ROBUST RECURSIVE ALGORITHM FOR NONLINEAR STATE ESTIMATION

ALEXEI R. PANKOV
Department of Applied Mathematics
Moscow Aviation Institute
Moscow,127080, Russia

ALEXEI V. BOSOV
Department of Applied Mathematics
Moscow Aviation Institute
Moscow,127080, Russia

ANDREI V. BORISOV
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117

Technical Report 92-27

April, 1992.

ROBUST RECURSIVE ALGORITHM FOR NONLINEAR STATE ESTIMATION

ALEXEI R. PANKOV, [1] ALEXEI V. BOSOV, [2] ANDREI V. BORISOV[3]

Abstract. The nonlinear recursive filtering algorithm for the discrete-time stochastic systems with observation outliers is considered. The results of numerical tests are given.

Key Words. Conditionally-minimax nonlinear filter, robust nonlinear recursive filter, optimal robust filtering.

1 Introduction.

It is well-known that the Kalman filter (KF) gives the optimal in a mean-square sense estimate of the linear system state only if the system and observation noises and the initial conditions are gaussian (Liptser, and Shiryayev, 1977). If for example the observation errors contain outliers, the KF may work very poor, i.e., the estimate accuracy is low. Many papers are devoted to the problem of robustifying the KF (Masreliez, and Martin, 1977), (Ershov 1978), (Subba Rao, and Yar, 1984), (Barton, and Poor, 1990). This fact shows the practical importance of the problem.
The problem of optimal robust filtering is a nonlinear filtering problem (Liptser, and Shiryayev, 1977), the solution of which leads to very complicated numerical algorithms. In this paper we shall construct the robust nonlinear recursive filter (RNRF) based on the conditionally-minimax nonlinear filter (CMNF) (Pankov, 1990). The last method makes it possible to solve the general nonlinear filtering problem It is based on the idea of conditionally-optimal nonlinear filter (Pugachev, 1979), (Pugachev, and Sinitsyn, 1990), (Raol, and Sinha, 1987). By taking into account the specific character of the robust filtering problem and using the CMNF method we obtain a rather effective RNRF algorithm, which is not a robust KF modification. In this paper we consider the RNRF structure for a linear difference stochastic observation system, the corresponding estimate, statistical properties and numerical results, which allow us to compare the accuracy of the RNRF and the robust KF (Ershov 1978). We also briefly consider the RNRF for nonlinear difference stochastic systems.

[1]Department of Applied Mathematics, Moscow Aviation Institute, Volokolamskoye shosse, 4, Moscow, 127080, Russia
[2]Department of Applied Mathematics, Moscow Aviation Institute, Volokolamskoye shosse, 4, Moscow, 127080, Russia
[3]Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor MI 48109-2117

2 Problem formulation.

We shall use the following notations: $||x|| = (x^T W x)^{1/2}$ for some known weight matrix $W : W = W^T$, $W \geq 0$; $p(x|m, S)$ - the multidimensional gaussian probability density with the mean m and covariance matrix S ; $P(m, S)$ - the set of all random vectors x with $\mathbf{E}\{x\} = m$ and $cov(x, x) = S$.

Let us consider the following stochastic observation model:

$$\left\{ \begin{array}{l} y_n = a_n y_{n-1} + w_n, \quad n = 1, 2, ...; \quad y_0 = \eta, \\ z_n = c_n y_n + v_n, \end{array} \right. \quad (1)$$

where $y_n \in R^p$ is a state vector, $\eta \in R^p$ is a vector of initial conditions with the density $p(x|m_\eta, R_\eta)$; $z_n \in R^q$ is an observation vector; a_n , c_n are known $(p * p)$ and $(q * p)$ matrices respectively; $\{w_n\}$ and $\{v_n\}$ are independent random white noises with the one-dimensional densities $f_w(x, n)$ and $f_v(x, n)$ of the following form:

