STATIONARY BEHAVIOR OF AN ANTI-WINDUP SCHEME FOR RECURSIVE PARAMETER ESTIMATION UNDER LACK OF EXCITATION

Magnus Evestedt and Alexander Medvedev

*Department of Information Technology, Uppsala University,
P. O. Box 337, SE-951 05, SWEDEN*

Abstract: Stationary properties of a recently suggested windup prevention scheme for recursive parameter estimation are investigated in the case of insufficient excitation. When the regressor vector contains data covering the whole parameter space, the algorithm has only one stationary point, the one defined by a weighting matrix. If the excitation is insufficient, the algorithm is shown to possess a manifold of stationary points and a parametrization of this manifold is given. However, if the past excitation conditions already caused the algorithm to converge to a certain point, the stationary solution would not be affected by current lack of excitation. This property guarantees good anti-windup properties of the studied parameter estimation algorithm. *Copyright© 2005 IFAC*

Keywords: Recursive estimation, Riccati equation, windup

1. INTRODUCTION

Consider the following regressor model

$$y(t) = \varphi^T(t)\theta + e(t) \tag{1}$$

where $y(t)$ is the scalar output measured at discrete time instances $t = [0, \infty)$, $\varphi \in R^n$ is the regressor vector, $\theta \in R^n$ is the parameter vector to be estimated and the scalar e is disturbance.

1.1 Recursive estimation

The estimation of θ is often performed by a linear recursive algorithm of the "prediction-correction" form

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\left(y(t) - \varphi^T(t)\hat{\theta}(t-1)\right) \tag{2}$$

where the first (prediction) term in the right hand part of the equation highlights the fact that the parameter vector is assumed to be constant in absence of information on parameters variations.

If $e(t)$ is white and the parameter vector is subject to the random walk model driven by a zero-mean white sequence $w(t)$

$$\theta(t) = \theta(t-1) + w(t) \tag{3}$$

the optimal, in the sense of minimum of the *a posteriori* parameter error covariance matrix, estimate is yielded by (2) with the Kalman gain

$$K(t) = P(t)\varphi(t)$$

where $P(t)$, $t = [1, \infty)$ is the solution to the Riccati equation

$$P(t) = P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{r(t) + \varphi^T(t)P(t-1)\varphi(t)} + Q(t) \tag{4}$$

for some $P(0) = P^T(0)$, $P(0) \geq 0$ describing the covariance of the initial guess of $\hat{\theta}(t), t = 0$. Optimality of the estimate is guaranteed only when Q is the covariance of w and $r(t) = \text{var}\, e(t)$, (Ljung and Gunnarsson, 1990). Since these quantities are seldom *a priori* known, they are usually treated as design parameters of the estimation algorithm and chosen as $Q(\cdot) \in R^{n \times n}, Q(\cdot) \geq 0, r(t) > 0$ in order to achieve some desired properties of the filter.

1.2 The role of excitation

Excitation properties of the regressor vector sequence play an important role in the dynamic behavior of (4). When the excitation in the input signal is insufficient, a phenomenon referred to as (covariance) windup (aka blow-up) can occur. This means that some eigenvalues of P tend to be very large. Here the term "insufficient excitation" refers to the situation when the regressor data do not cover the whole parameter space. The more commonly used term "persistent excitation" also assumes that the excitation is sufficient under a significant period of time and therefore is more restrictive. In Section 3, the notion of sufficient excitation is defined in mathematical terms.

The mechanism behind windup in the Riccati equation can be explained from e. g. the random walk model. Indeed, when $Q > 0$, all the elements in θ vary. At the same time, the fact that the excitation is not sufficient means that variations in some of the parameters cannot be observed from the system output y. Since $P(t)$ describes the covariance of the estimation error $\theta - \hat{\theta}$, the eigenvalues of $P(\cdot)$ corresponding to the unobservable elements of θ grow because the uncertainty of the corresponding estimates is increasing at each step.

Another complication caused by lack of excitation is that the solutions to Riccati equation (4) do not have to be symmetric and unique. This also has implications for the standard proof of stability of the Kalman filter given in (Jazwinski, 1970), where $P^{-1}(t)$ is used to form a Lyapunov function. Furthermore, the stabilizing solution provides the upper bound for all real *symmetric* solutions of the Riccati equation and *asymmetric* solutions do not have to obey this bound. Since the Riccati equation has to be solved on-line in the Kalman filter, special precautions have to be taken to avoid convergence to undesirable solutions. This can be easily achieved by e. g. propagating only the elements of $P(t)$ on and over the main diagonal. However, asymmetrical solutions have been found to be useful in control applications,(Cloutier and Stansbery, 1999). No similar examples in state estimation are available to the authors.

