# The equation of optimal filtering, Kalman's filter and Theorem on normal correlation.

Markovich Liubov

Institute of Control Sciences
Moscow, Russia.

26 июня 2014 г.

# Outlook

- Definitions and problems
- The equation of optimal filtering
- Part 1: Equivalence of equation of optimal filtering and Kalman's filter
  - Auxiliary result 1: The explicit form of the conditional density
- Part 2: Equivalence of equation of optimal filtering and Theorem on normal correlation
  - Auxiliary result 2: explicit covariance matrix inversion
- References

# Definitions

Let $(S_n, X_n)_{n\geq 1}$ be a partly observable Markov random sequence, where

- $S = (S_n)_{n\geq 1}$ is an unobservable part (useful signal),
- $X = (X_n)_{n\geq 1}$ is an observable part.

Relation between them is given by

$$X_n = \varphi(S_n, \eta_n), \tag{1}$$

where $(\eta_n \in \mathbb{R})_{n\geq 1}$ is an i.i.d random sequence and $(S_n)_{n\geq 1}$ is a Markov sequence.

Realizations of r.v.s $S_n \in \mathcal{S}_n \subseteq \mathbb{R}$ and $X_n \in \mathcal{X}_n \subseteq \mathbb{R}$ are denoted as $s_1^n = (s_1, \ldots, s_n)^T$ and $x_1^n = (x_1, \ldots, x_n)^T$ respectively.

# Definitions

Let us define the random sequence $\vartheta_n = Q(S_n)$, where $Q : \mathcal{S}_n \to \Theta_n$ is an one-to-one function and $\vartheta_n \in \Theta_n \subseteq \mathbb{R}$.

The random sequence $(\vartheta_n)_{n \geq 1}$ is also a Markov sequence.

To estimate $\vartheta_n$ we use estimator in the form of a conditional mean

$$\widehat{\vartheta}_n = \mathsf{E}(Q(S_n)|x_1^n) = \int_{\mathcal{S}_n} Q(s_n) w_n(s_n|x_1^n) ds_n. \tag{2}$$

where $w_n(s_n|x_1^n)$ is the posterior probability density function (Stratonovich: 1966).

# Problem

$w_n(s_n|x_1^n)$ satisfies the recurrence equation (Stratonovich: 1966)

$$w_1(s_1|x_1) = \frac{f(x_1|s_1)p(s_1)}{\int\limits_{\mathcal{S}_1} f(x_1|s_1)p(s_1)ds_1}, \quad \text{and for} \quad n \geq 2$$

$$w_n(s_n|x_1^n) = \frac{f(x_n|s_n)}{f(x_n|x_1^{n-1})} \int\limits_{\mathcal{S}_{n-1}} p(s_n|s_{n-1}) w_{n-1}(s_{n-1}|x_1^{n-1}) ds_{n-1}. \quad (3)$$

- $p(s_n|s_{n-1})$ is the transition probability density functions of the Markov sequence $(S_n)_{n\geq 1}$,
- $f(x_n|x_1^{n-1})$ and $f(x_n|s_n)$ are conditional densities.

Problem:

Since $w_n(s_n|x_1^n)$ depends on an unknown prior distribution function $p(s_1)$ and the unknown transition probability $p(s_n|s_{n-1})$ of the Markov sequence $(S_n)_{n>1}$, we can not use formula (2) to estimate $\widehat{\vartheta}_n$.