$$f_w(x, n) = p(x|m_v, R_v), \quad (2)$$

$$f_v(x, n) = (1 - \delta)p(x|m_1, Q_1) + \delta p(x|m_2, Q_2), \quad (3)$$

where m_w , R_w , m_i , $Q_i > 0$ $(i = 1, 2)$ are known multidimetsional parameters; $\delta \in [0, 1]$ is a probability of outliers in observation (the density of outliers is $p(x|m_2, Q_2)$).
From (1),(2) it follows that $\{y_n\}$ is the gaussian process. Equations (2),(4) define the standard observation model with outliers (Kassam, and Poor,1985). It should be mentioned that in (2),(3) all distribution parameters depend on n , but we shall omit this for simplicity.
Consider the problem of a robust recursive process $\{y_n\}$ estimation given the vector of observations $Z^n = col(z_n, ..., z_1)$ in the context of conditionally-minimax filtering theory (Pankov,1990).
Let $\hat{y}_{n-1}$ be the CMNF estimate of y_{n-1} given Z^{n-1}. The estimate $\hat{y}_n$ of y_n takes the form

$$\left\{ \begin{array}{l} \tilde{y}_n = a_n \hat{y}_{n-1} + m_w, \\ \hat{y}_n = \tilde{y}_n + \phi_n^*(\zeta_n(\tilde{y}_n, z_n)), \end{array} \right. \quad (4)$$

where $\zeta_n(\tilde{y}_n, z_n)$ is some known basic correction function (BCF) (the choice of BCF is considered in section 3); $\tilde{y}_n$ is a prediction of y_n given Z_{n-1} which is based on the equation (1); $\phi_n^*(.)$ is the vector-function obtained from the optimality condition

$$\phi_n^*(\nu_n) = arg \min_{\phi \in \Phi} \max_{P(m,S)} \mathbf{E}\{||y_n - \tilde{y}_n - \phi_n(\nu_n)||^2\}, \quad (5)$$

where Φ is the set of all measurable square integrable functions; $x = col(\Delta\tilde{y}_n, \nu_n) \in P(m, S)$ with known m, S ; $\Delta\tilde{y}_n = y_n - \tilde{y}_n$ is a prediction error.

The equation (5) shows that the CMNF belongs to the class of algorithms with a "predictor-corrector" structure, and $\phi_n^*(.)$ provides the mean-square optimal correction of $\tilde{y}_n$ under the additional condition that the correcting term is an arbitrary function of the BCF ν_n. The probability distribution of the random vector x is unknown, but the moments of the first and second order m, S are known (or can be calculated in some way). Hence, $\phi_n^*(.)$ is a conditionally-minimax correction on the set of distributions $P(m, S)$. If the distribution of x is known, we can obtain the optimal correction $\phi_n^o(\nu_n) = \mathbf{E}\{\Delta\tilde{y}_n|\nu_n\}$, but numerical evaluation of it has the same level of complexity as the original nonlinear filtering problem.
If the solution of (5) exists, the equation (4) defines the recursive CMNF. To obtain the robust modification of CMNF we need to choose the appropriate $\zeta_n(y, z)$ and determine the function $\phi_n^*(.)$ from (5). The corresponding results are given in the next section.

3 RNRF for the linear model.

In order to choose $\nu_n = \zeta_n(\tilde{y}_n, z_n)$ we consider the additional problem of the random vector X estimation given the observation vector $Y = CX + V$, where X has the density $p(x|m_x, R_x)$ (i.e. X is gaussian) ; V is a vector of observation error which is gaussian with outliers:
$f_V(x|n) = (1-\delta)p(x|m_1, Q_1) + \delta p(x|m_2, Q_2)$.
Let us obtain the mean-square optimal estimate $\hat{X}$ of X given Y:
$\hat{X} = \gamma^*(Y)$, where

$$\gamma^*(Y) = arg \min_{\gamma \in \Phi} \mathbf{E}\{||X - \gamma(Y)||^2\}$$

.

LEMMA 1. *Let X and V be independent, then*

$$\left\{ \begin{array}{l} \gamma^*(Y) = (1-\chi(Y))\hat{X}_1(Y) + \chi(Y)\hat{X}_2(Y), \\ \chi(Y) = \delta f^2(Y)[(1-\delta)f^1(Y) + \delta f^2(Y)]^{-1}; \end{array} \right. \quad (6)$$

$$\left\{ \begin{array}{l} f^i(y) = p(y|Cm_x + m_i, CR_xC^T = Q_i), \quad i = 1,2; \\ \hat{X}_i(Y) = m_x + R_xC^T(CR_xC^T + Q_i)^+(y - Cm_x - m_i). \end{array} \right. \quad (7)$$

Proof of Lemma: see Appendix A.