1.3 Previous work

As pointed out in (Cao and Schwartz, 2004), the windup phenomenon in the Kalman filter has not been much analyzed until recently. Therefore most of the suggested anti-windup schemes for Kalman filter parameter estimation are of *ad hoc* nature and are lacking strict proof of non-divergence under insufficient excitation, see e. g. (Hägglund, 1983), (Bittanti *et al.*, 1990).

In the approach taken in (Stenlund and Gustafsson, 2002), in the sequel referred to as the Stenlund-Gustafsson (SG) algorithm, a special choice of $Q(t)$ is used to control the convergence point of the P-matrix:

$$Q(t) = \frac{P_d \varphi(t) \varphi^T(t) P_d}{r(t) + \varphi^T(t) P_d \varphi(t)} \quad (5)$$

where $P_d \in R^{n \times n}, P_d > 0$.

The structure of (5) is designed to reflect the fact that Q, loosely speaking, specifies in what direction the solution to the Riccati equation has to be updated at each step. If the sequence of $\varphi(t), t \in [0, \ldots \infty)$ is persistently exciting, the matrix Q can be chosen to be positive definite since the whole parameter space R^n is covered by the excitation. However, dealing with non-stationary data, persistent excitation can not always be guaranteed. Therefore, it is reasonable to choose $Q(t)$ so that $P(t)$ is updated only in the subspace where excitation is present, that is $\text{Im} \, Q(t) = \text{Im} \, \varphi \varphi^T$. Addition of $r > 0$ in the denominator of (5) prevents division by zero when $\varphi(t) = 0$ for some t.

A formal proof of the fact that (4) with the free term chosen according to (5) is non-diverging both for the cases of sufficient and insufficient excitation can be found in (Medvedev, 2003; Medvedev, 2004). However, stationary properties of the scheme are not considered there. The weighting matrix, P_d, constitutes the main degree of freedom in the SG-algorithm. A good choice of P_d is the stationary point of the Kalman filter for the specific problem. If a non-stationary data set can be considered as piecewise stationary, a number of different P_d can be used, one for each interval of stationarity, (Evestedt and Medvedev, 2005).

The SG algorihm can be seen as a generalization of the normalized least mean squares (N-LMS), a method that is well-known and widely used in engineering practice. In (Ljung and Gunnarsson, 1990), it is shown that the N-LMS can be obtained as a special case of the Kalman filter with the parameters

$$P_d = \alpha I, \alpha \in R^+; \; P(0) = P_d; \; r(t) = 1 \quad (6)$$

in the equations (4), (5). In the N-LMS, the Riccati equation becomes redundant since it is initiated at its stationary point. Thus, the resulting filter (2),(4),(5),(6) becomes insensitive to loss of excitation. Similarly, in the SG algorithm, once the Riccati equation has converged to the stationary point P_d, it becomes robust against insufficient excitation, in the sense that the solution does not diverge. The mechanism behind this behavior is explained in more detail in Section 3 of this paper.

Interestingly, a directional tracking algorithm presented in (Cao and Schwartz, 2004), the one identified as Algorithm 1, also is very close to the N-LMS. The suggested choice of the free term in (4) is

$$Q(t) = \frac{\gamma \varphi(t) \varphi^T(t)}{\epsilon + \varphi^T(t) \varphi(t)}$$

where $\gamma > 0$ and $\epsilon > 0$ are some scalars. This algorithm becomes equivalent to the N-LMS with $\epsilon \gamma = 1$ and being initiated at the stationary point of (4), i. e. $P(0) = \gamma I$. Thus, it is also a special case of the SG algorithm.

It is as well worth to note at this point that the proofs of boundedness of the recursive estimation algorithms in (Cao and Schwartz, 2004) are based on the assumption that all considered solutions to the Riccati equation are positive semidefinite. As it is already mentioned before, this cannot be guaranteed under lack of excitation.

1.4 Contribution of the present paper

The main result of the paper is formulated in Proposition 4 and provides an explicit parametrization of all stationary solutions to the Riccati equation arising in the SG algorithm. Under sufficient excitation, the parametrization implies that the stationary point is unique and is pre-assigned by the matrix P_d. When the excitation is insufficient, the parametrization defines the manifold of all possible solutions, as well asymmetric ones.

