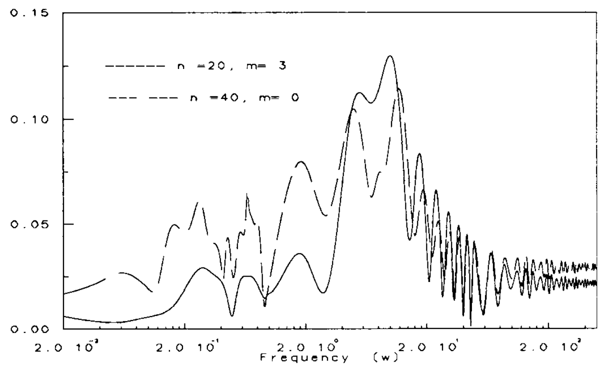

Fig. 1. Plot of $|\hat{h}_{id}^{N}(j\omega) - \hat{h}(j\omega)|$ for the example using the nonlinear algorithm.

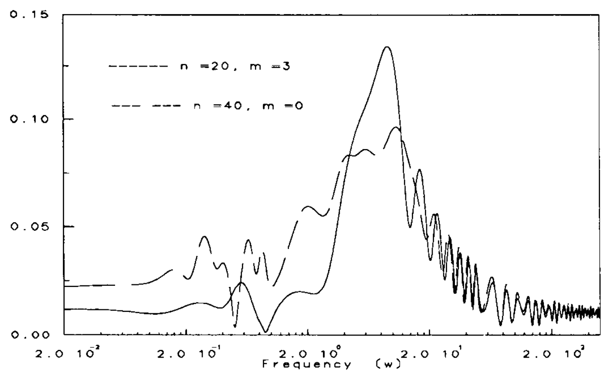

Fig. 2. Plot of $|\hat{h}_{id,L}^{N}(j\omega) - \hat{h}(j\omega)|$ for the example using the linear algorithm.

where $\tilde{h}_k^N$ is computed from inverse DFT as in (2.8). If the infinity norm of the anticausal part of the preidentified model is small, the use of the linear algorithm is justifiable.

IV. Illustrative Example

In this section, we use an example to illustrate the proposed identification algorithms with a window function in (3.21). The noise is generated according to $\eta_k = e^{j\theta_k}$, with θ_k a random variable uniformly distributed on the interval $[0, 2\pi]$. We chose $n = 20, 40$ and $N = 256$ as the order of the identified model and the number of data points, respectively.

Example: The system transfer function is given by (see also [11])

$$\hat{h}(s) = \frac{e^{-s}}{\sqrt{s^2 + \sqrt{3}s + 1}}.$$

The simulation results for the nonlinear algorithm and the linear algorithm can be found in Figs. 1 and 2, respectively. It should be noted that the use of the linear algorithm leads to fairly small identification errors. The parameter m used is $m = 3, 0$ for $n = 20, 40$, respectively.

V. Conclusions

In this note, we have shown that the class of systems with a lower bound on the relative stability, an upper bound on the steady state gain, and an upper bound on the roll-off rate is admissible. This led us to develop a class of robustly convergent nonlinear algorithms. Explicit error bounds were derived for a particular algorithm. It is concluded that the identification problem is reduced to the optimal design of the window function for the aforementioned admissible class of transfer functions.

References

[1] G. Gu and P. P. Khargonekar, "Linear and nonlinear algorithms for identification in $\mathcal{H}_\infty$ with error bounds," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 953–963, 1992; also in *Proc. 1991 Amer. Contr. Conf.*, Boston, MA, 1991, pp. 64–69.

[2] ——, "A class of algorithms for identification in $\mathcal{H}_\infty$," *Automatica*, vol. 28, pp. 229–242, 1992; also, abridged version in *Proc. 30th IEEE Conf. Dec. Contr.*, to appear.

[3] A. J. Helmicki, C. A. Jacobson, and C. N. Nett, "Identification in $\mathcal{H}_\infty$: A robustly convergent nonlinear algorithm," in *1989 American Contr. Conf. 1989 Int. Math. Theory Symp. Networks Syst.*, *Proc. 1990 Amer. Contr. Conf.*, 1990, pp. 386–391.