Denote $m_x = \mathbf{E}\{\Delta\tilde{y}_n\}$, $R_x = cov(\Delta\tilde{y}_n, \Delta\tilde{y}_n)$ and $C = c_n$. Let us take the BCF $\zeta_n(\tilde{y}_n, z_n)$ as follows

$$\zeta_n(\tilde{y}_n, z_n) = \gamma_n^*(z_n - c_n\tilde{y}_n), \quad (8)$$

where $\gamma_n^*(.)$ is defined in (6),(7). Taking into account $z_n - c_n\tilde{y}_n = c_n\Delta\tilde{y}_n + v_n$ we obtain from Lemma 1 that the BCF (8) provides the optimal correction $\tilde{y}_n$ under the condition of gaussianity of $\Delta\tilde{y}_n$. Hence, $\zeta_n(\tilde{y}_n, z_n)$ as in (8) provides the robust property of the estimation algorithm. The real distribution law of $\Delta\tilde{y}_n$ slightly differs from gaussian, hence the filter accuracy may be raised by the nonparametric filter optimization in conformity with (5).
First we state a preliminary result concerning the minimax problem (5). Let $z = col(x, y) \in P(m, S)$. Denote $m_x = \mathbf{E}\{x\}$, $m_y = \mathbf{E}\{y\}$, $S_x = cov(x, x)$, $S_y = cov(y, y)$, $S_{xy} = cov(x, y) = Syx^T$.
LEMMA 2. *Let $J(\phi, F) = \mathbf{E}\{||x - \phi(y)||^2\}$. If z has the distribution $F \in P(m, S)$ then*
$J(\phi^*, F) \leq J(\phi^*, F^*) \leq J(\phi, F^*)$
where$\phi^*(y) = S_{xy}S_y^+ y + (m_x - S_{xy}S_y^+ m_y)$, and F^* is the gaussian distribution with the expectation m and covariance S.

Proof of Lemma 2: see (Pankov, Bosov, and Borisov, 1992).

THEOREM 1. *Let $y_0 = \mathbf{E}\{\eta\}$, $\nu_n = \zeta_n(\tilde{y}_n, z_n)$ satisfy (7)-(9), then*
i) the solution of (5) exists for all $n \geq 1$ and

$$\begin{cases} \phi_n^*(\nu_n) = H_n\nu_n + h_n, \\ H_n = cov(\Delta\tilde{y}_n, \nu_n)cov^+(\nu_n, \nu_n); \\ h_n = -H - n\mathbf{E}\{\nu_n; \end{cases} \quad (9)$$

ii)$\hat{y}_n$ is an unbiased estimate of y_n with the covariance matrix $\hat{K}_n$ of the estimate error $\Delta\hat{y}_n = y_n - \hat{y}_n$, where

$$\hat{K}_n = a_n\hat{K}_{n-1}a_n^T + R_w - H_n cov(\nu_n, \Delta\tilde{y}_n), \quad n \geq 1; \quad \hat{K}_0 = R_\eta. \quad (10)$$

Proof of Theorem 1: see Appendix B.

Corollary. Let $\mathbf{E}$ { —— $y_n - \mathbf{E}\{y_n\}||^2\} \leq L < \infty, \quad n \geq 0$, then RNRF (4) is nondivergent, i.e. $\mathbf{E}\{||\Delta\tilde{y}_n||^2\} \leq L < \infty$.

Proof of the statement follows from (10).