The paper is organized as follows. First an equivalent linear time-varying form of the Riccati equation in the SG algorithm, (4),(5) is provided. The equation itself was used before, see e. g. (Stenlund and Gustafsson, 2002) and (Medvedev, 2003; Medvedev, 2004), but only a proof for the case when $P > 0$ had been originally given in (Stenlund and Gustafsson, 2002). Then, using the linear form, stationary points of (4),(5) are investigated and some results on the behavior of the algorithm under insufficient excitation are presented.

2. SYLVESTER EQUATION FORM

In (Stenlund and Gustafsson, 2002), it is shown that, for non-singular $P(\cdot)$, the difference $E(t) = P(t) - P_d$ obeys the recursion

$$E(t+1) = A_t^{-1}(P(t))E(t)A_t^{-T}(P_d) \qquad (7)$$

where $A_t(X) = I + r^{-1}(t)X\varphi(t)\varphi^T(t)$. When excitation is insufficient, positive definiteness of the solution to the Riccati equation cannot be guaranteed. Therefore, before analyzing anti-windup properties of the SG algorithm, it is important to check whether (7) also holds under milder conditions.

It can be proved that the quadratic Riccati equation for the SG algorithm is equivalent to a linear time-varying matrix equation without restricting the solutions to the non-singular ones.

Proposition 1. Equation (4) can, with the special choice of Q as (5), be rewritten as the following discrete Sylvester difference equation

$$P(t) = A_t^{-1}(P(t-1))P(t-1)A_t^{-T}(P_d) \qquad (8)$$
$$- A_t^{-1}(P(t-1))P_d A_t^{-T}(P_d) + P_d$$

where $A_t(X) = I + r^{-1}(t)X\varphi(t)\varphi^T(t)$, $X \in R^{n \times n}$.

Proof: See Appendix. □

The equation above is a discrete Sylvester difference equation, linear in P if the dependence of $A_t(P(t))$ can be seen as a general time variance. This way of thinking is widely used in e. g. handling non-linear systems via linear time-varying models. Riccati equation (4) is being embedded into a broader class of linear time-varying matrix equations

$$P(t) = A_t^{-1}(Y(t))P(t-1)A_t^{-T}(P_d) \qquad (9)$$
$$- A_t^{-1}(Y(t))P_d A_t^{-T}(P_d) + P_d$$

for arbitrary $Y(t) \in R^{n \times n}, t = 0, 1, \ldots$. Trajectories of (4) are the same as those of (9) only when $Y(t) = P(t-1)$. It is also clear that the structure of (9) does not necessarily imply that the solution is symmetric.

The linear character of (7) becomes more obvious when it is written in vectorized form with respect to $e(\cdot) = \text{vec } E(\cdot)$ (see e. g. (Horn and Johnson, 1991))

$$e(t+1) = M(P_d, P(t))e(t)$$
$$M(P_d, P(t)) = A_t^{-1}(P_d) \otimes A_t^{-1}(P(t)) \qquad (10)$$

where $\otimes$ denotes Kronecker (tensor) product.

3. STATIONARY POINTS

The purpose of this section is to study stationary points of Riccati equation (4) with the special choice of the free term (5) arising in the SG algorithm. The stationary solutions are evaluated both for the case of complete and insufficient excitation.

Consider a stationary point of (7)

$$E = E(t+1) = E(t)$$

Then the following algebraic condition holds

$$E = A_t^{-1}(P(t))EA_t^{-T}(P_d) \qquad (11)$$

In vectorized form (11) becomes

$$(M(P_d, P(t)) - I)e = 0, \ e = \text{vec } E \qquad (12)$$

In order to separate the direction of excitation at each particular time instant from its intensity, introduce a re-parametrization of the matrix function $A_t(X)$,

$$A_t(X) = I + \gamma X U(t) \qquad (13)$$

where $\gamma(t) = r^{-1}(t)\varphi(t)^T\varphi(t)$ and

$$U(t) = \frac{\varphi(t)\varphi^T(t)}{\varphi^T(t)\varphi(t)}$$

The matrix $U(t)$ is a Hermitian projection with rank $U(t) = 1$. Define the normalized eigenvectors of $U(t)$ as $\xi_i(t), i = 1, \ldots, n$, where $\xi_1(t)$ corresponds to the unit eigenvalue of $U(t)$ and $\xi_2(t), \ldots, \xi_n(t)$ correspond to the zero eigenvalues of $U(t)$. Then $\gamma(t)$ describes the energy in the regressor vector at time t and $\xi_1(t)$ characterizes the direction.