[4] ——, Worst-Case/Deterministic Identification in H_∞: The Continuous-Time Case, *IEEE Trans. Automat. Contr.*, vol. 37, no. 5, pp. 604–610, 1992.

[5] ——, "Identification in $\mathcal{H}_\infty$: Linear algorithms," in *Proc. 1990 Amer. Contr. Conf.*, pp. 2418–2423.

[6] ——, "Fundamentals of control-oriented system identification and their application for identification in $\mathcal{H}_\infty$," in *Proc. 1991 Amer. Contr. Conf.*, Boston, MA, 1991, pp. 89–99.

[7] P. M. Makila and J. R. Partington, "Robust approximation and identification in $\mathcal{H}_\infty$," Åbo Akademi, Report 91-2, 1991.

[8] J. R. Partington, "Robust identification in $\mathcal{H}_\infty$," in *J. Math. Anal. Appl.*, vol. 166, pp. 428–441, 1992.

[9] ——, "Robust identification and interpolation in $\mathcal{H}_\infty$," *Int. J. Contr.*, vol. 54, pp. 1281–1290, 1991.

[10] J. R. Partington and P. M. Makila, "Robust identification of stabilizable systems," preprint, presented at Univ. of Leeds, 1991.

[11] N. E. Wu, G. Gu, "Discrete Fourier transform and $\mathcal{H}_\infty$—approximation," *IEEE Trans. Automat. Contr.*, vol. 35, 1044–1046, Sept. 1990.

[12] N. J. Young, *An Introduction to Hilbert Space*. Cambridge, UK: Cambridge University Press, 1988.

[13] A. Zygmund, *Trigonometric Series*. Cambridge, UK: Cambridge University Press, 1959.

The Iterated Kalman Filter Update as a Gauss–Newton Method

Bradley M. Bell and Frederick W. Cathey

***Abstract*—We show that the iterated Kalman filter (IKF) update is an application of the Gauss–Newton method for approximating a maximum likelihood estimate. We also present an example in which the iterated Kalman filter update and maximum likelihood estimate show correct convergence behavior as the observation becomes more accurate, whereas the extended Kalman filter update does not.**

Manuscript received May 31, 1991; revised January 3, 1992.

B. M. Bell is with the Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, Seattle, WA 98105.

F. W. Cathey is with Avionics Technology, Boeing Defense and Space Group, Seattle, WA 98124.

IEEE Log Number 9203097.

I. Introduction

The basic filtering solution to the state estimation problem for a discrete-time dynamical system can be described as a two-stage recursive process of prediction and update. In particular, this is true of the Kalman filter and its variations—the *extended Kalman filter* (EKF) and the *iterated Kalman filter* (IKF). The two variations handle the case of nonlinear measurement functions. The update method of the IKF reduces to that of the EKF in the case of a single iterate, and both methods reduce to the ordinary Kalman update when the measurement function is affine. At the expense of more computation, the IKF generally performs better than the EKF. For a discussion of these filters, see, e.g., [3], [4], and [6].

We formulate the maximum likelihood/least squares approach to the (nonlinear) update problem and show that the update method of the IKF is an application of the Gauss–Newton method to approximate a solution. This unifies the treatment of the IKF in [3] with the treatment of the modified Newton methods found in [1]. We also present an example of a *bistatic* ranging problem in which the IKF and maximum likelihood solutions are given in algebraic closed form. In the example, we have correct convergence of the maximum likelihood and IKF solutions as the observation becomes more accurate. On the other hand, the EKF update converges to a biased value, while its error covariance converges to zero. This convergence/false convergence result is true in general for nonlinear measurements with "enough dimensionality," i.e., when an exact measurement completely determines state.

II. The Update Problem

The update problem for a dynamical system is equivalent to the static estimation problem: make a correction to the current (predicted) state estimate based on the current observation. The dynamics of the system do not figure into this problem.

We adopt the following notation and conventions. The state and measurement spaces are real vector spaces of dimensions n and m, respectively. Let $x \in \mathbb{R}^n$, $\hat{x} \in \mathbb{R}^n$, and $z \in \mathbb{R}^m$ denote the current state, the current state estimate, and the observation, respectively. The measurement function $h:\mathbb{R}^n \to \mathbb{R}^m$, where $h(x)$ models a noiseless measurement, is twice differentiable and its first derivative is denoted by h'. Throughout this note, $\xi \in \mathbb{R}^n$ is a free variable vector.