Let $W = 1$ and $\hat{K}_{n-1} \leq L$. Then

$$\mathbf{E}\{||\Delta\hat{y}_n||^2\} = \mathbf{tr}\{\hat{K}_n\} \leq \mathbf{tr}\{a_n\hat{K}_{n-1}a_n^T\} \leq$$
$$\mathbf{tr}\{a_n cov(y_{n-1}, y_{n-1})a_n^T + R_w\} = \mathbf{tr}\{cov(y_n, y_n)\} = \mathbf{E}\{||y_n - \mathbf{E}\{y_n\}||^2\} \leq L.$$

Hence, the estimate error variance of the arbitrary component of y_n is less or equal to the variance of this component for any $n \geq 0$. From Theorem 1 it follows that $\tilde{y}_n$ is also an unbiased estimate of y_n as $\Delta\tilde{y}_n = a_n\Delta\hat{y}_{n-1}$ and $\mathbf{E}\{\Delta\hat{y}_{n-1}\} = 0$. The

accuracy of $\tilde{y}_n$ is determined by its error covariance matrix $\tilde{K}_n$.
Consider the final system of equations, which defines the RNRF for system (1)-(3):

$$\tilde{y}_n = a_n \hat{y}_{n-1} + m_w, \quad n \geq 1 \quad \hat{y}_0 = m_\eta; \tag{11}$$

$$\begin{cases} \hat{y}_n = \tilde{y}_n + H_n \nu_n + h_n, \\ \nu_n = (1 - \chi(\epsilon_n))\hat{X}_n^1(\epsilon_n) + \chi(\epsilon_n)\hat{X}_n^2(\epsilon_n); \;\; \epsilon_n = z_n - c_n \tilde{y}_n; \end{cases} \tag{12}$$

$$\begin{cases} \hat{X}_n^i(\epsilon_n) = \tilde{K}_n n c_n^T (c_n \tilde{K}_n c_n^T + Q_i)^{-1}(\epsilon_n - m_i); \;\; \tilde{K}_n = a_n \hat{K}_{n-1} a_n^T + R_w; \\ \chi(\epsilon_n) = \delta f_n^2(\epsilon_n)[(1-\delta) f_n^1(\epsilon_n) + \delta f_n^2(\epsilon_n)]^{-1}; \\ f_n^i(\epsilon_n) = p(\epsilon_n | m_i, c_n \tilde{K}_n c_n^T + Q_i); \;\; i = 1, 2; \end{cases} \tag{13}$$

$$\begin{cases} H_n = cov(\Delta \tilde{y}_n, \nu_n) cov^+(\nu_n, \nu_n); \quad h_n = -H_n \mathbf{E}\{\nu_n\}; \\ \hat{K}_n = a_n \tilde{K}_n - H_n cov(\nu_n, \Delta \tilde{y}_n), \quad n \geq 1; \quad \hat{K}_0 = R_\eta. \end{cases} \tag{14}$$

Formula (11) defines the prediction, (12) defines the correction procedure, equations (13) define the robust BCF, and (14) allows one to calculate the optimal filter coefficients H_n , h_n and its error covariance $\hat{K}_n$.
It should be mentioned that in the case of gaussian observation errors without any outliers (i.e. $\delta = 0$) algorithm (1)-(14) coincides with the KF algorithm, i.e.

$$\hat{y}_n = \tilde{y}_n + \tilde{K}_n c_n^T (c_n \tilde{K}_n c_n^T + Q_i)^{-1} \tilde{\epsilon}_n,$$

where $\{\tilde{\epsilon}_n\} = \epsilon_n - m_1$ is the innovation sequence.
There is another interesting case, which leads to the minimax interpretation of the KF algorithm. From Lemma 2 it follows that if we use the linear KF equations (Sage,and Melsa, 1972) together with the model (1)-(3), but considering the observation error $\{v_n\}$ as a gaussian white noise with the expectation $(1-\delta)m_1 + \delta m_2$ and the covariance matrix $(1-\delta)Q_1 + \delta Q_2$, we obtain the estimate $\hat{y}_n^k$, which possesses the minimax property, and the $\hat{y}_n^k$ accuracy is guaranteed for given values of δ , m_1 , m_2 , Q_1 and Q_2.