The regression vector $\varphi(t)$ is called persistently exciting in m steps, (Söderström and Stoica, 1989),

if there exists a constant $0 < c < \infty$ and an integer $m > 0$ such that for all t,

$$\sum_{k=t+1}^{t+m} \varphi(k)\varphi^T(k) \geq cI \qquad (14)$$

This means that when $\varphi(t)$ is persistently exciting, the space R^n is spanned by $\varphi(t)$ in m steps.

Excitation is called sufficient at time t when the following rank condition is satisfied

$$\text{rank} \begin{bmatrix} \xi_1(t+n-1) & \dots & \xi_1(t) \end{bmatrix} = n \qquad (15)$$

In terms of (14) this means that $m = n$ and (15) does not have to hold for *all* t.

The argument of $\xi_1(t)$ and $\gamma(t)$ is often suppressed in the sequel for brevity when it does not lead to confusion.

Define the spectrum of $X \in R^{n \times n}$ as

$$\sigma(X) = \{\lambda_i(X), i = 0, \dots, n\}$$

Due to the Kronecker product structure of $M(\cdot, \cdot)$, the spectrum of it is easy to evaluate.

Proposition 2. The matrix

$$N(P_d, P(t)) = M(P_d, P(t)) - I$$

in (12) has the eigenvalues

$$\sigma(N(P_d, P(t))) = \Big\{ \frac{1 - (\gamma\xi_1^T P(t)\xi_1)(\gamma\xi_1^T P_d\xi_1)}{(\gamma\xi_1^T P(t)\xi_1)(\gamma\xi_1^T P_d\xi_1)},$$

$$\underbrace{\frac{-\gamma\xi_1^T P(t)\xi_1}{1 + \gamma\xi_1^T P(t)\xi_1}, \dots, \frac{-\gamma\xi_1^T P(t)\xi_1}{1 + \gamma\xi_1^T P(t)\xi_1}}_{n-1},$$

$$\underbrace{\frac{-\gamma\xi_1^T P_d\xi_1}{1 + \gamma\xi_1^T P_d\xi_1}, \dots, \frac{-\gamma\xi_1^T P_d\xi_1}{1 + \gamma\xi_1^T P_d\xi_1}}_{n-1}, \underbrace{0, \dots, 0}_{(n-1)^2} \Big\} \qquad (16)$$

Proof: See Appendix. □

The eigenvectors of $N(\cdot, \cdot)$ are as well easily obtained.

Proposition 3. The eigenvectors corresponding to the zero eigenvalues of $N(P_d, P(t))$ are $x_k = \xi_i \otimes \xi_j$, $i = 2, \dots, n, j = 2, \dots, n, k = 1, \dots, (n-1)^2$, where ξ_i, $i = 2, \dots, n$ are the eigenvectors corresponding to the zero eigenvalues of $U(t)$.

Proof: See Appendix. □

Now all the necessary partial results are in place to formulate the main contribution of the paper. The proposition below completely characterizes the space of all possible stationary solutions of (4), (5).

Proposition 4. A stationary solution $P(t) = P^* = $ const of (8) can, for a given P_d, be decomposed as

$$P^* = P_d + \sum_{i=2}^{n} \sum_{j=2}^{n} k_{ij}\xi_i\xi_j^T \qquad (17)$$

where k_{ij} are scalars. When the input signal is sufficiently exciting, the stationary solution is exactly P_d, i. e. $k_{ij} = 0$, $i = 2, \dots, n$, $j = 2, \dots, n$.

Proof: See Appendix. □

With each time step (4),(5) converges towards one of the possible solutions of (11). When the input signal is persistently exciting, the vectors spanning $\text{Ker } U(t)$ at each time instant are linearly independent and information about the whole parameter space is eventually collected by the algorithm. Thus, the solution will converge to P_d which is the stationary point of (4), (5). If, however, the input signal is not persistently exciting, convergence to P_d cannot be guaranteed.