All vectors will be regarded as column vectors, i.e., matrices of column dimension 1. The transpose of a matrix M is denoted by M^*.

We regard z and $\hat{x}$ as realizations of independent random vectors with multivariate normal distributions

$$z \sim N(h(x), R), \qquad \hat{x} \sim N(x, P). \tag{1}$$

Values for z, $\hat{x}$, R, and P are given. The update problem is to find a better state estimate $\hat{x}^+$ and corresponding covariance P^+ using the given information.

III. The IKF and EKF Update Methods

We briefly recall the update formulas used in the iterated and extended Kalman filters. For any natural number i, the IKF provides the update

$$\hat{x}^+ = x_i, \qquad P^+ = P_i.$$

The sequences $\{x_i\}$ and $\{P_i\}$ are defined inductively as follows

$$x_0 = \hat{x}, \qquad P_0 = P$$

$$x_{i+1} = \hat{x} + K_i(z - h(x_i) - H_i(\hat{x} - x_i)) \tag{2}$$

$$P_{i+1} = (I - K_i H_i)P \tag{3}$$

where

$$H_i = h'(x_i), \qquad K_i = PH_i^*(H_i P H_i^* + R)^{-1}.$$

For a single iteration, setting $i = 0$ in (2) and (3) above, we obtain the EKF update formulas

$$\hat{x}^+ = \hat{x} + K(z - h(\hat{x})) \tag{4}$$

$$P^+ = (I - KH)P \tag{5}$$

where

$$H = h'(\hat{x}), \qquad K = PH^*(HPH^* + R)^{-1}.$$

IV. The Maximum Likelihood Update

For convenience, we lump the current observation and state estimate into a single "observation" vector. Thus, we form the *augmented* observation and measurement function

$$Z = \begin{bmatrix} z \\ \hat{x} \end{bmatrix}, \qquad g(x) = \begin{bmatrix} h(x) \\ x \end{bmatrix}. \tag{6}$$

It follows from (1) and the independence assumption that

$$Z \sim N(g(x), Q) \qquad \text{where} \quad Q = \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}. \tag{7}$$

The update problem can now be restated as follows: compute a state estimate $\hat{x}^+$ and corresponding covariance P^+ given Z, Q, and g.

Recall that the *likelihood function* $L(\xi)$ is just the probability density of Z with x replaced by a free variable ξ (see, e.g., [5]). Thus

$$L(\xi) = \frac{1}{\sqrt{(2\pi)^{m+n}|Q|}} \cdot \exp\left(-\frac{1}{2}(Z - g(\xi))^* Q^{-1}(Z - g(\xi))\right). \tag{8}$$

The *maximum likelihood estimate* for x and an approximation of its covariance are given by

$$\hat{x}^+ = \operatorname{argmax}(L(\xi)) \tag{9}$$

$$P^+ = (G^* Q^{-1} G)^{-1} \qquad \text{where} \quad G = g'(\hat{x}^+). \tag{10}$$

Since maximizing $L(\xi)$ amounts to minimizing the negative of its logarithm, we have the following equivalent formulation of (9):

$$\hat{x}^+ = \operatorname{argmin}(q(\xi)) \tag{11}$$

where

$$q(\xi) = \frac{1}{2}(Z - g(\xi))^* Q^{-1}(Z - g(\xi)). \tag{12}$$

The differential condition $q'(\xi) = 0$, necessary for an extremum, is called the *maximum likelihood equation* for x. For the case at hand, it is given by

$$0 = g'(\xi)^* Q^{-1}(Z - g(\xi)). \tag{13}$$

The Gauss–Newton technique for approximating $\hat{x}^+$ and its equivalence with the IKF update method are discussed in the next two sections. We conclude here with a derivation of (10) and its relationship to the IKF covariance equation [see (3)].

