that

$$\nabla F = 2(u')^T u\,. \quad (15)$$

Letting $u_j''(\eta)$ denotes the $n \times n$ Hessian matrix whose entries are the second partial derivatives of the jth variable of u, we obtain

$$(\nabla F)'(\eta) = 2u'(\eta)^T u'(\eta) + 2\sum_{j=1}^{n} u_j(\eta)\, u_j''(\eta)\,. \quad (16)$$

By dropping the second term, we obtain from (9)

$$x^{i+1} = x^i - (u'(x^0)^T u'(x^0))^{-1} u'(x^i)^T u(x^i)\,. \quad (17)$$

The term

$$2\sum_{j=1}^{n} u_j(x^0)\, u_j''(x^0)$$

is called the truncation error or perturbation matrix of the analytic Hessian at x^0.

Adding $\hat{x}_k$ to y_k, we can define an augmented $(m+n)$ dimensional observation vector

$$Y_k = \begin{bmatrix} y_k \\ \hat{x}_k \end{bmatrix}. \tag{18}$$

Then it can easily be shown that

$$Y_k \sim N(d_k(x_k), B_k) \tag{19}$$

where

$$d_k(x_k) = \begin{bmatrix} h_k(x_k) \\ x_k \end{bmatrix}, \quad B_k = \begin{bmatrix} R_k & 0 \\ 0 & P_k \end{bmatrix}. \tag{20}$$

Suppose that P_k, R_k, and $H_k^T R_k^{-1} H_k + P_k^{-1}$ are all invertible. The likelihood function is

$$L(\eta_k) = C \exp\left\{-\frac{1}{2}(Y_k - d_k(\eta_k))^T B_k^{-1}(Y_k - d_k(\eta_k))\right\} \tag{21}$$

where η_k is a free variable replacing x_k and $C = (2\pi)^{-(m+n)/2}|B_k|^{-\frac{1}{2}}$ is a constant independent of η_k. The logarithm of this likelihood function is

$$\ln L(\eta_k) = -\frac{1}{2}(Y_k - d_k(\eta_k))^T B_k^{-1}(Y_k - d_k(\eta_k)) + \ln C\,. \tag{22}$$

The method of maximum likelihood is one of choosing an estimate $\hat{x}_k^+$ for x_k which maximizes $L(\eta_k)$. Notice that maximizing $L(\eta_k)$ is equivalent to minimizing the quadratic function

$$\min(Y_k - d_k(\eta_k))^T B_k^{-1}(Y_k - d_k(\eta_k))\,. \tag{23}$$

Thus a stochastic optimization problem described by maximum likelihood is diverted into a deterministic minimization problem represented by (23) [8].

We shall now derive the modified iterated Kalman filter by applying the modified Newton method (17) to the problem of finding the maximum likelihood estimate of x_k. Comparing (23) with (14), we can define

$$\nabla F(\eta_k) = (Y_k - d_k(\eta_k))^T B_k^{-1}(Y_k - d_k(\eta_k))\,. \tag{24}$$

Consequently, $u : R^n \to R^{m+n}$ and its first derivative are obtained as

$$u(\eta_k) = B_k^{-\frac{1}{2}}(Y_k - d_k(\eta_k)) \tag{25}$$

$$u'(\eta_k) = -B_k^{-\frac{1}{2}} d_k'(\eta_k)\,. \tag{26}$$

Substituting (25) and (26) into (17), we obtain

$$x^{i+1} = x^i + (d_k'(x_k^0)^T (B_k^{-\frac{1}{2}})^T B_k^{-\frac{1}{2}} d_k'(x_k^0))^{-1} d_k'(x_k^i)^T (B_k^{-\frac{1}{2}})^T B_k^{-\frac{1}{2}} (Y_k - d_k(x_k^i)) \tag{27}$$

where $x_k^0 = \hat{x}_k$. Further simplification is obtained by denoting

$$H_{k,i} = h_k'(x_k^i), \quad H_k = h_k'(x_k^0)\,. \tag{28}$$

We then have

$$d_k'(x_k^i) = \begin{bmatrix} h_k'(x_k^i) \\ I \end{bmatrix} = \begin{bmatrix} H_{k,i} \\ I \end{bmatrix}. \tag{29}$$