Examining the structure of (11), one can conclude that a stationary point of (4),(5) does not have to be a semi-definite or even symmetric matrix. Neither it has to be bounded. However, the only source of disturbance in solving the Riccati equation is numerical errors and it is unlikely that their structure will fit $\xi_i\xi_j^T, i = 2, \dots, n, j = 2, \dots, n$ and their magnitude will be significant.

Example 1. Proposition 4 implies that all possible solutions (17) are symmetric for a regressor equation of dimension two. However, as the numerical example below demonstrates, asymmetrical solutions can appear already for a third order equation.

Let P_d be an identity matrix. Further assume that the excitation is not sufficient and the regressor vector is of the form $\varphi(t) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. Then it is easy to check that there are stationary solutions to (4) and (5) of the form

$$P^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \kappa \\ 0 & 0 & 1 \end{bmatrix} \qquad (18)$$

where κ is any number. Clearly $||P^*||$ is unbounded. The above matrix is asymmetrical for $\kappa \neq 0$.

Ending up the discussion on stationary solutions, it is proved that all the stationary points of (4),(5) result in one and the same Kalman gain.

Proposition 5. All the stationary points P^*, yield the same Kalman gain $K(t) = P_d\varphi(t)$.

Proof: See Appendix. □

4. CONCLUSION

Stationary properties of a recently suggested windup prevention method for recursive parameter estimation are studied in the case of non-persistently exciting data. A particular choice of the free term of the Riccati equation suggested by the method imposes linear

dynamics on the difference Riccati equation and simplifies its analytical analysis.

Generalizing a known result, it is shown that the resulting Riccati equation can *always* be written as a Sylvester equation. By a direct use of this parametrization, the manifold of all stationary solutions of the Riccati equation is evaluated and demonstrated to include as well indefinite and non-symmetric matrices. The corresponding Kalman gain is though unique at each step.

When the excitation is persistent, the stationary point is unique and equal to a pre-defined matrix.

ACKNOWLEDGEMENTS

REFERENCES

Bittanti, S., P. Bolzern and M. Campi (1990). Convergence and exponential convergence of identification algorithms with directional forgetting factor. *Automatica* **26**(5), 929–932.

Cao, L. and H.M. Schwartz (2004). Analysis of the Kalman filter based estimation algorithm: an orthogonal decomposition approach. *Automatica* **40**, 5–19.

Cloutier, J. and D. Stansbery (1999). Control of a continuously stirred tank reactor using an asymmetric solution of the state-dependent Riccati equation. In: *Proceedings of the 1999 IEEE Conference on control applications*. Hawaii, USA. pp. 893–898.

Evestedt, M. and A. Medvedev (2005). Windup properties of recursive parameter estimation algorithms in acoustic echo cancellation. In: *Proceedings of the 16th IFAC World Congress*.

Hägglund, T. (1983). *New estimation techniques for adaptive control*. Lund Institute of Technology, Sweden.

Horn, R.A. and C.R. Johnson (1991). *Topics in matrix analysis*. Cambridge University Press.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Ljung, L. and S. Gunnarsson (1990). Adaptation and tracking in system identification - a survey. *Automatica* **26**(1), 7–21.

Medvedev, A. (2003). Stability of a windup prevention scheme in recursive parameter estimation. In: *Proceedings of the 42nd IEEE Conference on Decision and Control*. Maui, Hawaii, USA.

Medvedev, A. (2004). Stability of a Riccati equation arising in recursive parameter estimation under lack of excitation. *IEEE Transactions on Automatic Control* **49**(12), 2275–2280.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice Hall.

Stenlund, B. and F. Gustafsson (2002). Avoiding windup in recursive parameter estimation. In: *Preprints of reglermöte 2002*. Linköping, Sweden. pp. 148–153.

Appendix A. PROOF OF PROPOSITION 1

First, consider the inverse of $A(X)$, that can be found using the matrix inversion lemma as follows,

$$
\begin{aligned}
A^{-1}(X) &= \left(I + r^{-1} X \theta \theta^T \right)^{-1} \qquad (A.1) \\
&= I - r^{-1} X \theta \left(1 + \theta^T r^{-1} X \theta \right)^{-1} \theta^T \\
&= I - \frac{X \theta \theta^T}{r + \theta^T X \theta}
\end{aligned}
$$