The approximation of the covariance of $\hat{x}^+ - x$ is based on the assumption that the given realization of $\hat{x}^+$ is close enough to x that we can replace g by its first-order approximant: i.e., we assume g affine on a neighborhood of $\hat{x}^+$ and x. Thus, $g(\xi) = g(\hat{x}^+) + G(\xi - \hat{x}^+)$ where $G = g'(\xi) = g'(\hat{x}^+)$ is constant. Let $V = Z - g(x)$, and note from (7) that $V \sim N(0, Q)$. Substituting $\hat{x}^+$ for ξ in (13) we obtain

$$\begin{aligned} 0 &= g'(\hat{x}^+)^* Q^{-1}(g(x) + V - g(\hat{x}^+)) \\ &= G^* Q^{-1}(G(x - \hat{x}^+) + V) \\ \hat{x}^+ - x &= (G^* Q^{-1} G)^{-1} G^* Q^{-1} V. \end{aligned} \tag{14}$$

Thus

$$\begin{aligned} P^+ &\overset{\text{def}}{=} E((\hat{x}^+ - x)(\hat{x}^+ - x)^*) \\ &= (G^* Q^{-1} G)^{-1} G^* Q^{-1} E(VV^*) Q^{-1} G (G^* Q^{-1} G)^{-1} \\ P^+ &= (G^* Q^{-1} G)^{-1}. \end{aligned} \tag{15}$$

We relate this expression to the usual formula involving the Kalman gain. Since

$$G = g'(\hat{x}^+) = \begin{bmatrix} H \\ I \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} R & 0 \\ 0 & P \end{bmatrix}, \quad \text{where} \quad H = h'(\hat{x}^+), \tag{16}$$

it follows that

$$P^+ = (G^* Q^{-1} G)^{-1} = (H^* R^{-1} H + P^{-1})^{-1}. \tag{17}$$

The matrix inversion lemma (see, e.g., [4])

$$(H^* R^{-1} H + P^{-1})^{-1} H^* R^{-1} = K = PH^*(HPH^* + R)^{-1} \tag{18}$$

holds for any invertible R and P and for any H of consistent dimension, as can be shown by multiplying on the left by $H^* R^{-1} H + P^{-1}$ and on the right by $HPH^* + R$. It follows from (17) and (18) that

$$\begin{aligned} P^+ &= (H^* R^{-1} H + P^{-1})^{-1}((H^* R^{-1} H + P^{-1})P - H^* R^{-1} HP) \\ &= P - (H^* R^{-1} H + P^{-1})^{-1} H^* R^{-1} HP \\ &= (I - KH)P. \end{aligned} \tag{19}$$

This is the usual covariance update formula in terms of the Kalman gain.

V. The Gauss–Newton Method

Using the notation of [2, definition 10.1.1], the *nonlinear least squares problem* has the form

$$\text{minimize} \quad f(\xi) = \tfrac{1}{2}\|r(\xi)\|^2 \tag{20}$$

where $r: \mathbb{R}^m \to \mathbb{R}^m$ is twice differentiable and $\|y\| = \sqrt{y^* y}$ is the usual norm on $y \in \mathbb{R}^m$. The Gauss–Newton method for solving this problem applies an approximate Newton method to the problem of finding a root of the gradient function $\nabla f(\xi) = r'(\xi)^* r(\xi)$. (Here $\nabla f: \mathbb{R}^n \to \mathbb{R}^n$ is given by the transpose of f'.)

Recall that Newton's method defines a sequence of approximations x_i as follows. Given an initial approximation x_0, inductively define

$$x_{i+1} = x_i - ((\nabla f)'(x_i))^{-1} \nabla f(x_i). \tag{21}$$

If f is quadratic, so that ∇f (equivalently r) is affine, then x_1 solves problem (20) in a single step. In general, if $\lim_{i \to \infty} x_i = c$ exists and $(\nabla f)'(c)$ is invertible, then $\nabla f(c) = 0$, showing that c is a critical point of f.

Let H_j denote the Hessian matrix of the jth component of r. Then

$$(\nabla f)'(\xi) = r'(\xi)^* r'(\xi) + D(\xi),$$

$$\text{where} \quad D(\xi) = \sum_{j=1}^{n} r_j(\xi) H_j(\xi).$$

The Gauss–Newton method drops the term D from this formula, yielding the sequence of approximations

$$x_{i+1} = x_i - (r'(x_i)^* r'(x_i))^{-1} r'(x_i)^* r(x_i). \tag{22}$$

Noting that if r is affine then $D = 0$, we observe that the Gauss–Newton method operates by performing successive minimizations

$$x_{i+1} = \operatorname{argmin}(\|\tilde{r}(\xi; x_i)\|^2) \tag{23}$$

where $\tilde{r}(\xi; x_i) = r(x_i) + r'(x_i)(\xi - x_i)$ is the affine approximation to r at x_i.