4 RNRF for the nonlinear system.

Consider the following nonlinear observation model:

$$y_n = a_n(y_{n-1}) + b_n(y_{n-1}) w_n, \quad n \geq 1; \quad y_0 = \eta; \tag{15}$$

$$z_n = g_n(y_n) + v_n, \tag{16}$$

where $a_n(y)$, $b_n(y)$, $g_n(y)$ are known nonlinear vector-functions. All properties of η, $\{w_n\}$, $\{v_n\}$ considered in section 2 hold.
To derive the robust CMNF let us introduce the second structural function $\xi_n(y)$ - the basic prediction function (BPF). Denote $\hat{\xi}_n = \xi_n(\hat{y}_{n-1})$ and consider the conditionally

minimax prediction $\tilde{y}_n$ of the following type

$$\tilde{y}_n = \psi_n^*(\hat{\xi}_n), \tag{17}$$

where $\psi_n^*(.)$ is defined by the condition

$$\psi_n^*(\hat{\xi}_n) = arg \min_{\psi \in \Phi} \max_{P(m_0,S_0)} \mathbf{E}\{||y_n - \psi(\hat{\xi}_n)||^2\}, \tag{18}$$

where $m_0 = \mathbf{E}\{x_0\}$; $S_0 = cov(x_0, x_0)$ and $x_0 = col(y_0, \hat{\xi}_n)$.
From (17),(18) it follows that $\tilde{y}_n$ is a function of BPF $\hat{\xi}_n$ and provides the mean square optimal prediction of y_n for the most unfavourable joint distribution of y_n and $\hat{\xi}_n$ from the set $P(m_0, S_0)$. Consider the sufficient conditions for the RNRF for (15) to exist and the structure of the corresponding filtering algorithm.

THEOREM 2. *Let for $a_n(y)$, $b_n(y)$, $g_n(y)$ from (15,(16) and $\xi_n(y)$ there exist $\alpha_n, \beta_n < \infty$ such that for all $y \in R^p$ and $n \geq 1$ the following inequalities are correct:*

a)

$$||a_n(y)|| + ||b_n(y)|| + ||g_n(y)|| + ||\xi_n(y)|| \leq \beta_n(1 + ||y||^{\alpha_n});$$

b) there exists

$$d_n(y) = \partial g_n(y)/\partial y,$$

and

$$||d_n(y)|| \leq \beta_n(1 + ||y||^{\alpha_n}).$$

Then the RNRF exists and is defined by the equations

$$\tilde{y}_n = F_n\hat{\xi}_n + f_n; \quad n \geq 1; \quad \hat{y}_0 = m_\eta, \tag{19}$$

where $F_n = cov(y_n, \hat{\xi}_n)cov^+(\hat{\xi}_n, \hat{\xi}_n)$, $f_n = \mathbf{E}\{y_n\} - F_n\mathbf{E}\{\hat{\xi}_n\}$, and formulae (12)-(14), where $\epsilon_n = z_n - g(\tilde{y}_n)$, $c_n = d(\tilde{y}_n)$ and $\tilde{K}_n = cov(y_n, y_n) - F_n cov(\hat{\xi}_n, y_n)$. The estimates $\tilde{y}_n$ and $\hat{y}_n$ are unbiased and obtain the error covariance matrices $\tilde{K}_n$ and $\hat{K}_n$ respectivly.

The proof of the Theorem 2 is based on the fact that under the conditions of Theorem 2, y_n and $\hat{y}_n$ have finite moments of arbitrary order for any $n \geq 0$. Consequently, it looks like the proof of Theorem 1, hence we omit the details.

Consider some obvious types of BPF $\xi_n(y)$
a) $\xi_n(\hat{y}_{n-1}) = \hat{y}_{n-1}$ -linear BPF;
b) $\xi_n(\hat{y}_{n-1}) = a_n(y_{n-1}) + b(y_{n-1})m_w$-the BPF based on the dynamic system equations (15);

c) $\xi_n(\hat{y}_{n-1}) = col(e_1(y_{n-1}), ..., e_N(y_{n-1}))$, where $\{e_i(y_{n-1})\}$ is a system of linearly independent functions on R^P (the BPF of the general type).