For an $m \times n$ matrix of the form

$$A = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

in which $\sigma_i > 0$, we define its pseudoinverse to be the $n \times m$ matrix

$$A^* = \begin{bmatrix} \sigma_1^{-1} & & & & & & \\ & \sigma_2^{-1} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r^{-1} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}.$$

In all of these matrices, elements not displayed are zeros. The pseudoinverse of a matrix is uniquely determined although the singular-value decomposition is not unique [6]. An $m \times n$ matrix A is said to have full rank if $\operatorname{rank}(A) = \min(m, n)$. It is assumed that $H_{k,i}$ is of full rank. Hence from (18), (20) and (29) we have

$$\begin{aligned} x_k^{i+1} &= x_k^i + (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} (H_{k,i}^T R_k^{-1}(y_k - h_k(x_k^i)) + P_k^{-1}(\hat{x}_k - x_k^i)) \\ &= (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} (H_{k,i}^T R_k^{-1}(y_k - h_k(x_k^i)) + H_k^T R_k^{-1} H_k x_k^i + P_k^{-1}\hat{x}_k) \\ &= \hat{x}_k + (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} H_{k,i}^T R_k^{-1} \\ &\quad \cdot (y_k - h_k(x_k^i) - R_k (H_{k,i}^T)^* H_k^T R_k^{-1} H_k (\hat{x}_k - x_k^i)). \end{aligned} \tag{30}$$

Let us derive the error covariance associated with the updated estimate $\hat{x}_k^+$. Differentiating equation (23) with respect to η_k and equating to zero, we obtain

$$d_k'(\eta_k)^T B_k^{-1}(Y_k - d_k(\eta_k)) = 0 \tag{31}$$

which is the well-known maximum likelihood equation for x_k. The following assumption will be needed to get the error covariance of $\hat{x}_k^+$. It is assumed that $\hat{x}_k^+$ is sufficiently close to x_k, so that $h_k(x_k)$ can be expanded in a power series about $\hat{x}_k^+$ retaining only first-order terms. The measurement function $h_k(x_k)$ may then be written as

$$h_k(x_k) = h_k(\hat{x}_k^+) + H_k(x_k - \hat{x}_k^+) \tag{32}$$

where $H_k = h'_k(\hat{x}_k^+)$ is a constant.

Since $\hat{x}_k^+$ is simply the solution of (31), we have

$$d'_k(\hat{x}_k^+)^T B_k^{-1}(Y_k - d_k(\hat{x}_k^+)) = H_k^T R_k^{-1}(H_k(x_k - \hat{x}_k^+) + v_k) + P_k^{-1}(\hat{x}_k - \hat{x}_k^+) = 0\,.$$

It is easily seen that

$$\hat{x}_k^+ - x_k = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1}(H_k^T R_k^{-1} v_k + P_k^{-1}(\hat{x}_k - x_k))\,. \tag{33}$$

Hence the error covariance of $\hat{x}^+$ turns out to be

$$\begin{aligned} P_k^+ &\equiv \operatorname{Cov}(\hat{x}_k^+ - x_k) \\ &= E[(\hat{x}_k^+ - x_k)(\hat{x}_k^+ - x_k)^T] \\ &= (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1}(H_k^T R_k^{-1} E[v_k v_k^T] R_k^{-1} H_k \\ &\qquad + P_k^{-1} E[(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T] P_k^{-1})(H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} \\ &= (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} \end{aligned} \tag{34}$$

where $P_k \equiv \operatorname{Cov}(\hat{x}_k - x_k)$. We next derive an approximate error covariance of x_k^{i+1}. It is also assumed that an initial approximation x_k^0 is sufficiently close to x_k and $\hat{x}_k^+$. Then it can be proven that x_k^{i+1} converges to x_k and $\hat{x}_k^+$. Hence the values of h'_k do not change much in a neighborhood of x_k^0 and the effect of replacing $H_{k,i}$ by H_k is small. By repeating the above process to get P_k^+, we obtain