See [2, theorem 10.2.1] for a statement of rate of convergence of the Gauss–Newton method.

VI. The IKF Update is a Gauss–Newton Method

With the notation of the preceding sections, let S be a "square root" of Q^{-1} (i.e., S satisfies $S^* S = Q^{-1}$), and define $r: \mathbb{R}^n \to \mathbb{R}^{n+m}$ by

$$r(\xi) = S(Z - g(\xi)). \tag{24}$$

Observe that the objective function of the maximum likelihood problem (12) and the objective function of the nonlinear least squares problem (20) are identical.

We evaluate the Gauss–Newton iterates of (22) using the substitution equation (24). Noting that $r'(\xi) = -Sg'(\xi)$, we obtain

$$x_{i+1} = x_i + (g'(x_i)^* S^* S g'(x_i))^{-1} g'(x_i)^* S^* S(Z - g(x_i)) \tag{25}$$

$$= (G_i^* Q^{-1} G_i)^{-1} G_i^* Q^{-1}(Z - g(x_i) + G_i x_i) \tag{26}$$

where $G_i = g'(x_i)$. Using (6), (16), (17), and the matrix inversion lemma (18), we obtain

$$\begin{aligned} x_{i+1} &= (H_i^* R^{-1} H_i + P^{-1})^{-1} \\ &\quad \cdot (H_i^* R^{-1}(z - h(x_i) + H_i x_i) + P^{-1}\hat{x}) && (27) \\ &= \hat{x} + (H_i^* R^{-1} H_i + P^{-1})^{-1} H_i^* R^{-1} \\ &\quad \cdot (z - h(x_i) - H_i(\hat{x} - x_i)) && (28) \\ &= \hat{x} + K_i(z - h(x_i) - H_i(\hat{x} - x_i)). && (29) \end{aligned}$$

By induction, the sequence of iterates generated by the Gauss–Newton method with initial estimate $\hat{x}$ and the sequence of iterates generated by the IKF are identical.

VII. An Illustrative Example

In this section, we present a two-dimensional bistatic ranging example in which the maximum likelihood EKF and IKF solu-

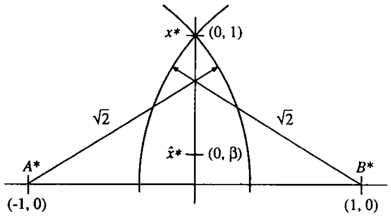

Fig. 1. Bistatic ranging geometry.

tions are given by algebraic expressions. The example demonstrates correct convergence of the maximum likelihood and IKF solutions as the observation becomes more accurate. The EKF update solution is shown to converge to a biased value even though its error covariance converges to zero.

The geometry of the problem is illustrated in Fig. 1. We have ranging stations located at position $A^* = (-1, 0)$ and $B^* = (+1, 0)$, an object being tracked that is currently situated at x, and a current estimate $\hat{x}$ of the object's location. For simplicity, we assume that observation data consist of one-half range squared as measured from each of the tracking stations. This makes the derivative of the measurement function as simple as possible without being constant. Thus, $h:\mathbb{R}^2 \to \mathbb{R}^2$ and its derivative are given by

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

$$\text{where } \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}. \tag{30}$$

The current object state, state estimate, and the current measurement are given by

$$x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \hat{x} = \begin{bmatrix} 0 \\ \beta \end{bmatrix}, \qquad z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{31}$$

We assume that $\hat{x}$ and z are independent and that their covariance matrices are given by

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad R = \rho\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{32}$$

where $\rho > 0$. The observation z represents a chance perfect measurement and is thus consistent with arbitrarily small values of ρ. P is consistent with the current estimate, provided $|\beta - 1|$ is not too large. Note that there are two points that satisfy the measurement equation $z = h(\xi)$, namely x and $-x$. We assume that the estimate $\hat{x}$ is closer to x, i.e., we assume $\beta > 0$.