5 Numerical examples.

From the equations (14) and (19) it follows thar RNRF parameters are independent of particular measurement trajectory $\{z_n\}$ and depend only on the moment characteristics of the vector $x = col(\Delta\tilde{y}_n, \nu_n)$ and $x_0 = col(y_n, \xi_n)$. Consequently they can be calculated in advance and store in a computer memory. There are several methods for determinating F_n , f_n , H_n and h_n (Raol and Sinha, 1987) , (Pugachev, and Sinitsin, 1990), developed for the conditionally-optimal filtering and based on the evaluation of the joint characteristic function of y_n and $\hat{y}_n$. These methods are very complicated and require many a priori computations. We think that much more effictive method is the method based on the joint computer recursive statistical modelling of y_n and $\hat{y}_n$ and the statistical treatment of the simulations results. The details of this method are given in previous papers (Pankov,1990). Using these methods we obtain several estimation results which allow us to compare the RNRF method with others finite-dimensional filtering algorithms with the optimal one.

Consider the following observation model:

$$\begin{cases} y_n = 0.9y_{n-1} + w_n, & ; y_0 = \eta; \\ z_n = 2y_n + v_n. \end{cases} \quad (20)$$

1. The distributions of w_n , v_n and η are given in section 2. We use the following concrete distribution parameters:
$m_\eta = m_w = m_1 = m_2 = 0.0$, $R_w = 0.49$, $R_\eta = 1.0$, $Q_1 = 1.0$, $Q_2 = 100.0$, $\delta = 0.2$. We simulated on a computer $N = 5000$ system (20) trajectories and obtain the following estimates.

$\hat{y}_n^0$ - the mean-square optimal estimate, determined by the special numerical algorithm of the spline-appriximation of the conditional density of the system state (Pankov, Bosov, and Borisov, 1992);

$\hat{y}_n^k$ - the minimax linear estimate; given by the KF algorithm (see section 2);

$\hat{y}_n^{rk}$ - the estimate of the robust KF (Ershov 1978).;

$\hat{y}_n$ - the RNRF estimate.

The corresponding numerical results are given in Table 1.

Table 1.

n	$\hat{K}_n^0$	$\hat{K}_n$	$\hat{K}_n^{rk}$	K_n^k
1	0.592	0.618	0.712	1.040
2	0.490	0.539	0.605	1.061
3	0.444	0.500	0.600	1.071
4	0.449	0.481	0.638	1.077
5	0.429	0.460	0.632	1.079
6	0.385	0.473	0.618	1.081
7	0.392	0.478	0.650	1.081
8	0.392	0.471	0.665	1.082
9	0.420	0.441	0.669	1.082
10	0.409	0.441	0.655	1.082

We see that $\hat{y}_n$ is much more accurate than $\hat{y}_n^{rk}$, and $\hat{y}_n^k$ is close to $\hat{y}_n^0$. The computation time of $\hat{y}_n$ is practically the same as of $\hat{y}_n^k$ and sufficiently smaller than of $\hat{y}_n^0$.

2. In order to check the importance of the normality condition for $\Delta\tilde{y}_n$ (see Lemma 1) we consider the system (20) with two types of nongaussian disturbances $\{w_n\}$ and initial condition η: the uniform distribution and the Laplace distribution. All parameters of these distributions were chosen as in (21). We calculate the RNRF-estimate $\hat{y}_n$ and the optimal one $\hat{y}_n^0$. the resuts are given in Table 2.

Table 2.

n	$\hat{K}_n^0(uniform)$	$\hat{K}_n(uniform)$	$\hat{K}_n^0(Laplace)$	$K_n(Laplace)$
1	0.547	0.614	0.680	0.726
2	0.431	0.484	0.510	0.598
3	0.392	0.464	0.477	0.551
4	0.397	0.422	0.493	0.518
5	0.381	0.421	0.464	0.505
6	0.375	0.432	0.480	0.492
7	0.388	0.422	0.453	0.489
8	0.422	0.465	0.472	0.511
9	0.411	0.426	0.479	0.500
10	0.390	0.461	0.471	0.497

From the obtained results it follows that the influence of the deviations in the $\Delta\tilde{y}_n$ distribution from the gaussian one is negligible.

6 Appendix A.

Proof of Lemma 1.