$$P_k^{i+1} = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1}\,.$$

The modified iterated Kalman filter formulae are summarized as follows:

$$x_k^{i+1} = \hat{x}_k + K_{k,i}(y_k - h_k(x_k^i) - R_k(H_{k,i}^T)^* H_k^T R_k^{-1} H_k(\hat{x}_k - x_k^i)) \tag{35}$$

$$P_k^{i+1} = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} \tag{36}$$

where

$$H_{k,i} = h'_k(x_k^i), \quad H_k = h'_k(x_k^0), \quad K_{k,i} = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} H_{k,i}^T R_k^{-1}\,.$$

The iteration process can be stopped when further improvement from additional iterations is small enough. The maximum number of iteration steps that the user will permit is necessary to avoid the possibility of the computation going into an infinite loop.

Remark. A convergence analysis of the MIKF is entirely analogous to that of Theorem 3.1.

5. THE EKF AND MIKF

This section establishes the relation between the modified iterated Kalman filter and the extended Kalman filter. From (35) and (36) with $i = 0$, we obtain

$$\hat{x}_k^+ = \hat{x}_k + K_k(y_k - h_k(\hat{x}_k)) \tag{37}$$

$$P_k^+ = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} \tag{38}$$

where

$$\hat{x}_k = x_k^0, \quad H_k = h_k'(\hat{x}_k), \quad K_k = (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} H_k^T R_k^{-1}.$$

By the matrix inversion lemma, the expression for K_k and P_k^+ can be rewritten as

$$\begin{aligned} K_k &= P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} && (39)\\ P_k^+ &= (I - K_k H_k) P_k. && (40) \end{aligned}$$

Thus the MIKF reduces to the EKF when a single iteration is performed.

We now obtain the MIKF through an iteration technique to improve the updated state estimate (37). The update equations producing the complete extended Kalman filtering are given by

$$\begin{aligned} \hat{x}_k^+ &= \hat{x}_k + K_k(y_k - \hat{h}_k(x_k)) && (41)\\ K_k &= -E[\tilde{x}_k(h_k(x_k) - \hat{h}_k(x_k))^T] \\ &\qquad \cdot \left\{E[(h_k(x_k) - \hat{h}_k(x_k))_k(x_k) - \hat{h}_k(x_k))^T] + R_k\right\}^{-1} && (42)\\ P_k^+ &= P_k + K_k E[(h_k(x_k) - \hat{h}_k(x_k))\tilde{x}_k^T] && (43) \end{aligned}$$

where the caret $(\hat{\ })$ denotes the expectation operation and $\tilde{x}_k = \hat{x}_k - x_k$ [4]. Taking the Taylor expansion of $h_k(x_k)$ around x_k^i produces the following

$$h_k(x_k) = h_k(x_k^i) + H_{k,i}(x_k - x_k^i) + \ldots$$

where $H_{k,i} = h_k'(x_k^i)$. Assuming that $\hat{x}_k$ is close enough to x_k, so that x_k^i converges to x_k and the effect of approximating $H_{k,i}$ by H_k is very small, we have

$$h_k(x_k) \cong h_k(x_k^i) + H_k(x_k - x_k^i). \qquad (44)$$

Multiplying the left side of H_k in (44) by $R_k(H_{k,i}^T)^* H_k^T R_k^{-1}$, we obtain

$$h_k(x_k) \cong h_k(x_k^i) + R_k(H_{k,i}^T)^* H_k^T R_k^{-1} H_k(x_k - x_k^i). \qquad (45)$$

Substituting (45) into (41), we obtain the iterative expression for the updated state estimate as follows:

$$\begin{aligned} x_k^{i+1} &= \hat{x}_k + K_{k,i}(y_k - \hat{h}_k(x_k)) \\ &= \hat{x}_k + K_{k,i}\left\{y_k - h_k(x_k^i) - R_k(H_{k,i}^T)^* H_k^T R_k^{-1} H_k(\hat{x}_k - x_k^i)\right\}. && (46) \end{aligned}$$