Then define the scalars $\alpha = \varphi^T P_d \varphi$ and $\beta = \varphi^T P(t-1) \varphi$. Starting from the equality

$$
\begin{aligned}
0 =\;& \frac{\alpha P(t-1)\varphi\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} - \frac{\alpha P(t-1)\varphi\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} \\
& + \frac{\beta P(t-1)\varphi\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} - \frac{\beta P(t-1)\varphi\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))}
\end{aligned}
$$

some algebra results in the following

$$
\begin{aligned}
0 =\;& \frac{P(t-1)\varphi\varphi^T P_d}{\beta + r(t)} - \frac{P(t-1)\varphi\varphi^T P_d}{\alpha + r(t)} \\
& + \frac{P(t-1)\varphi\beta\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} - \frac{P(t-1)\varphi\alpha\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))}
\end{aligned}
$$

Adding and subtracting P_d to the equation above, equation (4) can be rewritten as,

$$
\begin{aligned}
P(t) =\;& P(t-1) - \frac{P(t-1)\varphi\varphi^T P(t-1)}{r(t) + \beta} \\
& + \frac{P_d \varphi\varphi^T P_d}{r(t) + \alpha} + \frac{P(t-1)\varphi\varphi^T P_d}{\beta + r(t)} - \frac{P(t-1)\varphi\varphi^T P_d}{\alpha + r(t)} \\
& + \frac{P(t-1)\varphi\beta\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} - \frac{P(t-1)\varphi\alpha\varphi^T P_d}{(\alpha + r(t))(\beta + r(t))} \\
& + P_d - P_d
\end{aligned}
$$

The above expression can be formulated as a sum of two matrix products and P_d:

$$
\begin{aligned}
P(t) =\;& \left(P(t-1) - \frac{P(t-1)\varphi\varphi^T P(t-1)}{\beta + r(t)} \right) \\
& \left(I - \frac{P_d \varphi\varphi^T}{\alpha + r(t)} \right)^T \\
& - \left(P_d - \frac{P(t-1)\varphi\varphi^T P_d}{\beta + r(t)} \right) \left(I - \frac{P_d \varphi\varphi^T}{\alpha + r(t)} \right)^T + P_d \\
=\;& \left(I - \frac{P(t-1)\varphi\varphi^T}{\beta + r(t)} \right) P(t-1) \left(I - \frac{P_d \varphi\varphi^T}{\alpha + r(t)} \right)^T \\
& - \left(I - \frac{P(t-1)\varphi\varphi^T}{\beta + r(t)} \right) P_d \left(I - \frac{P_d \varphi\varphi^T}{\alpha + r(t)} \right)^T + P_d
\end{aligned}
$$

Following equation (A.1), the bracketed matrices in the products are the inverses of $A_t(\cdot)$ which concludes the proof.

Appendix B. PROOF OF PROPOSITION 2

From Proposition 1 in (Medvedev, 2003; Medvedev, 2004) and since $A_t^{-1}(P) > 0$ we have

$$\sigma(A_t^{-1}(P)) = \left\{ \frac{1}{1 + \gamma \xi_1^T P \xi_1}, 1, \ldots, 1 \right\}$$

and

$$\sigma(A_t^{-1}(P_d)) = \left\{ \frac{1}{1 + \gamma \xi_1^T P_d \xi_1}, 1, \ldots, 1 \right\}$$

A well-known result on the eigenvalues and eigenvectors of the Kronecker product of matrices formulated below is necessary for the analysis in the sequel.

Lemma 1. Let $A \in R^{n \times n}$ and $B \in R^{m \times m}$. If $\lambda \in \sigma(A)$ and $x \in R^n$ is a corresponding eigenvector of A, and if $\mu \in \sigma(B)$ and $y \in R^m$ is a corresponding eigenvector of B, then $\lambda\mu \in \sigma(A \otimes B)$ and $x \otimes y \in R^{nm}$ is a corresponding eigenvector of $A \otimes B$.

Proof: See page 245 in (Horn and Johnson, 1991). $\square$

Now consider the Kronecker product matrix $N(P_d, P(t))$. From Lemma 1 the corresponding eigenvalues can be calculated as (16).

Appendix C. PROOF OF PROPOSITION 3

From Proposition 1 in (Medvedev, 2003; Medvedev, 2004) it is known that the eigenvectors corresponding to the unit eigenvalues of $A(X)$ are ξ_i, $i = 2, \ldots, n$. Now from Lemma 1 we can calculate the eigenvectors corresponding to the zero eigenvalues of $N(P_d, P(t))$ as $x_l = \xi_i \otimes \xi_j$, $i = 2, \ldots, n$, $j = 2, \ldots, n$, $l = 1, \ldots, n - 1$.