Define $f_X(x) = p(x|m_x, R_x)$, $f_i(x) = p(x|m_i, Q_i)$, $i = 1,2$. Then the probability density of V is $f_V(x) = (1-\delta)f_1(x) + \delta f_2(x)$. It could be easily checked that in this case Y has the probability density

$$f_Y(y) = (1-\delta)f^1(y) +^2 (y), \tag{21}$$

where $f^i(y) = p(y|Cm_x + m_i, CR_xC^T + Q_i)$, $i = 1,2$.
The conditional density $f(x|y)$ of X given Y may be expressed in the form $f(x|y) = f(y|x)f_X(x)[f_Y(y)]^{-1}$, hence

$$\gamma^*(y) = \int_{R^p} xf(x|y)dx = \int_{R^p} xf(y|x)f_X(x)dx[f_Y(y)]^{-1}.$$

Obviously $f(y|x) = (1-\delta)f_1(Y - Cx) + \delta f_2(y - Cx)$, then

$$f(y|x)f_X(x) = (1-\delta)f^1(y)f_1(x|y) + \delta f^2(y)f_2(x|y),$$

where $f_i(x|y)$ is the conditional density of X given Y under the assumption that V has the density $f_i(x)$, $i = 1,2$. Then

$$\gamma^*(y) = [f_Y(y)]^{-1}[(1-\delta)f^1(y)\int_{R^p} xf_1(x|y)dx + \delta f^2(y)\int_{R^p} xf_2(x|y)dx] =$$

$$[f_Y(y)]^{-1}[(1-\delta)\gamma_1^*(y) + \delta\gamma_2^*(y)],$$

where γ_i^* is the optimal estimator of X given Y if V has the density $f_i(x)$, $i = 1,2$.
From the normal correlation theorem (Liptser, and Shiryayev, 1977) it follows that if $Q_i > 0$

$$\gamma_i^*(y) = m_x + R_xC^T(CR_xC^T + Q_i)^{-1}(y - Cm_x - m_i)), \quad i = 1,2. \tag{22}$$

Denote $\chi(y) = \delta f^2(y)[f_Y(y)]^{-1}$, $\hat{X}_i(Y) = \gamma_i^*(Y)$, then for the optimal estimate $\hat{X} = \gamma^*(Y)$ from (21) , (22) we obtain

$$\hat{X} = (1 - \chi(Y))\hat{X}_1(Y) + \chi(y))\hat{X}_2(Y)$$

.

7 Appendix B.

Proof of Theorem 1.

From the conditions of Theorem 1 it follows that $\mathbf{E}\{||y_0||^2\} = tr\{WR_\eta\} + ||m_\eta||^2 < \infty$, then for all $n \geq 1$

$$\mathbf{E}\{||y_n||^2\} \leq 2(||a_n||^2\mathbf{E}\{||y_{n-1}||^2\} + \mathbf{E}\{||w_n||^2) < \infty,$$