Substituting (45) into (42), and using the matrix inversion lemma, we also have

$$\begin{aligned} K_{k,i} &= P_k H_k^T R_k^{-1} H_k H_{k,i}^* R_k (R_k (H_{k,i}^T)^* H_k^T R_k^{-1} H_k P_k H_k^T R_k^{-1} H_k H_{k,i}^* R_k + R_k)^{-1} \\ &= (H_k^T R_k^{-1} H_k H_{k,i}^* R_k R_k^{-1} R_k (H_{k,i}^T)^* H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} H_k^T R_k^{-1} H_k H_{k,i}^* R_k R_k^{-1} \\ &= \left(H_k^T R_k^{-1} H_k + \left(H_k^T R_k^{-1} H_k H_{k,i}^* R_k (H_{k,i}^T)^*\right)^{-1} P_k^{-1}\right)^{-1} \\ &\qquad \cdot \left(H_k^T R_k^{-1} H_k H_{k,i}^* R_k (H_{k,i}^T)^*\right)^{-1} H_k^T R_k^{-1} H_k H_{k,i}^* \\ &= \left(H_k^T R_k^{-1} H_k + (H_k^T R_k^{-1} H_k H_{k,i}^* R_k (H_{k,i}^T)^*)^{-1} P_k^{-1}\right) H_{k,i}^T R_k^{-1}. && (47) \end{aligned}$$

Note that (47) differs from the gain matrix of the MIKF. In order to get asymptotic result, it is necessary to approximate $H_k^T R_k^{-1} H_k H_{k,i}^* R_k (H_{k,i}^T)^*$ by identity matrix under hyperthesis x_k^{i+1} is sufficiently close to x_k°. Thus we only obtain the following approximate gain matrix

$$K_{k,i} \cong (H_k^T R_k^{-1} H_k + P_k^{-1}) H_{k,i}^T R_k^{-1} . \tag{48}$$

Then we have

$$\begin{aligned} P_k^{i+1} &= P_k + K_{k,i} E\left[(h_k(x_k) - \hat{h}_k(x_k))\tilde{x}_k^T\right] \\ &= P_k - K_{k,i} R_k (H_{k,i}^T)^* H_k^T R_k^{-1} H_k P_k \\ &= P_k - (H_k^T R_k^{-1} H_k + P_k^{-1})^{-1} H_k^T R_k^{-1} H_k P_k \\ &= \left(H_k^T R_k^{-1} H_k + P_k^{-1}\right)^{-1} . \end{aligned} \tag{49}$$

6. DAMPING METHOD FOR THE MIKF

Although the MIKF shows efficient convergence behavior, it requires many calculations for treating system nonlinearities. To ensure convergence, we need to have a sufficiently close initial estimate to the state variable and the condition for being sufficiently close is not easy to check. To overcome these disadvantages, we apply the damping method to the MIKF.

Let z be an $n \times 1$ column vector, $A = (a_{ij})$ be an $n \times n$ matrix and define the maximum norms of z and A by

$$\begin{aligned} \|z\|_\infty &= \max_{1 \le i \le n} |z_i| \\ \|A\|_\infty &= \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|. \end{aligned}$$

Smooke [9] presented an error estimate which bounds the size of the sequence of modified Newton iterates by the terms of a recurrence sequence scaled by the initial Newton step. From an error estimate for the modified Newton method, we take

$$\|x_k^{i+1} - x_k^i\|_\infty \le w(c) \|x_k^i - x_k^{i-1}\|_\infty, \quad i = 1, 2, \dots \tag{50}$$

where $w(c)$ is a function of c and c is the constant satisfying the Kantorovich hypotheses. In damping, we use the relationship (50) as a repetition criteria to accelerate the convergence of the MIKF. If, after $(i+1)$th iteration procedure, we find that $\|x_k^{i+1} - x_k^i\|_\infty$ does not satisfy the inequality (50), then we choose x_k^i as a new initial estimate and restart the MIKF algorithm.