The IKF update is prescribed by (2) and (3), which we rewrite, using (17) and the matrix inversion lemma (18), as

$$x_{i+1} = \hat{x} + P_{i+1}H_i^*R^{-1}(z - h(x_i) - H_i(\hat{x} - x_i)) \tag{33}$$

$$P_{i+1} = (H_i^*R^{-1}H_i + P^{-1})^{-1}. \tag{34}$$

These expressions simplify to the forms

$$x_{i+1} = \beta_{i+1}x$$

$$P_{i+1} = \rho\begin{bmatrix} (2+\rho)^{-1} & 0 \\ 0 & (2\beta_i^2 + \rho)^{-1} \end{bmatrix}$$

where

$$\beta_{i+1} = \frac{1 + \beta_i^2}{2\beta_i^2 + \rho}\beta_i + \frac{\rho}{2\beta_i^2 + \rho}\beta_0, \qquad \beta_0 = \beta. \tag{35}$$

We show this by induction on the assumption $x_i = \beta_i x$:

$$h(x_i) = \tfrac{1}{2}(1 + \beta_i^2)z \tag{36}$$

$$h'(x_i) = H_i = \begin{bmatrix} 1 & \beta_i \\ -1 & \beta_i \end{bmatrix} \tag{37}$$

$$H_i^*R^{-1}H_i = \frac{2}{\rho}\begin{bmatrix} 1 & 0 \\ 0 & \beta_i^2 \end{bmatrix} \tag{38}$$

$$H_i^*R^{-1}(z - h(x_i) - H_i(\hat{x} - x_i)) = \frac{1}{\rho}\beta_i(1 + \beta_i^2 - 2\beta_i\beta_0)x \tag{39}$$

$$\hat{x} + P_{i+1}H_i^*R^{-1}(z - h(x_i) - H_i(\hat{x} - x_i)) = \beta_0 x + \frac{\beta_i(1 + \beta_i^2 - 2\beta_i\beta_0)}{2\beta_i^2 + \rho}x. \tag{40}$$

It follows that $x_{i+1} = \beta_{i+1}{}^x$, which completes the induction.

Letting $\rho \to 0$, we see that

$$\beta_{i+1} = \frac{1}{2\beta_i} + \frac{\beta_i}{2}, \qquad \beta_0 = \beta \tag{41}$$

which is the well-known Newton method approximating sequence for $\sqrt{1}$. This sequence converges to one for any value of $\beta > 0$, thus showing the convergence of the IKF/Gauss–Newton method as the observation becomes more accurate.

On the other hand, since the EKF update is the first iterate of the IKF, and since $\beta_0 = \beta$, it follows that

$$x_1 \to \frac{\beta^2 + 1}{2\beta}x, \qquad P^+ \to 0 \quad \text{as} \quad \rho \to 0.$$

Thus the EKF update converges to a biased value if $\beta \neq 1$, even though the error estimate converges to zero.

Finally, consider the maximum likelihood update equations (10) and (11). These take the form

$$\hat{x}^+ = \xi_2 x \tag{42}$$

$$P^+ = \rho\begin{bmatrix} (2+\rho)^{-1} & 0 \\ 0 & (2\xi_2^2 + \rho)^{-1} \end{bmatrix} \tag{43}$$

where λ is the largest root of the cubic polynomial

$$\xi_2^3 + (\rho - 1)\xi_2 - \beta\rho.$$

Derivation of the polynomial follows directly from the fact that $\hat{x}^+$ is a solution to the maximum likelihood equation (13). The calculation of P^+ can be made using (17).

Note that the largest root of the cubic polynomial approaches one as $\rho \to 0$, showing that the maximum likelihood estimate is convergent independent of β.

References

[1] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, p. 343, 1973.

[2] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[3] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: M.I.T. Press, p. 190, 1974.

[4] H. W. Sorenson, "Kalman filtering techniques," in *Kalman Filtering: Theory and Application*, H. W. Sorenson, Ed. New York: IEEE Press, p. 90, 1966.

[5] C. Radhakrishna Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.

[6] R. P. Wishner, J. A. Tabaczynski, and M. Athans, "A comparison of three nonlinear filters," *Automatica*, vol. 5, pp. 487–489, 1969.