Let for some $n \geq 1$ we have
$\mathbf{E}\{||y_{n-1}||^2\} < \infty$, $\mathbf{E}\{y_{n-1} - \hat{y}_{n-1}\} = 0$.
For $\Delta\tilde{y}_n = y_n - \tilde{y}_n$ we obtain $\mathbf{E}\{||\Delta\tilde{y}_n||^2\} \leq 2\mathbf{E}\{||y_n||^2 + ||\tilde{y}_n||^2\} < \infty$.
Let us show that in this case
$\mathbf{E}\{||\zeta_n(\tilde{y}_n, z_n)||^2\} = \mathbf{E}\{||\gamma^*(\epsilon_n)||^2\} < \infty$,
where $\epsilon_n = z_n - c_n\tilde{y}_n = c_n\Delta\tilde{y}_n + v_n$. Obviously, $\mathbf{E}\{||\epsilon_n||^2\} < \infty$ since we have $\mathbf{E}\{||\Delta\tilde{y}_n||^2\} < \infty$ and $\mathbf{E}\{||v_n||^2\} = tr\{W[(1-\delta)Q_1+\delta Q_2]\} + ||(1-\delta)m_1+\delta m_2||^2 < \infty$.
$\gamma^*(\epsilon_n) = (1 - \chi(\epsilon_n))\hat{X}_1(\epsilon_n) + \chi(\epsilon_n)\hat{X}_2(\epsilon_n)$
and besides $\chi(\epsilon_n) \in [0,1]$ with probability 1. Hence, it is sufficient to show that $\mathbf{E}\{||\hat{X}_i(\epsilon_n)||^2\} < \infty$, $i = 1,2$.
$\hat{X}_i(\epsilon_n) = \hat{K}_n c_n^T(c_n\hat{K}_n c_n^T + Q_i)^{-1}(\epsilon_n - m_i)$,
hence there exists $D_n^i < \infty$, such that $||\hat{X}_i(\epsilon_n)||^2 \leq D_n^i(1 + ||y||^2)$.
So, $\mathbf{E}\{||\hat{X}_i(\epsilon_n)||^2\} < D_n^i(1 + ||\mathbf{E}\{\epsilon_n||^2\}) < \infty$. Then for $x = \Delta\tilde{y}_n$, $y = \gamma^*(\epsilon_n)$ the random vector $z = col(x,y) \in P(m,S)$ where $m = \mathbf{E}\{z\}$, $S = cov(z,z)$ with $tr\{WS\} + ||m||^2 < \infty$. Then from Lemma 1 it follows that
$\hat{y}_n = \tilde{y}_n + \phi_n^*(\zeta_n(\tilde{y}_n, z_n)) = \tilde{y}_n + H_n\gamma^*(\epsilon_n) + h_n$, where $H_n = cov(x,y)cov^+(y,y)$, $h_n = -H_n\mathbf{E}\{y\}$.
It should be mentioned that in this special case $m_x = \mathbf{E}\{\Delta\tilde{y}_n\} = a_n\mathbf{E}\{y_{n-1} - \tilde{y}_{n-1}\} = 0$.
Let now $n = 1$ and $\hat{y}_0 = m_\eta$. Hence $\mathbf{E}\{||\hat{y}_0||^2\} = ||m_\eta||^2 < \infty$ and $\mathbf{E}\{y_0 - \tilde{y}_0\} = 0$.
Now the result follows from the mathematical induction principle.

References

Barton,R.J., and Poor,N.V., 1990 Robust estimation and signal processing. *IEEE Trans. on Inform. theory,* IT-36, 485-501.

Ershov,A.A., 1978 Stable methods of estimating parameters (survey).. *Automat. and Remote Contr.*, 39, Number 3

Kassam,S.A., and Poor,N.V., 1985 Robust techniques for signal processing: a survey. *Proceedings IEEE,* 73, Number 3, 433-481.

Liptser,R.S., and Shiryayev,A.N., 1977 em Statistics of random processes. N.-Y.,Springer Verlag.

Masreliez C., and Martin,R., 1977 Robust bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. on Autom. Control*, AC-27, 361-371.

Pankov,A.R., 1990 Synthesis of conditionally-optimal filters by computer simulation. *Analysis and synthesis of dynamic systems under uncertainty*, Moscow, MAI Press, 69-75 (in Russian).

Pankov,A.R., 1990 Recursive dynamic system trajectory estimation by means of regression nonlinear filter. *Statistical methods in flight control*, Moscow, MAI Press, 117-118 (in Russian).

Pankov,A.R., 1990 Synthesis of conditionally-minimax filters for nonlinear stochastic systems by method of modeling. *The All-Union Conference on Dynamic, Control and Operation Research*, Moscow, MAI Press, 117-118 (in Russian).

Pugachev,V.S., 1979 Estimation of variables and parameters in discrete-time nonlinear systems. *Automat. and Remote Control*, 40,512-521.

Pugachev,V.S., and Sinitsyn,I.N., 1990 Stochastic differential systems, London, J.Wiley & Sons.

Raol,S.R., and Sinha,N.K., 1987 Conditionally optimal state estimation for systems, governed by difference equations. *Canadian Electr. Engin. J.*, 12, 71-77.

Sage,A.P., and Melsa,J.L. 1972 Estimation theory with application to communication and control. New York, McGrow-Hill.

Subba Rao T., and Yar M. 1984 Linear and nonlinear filters for linear but nongaussian processes. *Int. J. Contr.*,39, Number 1, 235-246.

Pankov,A.R., Bosov,A.V., and Borisov,A.V. 1992 Finite-dimensional algorithms of nonlinear system state estimation. University of Michigan, Technical Report 92-13.