The damping parameter $w(c)$ is to be chosen to make x_k^i converge to x_k as rapidly as possible. The exact bounds of $w(c)$ for the MIKF was not found. However, carrying out experiment with several values of $w(c)$ and comparing the effect on the convergence speed, we can choose an optimal value of $w(c)$. As a result, we can encourage the speed of convergence or may even avoid the divergence caused by a poor initial estimate.

7. A NUMERICAL EXAMPLE

To compare the damped modified iterated Kalman filter with the extended and iterated Kalman filters, we consider a two-dimensional bistatic ranging problem when noisy measurements are taken. This example was studied algebraically by Bell [1]. The iterated Kalman filter formulae are given as follows:

$$x_k^{i+1} = \hat{x}_k + K_{k,i}(y_k - h_k(x_k^i) - H_{k,i}(\hat{x}_k - x_k^i)) \tag{51}$$

$$P_k^{i+1} = (I - K_{k,i}H_{k,i})P_k \tag{52}$$

where

$$H_{k,i} = h_k'(x_k^i), \quad K_{k,i} = (H_{k,i}^T R_k^{-1} H_{k,i} + P_k^{-1})^{-1} H_{k,i}^T R_k^{-1}.$$

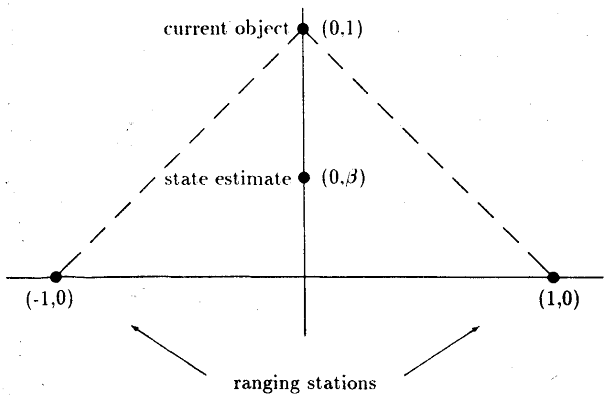

Fig. 1.

The object being tracked is currently located at x and two ranging stations are sited at $(-1,0)$ and $(+1,0)$. The current object state, state estimate, and the current measurement are taken as

$$x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} 0 \\ \beta \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{53}$$

where $\beta \sim N(2.0, 0.1^2)$ or $\beta \sim N(0.5, 0.1^2)$. The covariance matrices of $\hat{x}$ and y are given by

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \tag{54}$$

where $\hat{x}$ and y are independent and $\rho \sim N(0.01, 0.001^2)$. The measurement function $h : R^2 \to R^2$ and its derivative are taken as

$$h(\xi) = \frac{1}{2}\begin{bmatrix} (\xi_1+1)^2 + \xi_2^2 \\ (\xi_1-1)^2 + \xi_2^2 \end{bmatrix}, \quad h'(\xi) = \begin{bmatrix} \xi_1+1 & \xi_2 \\ \xi_1-1 & \xi_2 \end{bmatrix} \tag{55}$$

where $\xi = [\xi_1 \quad \xi_2]^T$. We chose $w(c) = 0.25$ and Monte Carlo simulation experiments of 100 runs were performed. The computations were performed on a SUN C computer.

The comparative convergence performances of several types of filtering algorithms are shown in Figure 2 and Figure 3 with $\beta \sim N(2.0,\, 0.1^2)$ and $\beta \sim N(0.5,\, 0.1^2)$, respectively.

This example shows that the convergence speed of the damped MIKF is slower than that of the IKF in the first few steps, but it becomes quite fast as the number of iterates increases by the damping method. Figure 3 also demonstrates that the damped MIKF converges while the MIKF diverges because of an inappropriate initial estimate. We obtain similar results for different values of $w(c)$, β and ρ.