Appendix D. PROOF OF PROPOSITION 4

Consider the matrix $N(P_d, P^*)$, where P^* is the stationary solution to (8). Note that $N(P_d, P^*)$ is time variant due to its dependence on the regressor vector $\varphi(t)$.

Let the vectors $\xi_2(t), \ldots, \xi_n(t)$ be the eigenvectors corresponding to the zero eigenvalues of $U(t)$, spanning $\mathrm{Ker}\, U(t)$ and let $\xi_1(t)$ be the eigenvector corresponding to the unit eigenvalue of $U(t)$, spanning $\mathrm{Im}\, U(t)$.

Then by Proposition (3), P^* can be written as

$$P^* = P_d + \sum_{i=2}^n \sum_{j=2}^n k_{ij} \xi_i \xi_j^T$$

for some scalars k_{ij}. Let ξ_{ij} denote the jth element in the ith eigenvector. Then the above equation can be rewritten as a sum of matrices,

$$P^* = P_d + \Big[\sum_{i=2}^n \sum_{j=2}^n (k_{ij} \xi_{j1}) \xi_i$$

$$\sum_{i=2}^n \sum_{j=2}^n (k_{ij} \xi_{j2}) \xi_i \ \ldots \ \sum_{i=2}^n \sum_{j=2}^n (k_{ij} \xi_{jn}) \xi_i \Big]$$

The columns of $P^* - P_d$ must therefore lie in $\mathrm{Ker}\, U(t)$.

For a sufficiently exciting signal we have, according to (15) that

$$\mathrm{rank}\, [\xi_1(t + n - 1) \ldots \xi_1(t)] = n$$

Now consider $P^* - P_d$ at the time instants $t = \tau, \ldots, \tau + n - 1$. Since $P^* - P_d$ is constant its columns must be in the intersection of the nullspaces $\bigcap_{t=\tau}^{\tau+n-1} \mathrm{Ker}\, U(t)$.

At time $t = \tau$, since $U = U^T$, $R^n = \mathrm{Ker}\, U(\tau) \oplus \xi_1(\tau)$ and at time $t = \tau + 1$, $R^n = \mathrm{Ker}\, U(\tau + 1) \oplus \xi_1(\tau + 1)$. Thus $R^n = (\mathrm{Ker}\, U(\tau) \cap \mathrm{Ker}\, U(\tau + 1)) \oplus \mathrm{Im}\, U(\tau) \oplus \mathrm{Im}\, U(\tau + 1)$. Proceeding in the same way for $t = \tau + 2, \ldots, \tau + n - 1$ we get

$$R^n = \bigcap_{t=\tau}^{\tau+n-1} \mathrm{Ker}\, U(t) \oplus \bigoplus_{t=\tau}^{\tau+n-1} \xi_1(t) \qquad \text{(D.1)}$$

Due to the sufficiently exciting signal the direct sum $\bigoplus_{t=\tau}^{\tau+n-1} \xi_1(t) = R^n$, which means

$$\bigcap_{t=\tau}^{\tau+n-1} U(t) = \oslash$$

according to (D.1).

The columns of $P^* - P_d$ can thus not lie in the same nullspace for $t = \tau \ldots \tau + n - 1$, which means $k_{ij} = 0$, $i = 2, \ldots, n$; $j = 2, \ldots, n$.

Appendix E. PROOF OF PROPOSITION 5

Consider the Kalman gain $K(t) = P(t)\varphi(t)$ in a stationary point P^* under insufficient excitation

$$K(t) = P^* \varphi(t)$$

which using Proposition 4 can be rewritten as

$$K(t) = P_d \varphi(t) + \left(\sum_{i=2}^n \sum_{j=2}^n k_{ij} \xi_i \xi_j^T \right) \varphi(t)$$

Since $U(t)$ is symmetrical, $R^n = \mathrm{Im}\, U(t) \oplus \mathrm{Ker}\, U(t)$. Now, the eigenvectors ξ_i, $i = 2, \ldots, n$ span $\mathrm{Ker}\, U(t)$ and $\varphi(t) = c\xi_1$, $c \in R$ spans $\mathrm{Im}\, U(t)$. Then $\varphi(t)$ and ξ_i $i = 2, \ldots, n$ are orthogonal and

$$K(t) = P_d \varphi(t)$$