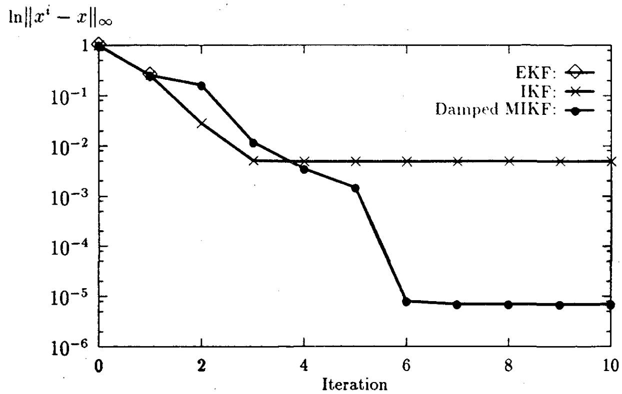

Fig. 2.

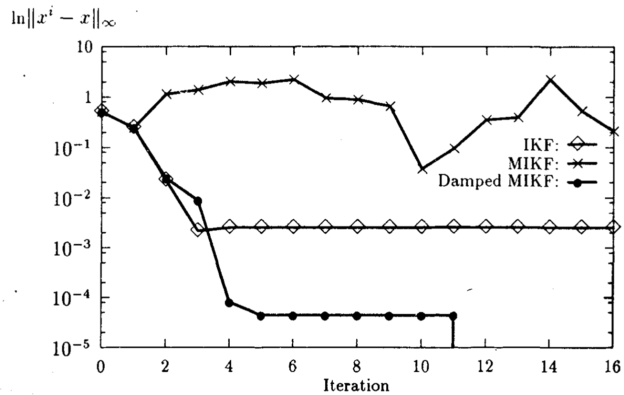

Fig. 3.

8. CONCLUDING REMARKS

We have derived the modified iterated Kalman filter from the modified Newton method to approximate a maximum likelihood estimate, and showed a convergence analysis of the MIKF. Applying the damping method to the MIKF, we can reduce the total CPU time needed to solve a given estimation problem or may even obtain a convergent scheme when it diverges.

A numerical example illustrates the difference between estimation accuracy and computational complexity, which demonstrates the effectiveness of the damped MIKF. Depending on several factors, which are required to the particular problem, such as convergence assurance and computational efficiency, one can choose an appropriate method.

A prediction error algorithm has also been derived by using the principle of maximum likelihood for the estimation problem and its asymptotic behavior is given in [7]. It remains further study to compare the damped MIKF with Ljung's prediction error method.

(Received May 18, 1995.)

REFERENCES

[1] M. B. Bell and F. W. Cathey: The iterated Kalman filter update as a Gauss–Newton method. IEEE Trans. Automat. Control *38* (1993), 2, 294–297.
[2] D. E. Catlin: Estimation, Control, and the Discrete Kalman Filter. Springer–Verlag, New York 1989.
[3] S. D. Conte and C. de Boor: Elementary Numerical Analysis. Third edition. McGraw–Hill, Singapore 1987.
[4] A. Gelb: Applied Optimal Estimation. MIT Press, Cambridge, Massachusetts 1974.
[5] Y. Hosoya and M. Taniguchi: A central limit theorem for stationary processes and the parameter estimation of linear processes. Ann. Statist. *10* (1982), 1, 132–153.
[6] D. Kincaid and W. Cheney: Numerical Analysis: Mathematics of Scientific Computing. Brooks/Cole Publishing Company, California 1990.
[7] L. Ljung: Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. IEEE Trans. Automat. Control *AC-24* (1979), 1, 36–50.
[8] N. E. Nahi: Estimation Theory and Applications. Wiley, New York 1969.
[9] M. D. Smooke: Error estimate for the modified Newton method with application to the solution of nonlinear, two–point boundary–value problems. J. Optim. Theory Appl. *39* (1983), 4, 489–511.
[10] F. Szidarovszky and S. Yakowitz: Principles and Procedures of Numerical Analysis. Plenum Press, New York 1978.

Dr. Myoungho Oh, Department of Computer Science, Korea Military Academy, P. O. Box 77, Gongneung-dong, Nowon-gu, Seoul, 139-799. Korea.

Dr. U Jin Choi, Department of Mathematics, Korea Advanced Institute of Science and Technology, Kusong-dong, Yusong-gu, Taejon, 305-701. Korea.