# Conditional density and its derivative

Integrating (3) over $s_n$

$$\int\limits_{S_n} w_n(s_n|x_1^n)ds_n = \int\limits_{\mathcal{S}_n} \frac{f(x_n|s_n)}{f(x_n|x_1^{n-1})} \int\limits_{\mathcal{S}_{n-1}} p(s_n|s_{n-1})w_{n-1}(s_{n-1}|x_1^{n-1})ds_{n-1}ds_n$$

and transferring $f(x_n|x_1^{n-1})$ to the left side of the equation we get

$$f(x_n|x_1^{n-1}) = \int\limits_{\mathcal{S}_n} f(x_n|s_n) \int\limits_{\mathcal{S}_{n-1}} p(s_n|s_{n-1})w_{n-1}(s_{n-1}|x_1^{n-1})ds_{n-1}ds_n. \quad (4)$$

Derivative of (4) by $x_n$ leads to

$$f'_{x_n}(x_n|x_1^{n-1}) = \int\limits_{\mathcal{S}_n} f'_{x_n}(x_n|s_n) \int\limits_{\mathcal{S}_{n-1}} p(s_n|s_{n-1})w_{n-1}(s_{n-1}|x_1^{n-1})ds_{n-1}ds_n. \quad (5)$$

# Exponential family

Let us assume that the conditional density $f(x_n|s_n)$ belongs to the exponential family of density distributions

$$f(x_n|s_n) = \widetilde{C}(s_n)h(x_n)\exp(T(x_n)Q(s_n)), \tag{6}$$

where $\widetilde{C}(s_n)$ is a normalization constant, $h(x_n), T(x_n), Q(s_n)$ are known functions.
Its derivative by $x_n$ is determined by

$$f'_{x_n}(x_n|s_n) = f(x_n|s_n)\left(\frac{h'_{x_n}(x_n)}{h(x_n)} + T'_{x_n}(x_n)Q(s_n)\right).$$

# Logarithmic derivative of conditional density $f(x_n|x_1^{n-1})$

Substituting $f'_{x_n}(x_n|s_n)$ in equation for $f'_{x_n}(x_n|x_1^{n-1})$ (5) we can deduce that

$$f'_{x_n}(x_n|x_1^{n-1}) = \frac{h'_{x_n}(x_n)}{h(x_n)} f(x_n|x_1^{n-1})$$
$$+ \quad T'_{x_n}(x_n) \int\limits_{\mathcal{S}_n} f(x_n|s_n) Q(s_n) \int\limits_{\mathcal{S}_{n-1}} p(s_n|s_{n-1}) w_{n-1}(s_{n-1}|x_1^{n-1}) ds_{n-1} ds_n.$$

Dividing the latter equation on $f(x_n|x_1^{n-1})$ and due to (3) we can write

$$\frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} = \frac{h'_{x_n}(x_n)}{h(x_n)} + T'_{x_n}(x_n) \int\limits_{\mathcal{S}_n} Q(s_n) w_n(s_n|x_1^n) ds_n.$$

# The equation of optimal filtering (Dobrovidov: 1983)

$$\mathsf{E}(Q(S_n)|x_1^n) \cdot T'_{x_n}(x_n) = \left(\ln\left(\frac{f(x_n|x_1^{n-1})}{h(x_n)}\right)\right)'_{x_n} \qquad (7)$$

The latter equation does not contain explicitly probabilistic characteristics $p(s_1)$ and $p(s_n|s_{n-1})$ of the unknown sequence $(S_n)_{n\geq 1}$. This allows us to find the optimal estimator $\widehat{\vartheta}_n = \mathsf{E}(Q(S_n)|x_1^n)$ knowing only observable quantities $x_1^n$.

Our objective is to derive that

equation (7) coincides with Kalman's filter and the theorem on normal correlation in case of gaussian density $f(x_n|s_n)$.

# Part 1:

Equivalence of equation of optimal filtering (7) and Kalman's filter.

# Equation (7) for gaussian density

As an example of the exponential family we can take gaussian density

$$f(x_n|s_n) = \frac{1}{\sqrt{2\pi}B} \exp\left(-\frac{(x_n - As_n)^2}{2B^2}\right).$$

Then the observation model is defined by a linear equation

$$X_n = AS_n + B\eta_n,$$

where $\eta_n$ is the i.i.d random sequence with Gaussian distribution, coefficients $A$ and $B$ are given real numbers.

Then the general filtration equation (7) has the form

$$\mathrm{E}(S_n|x_1^n) = \frac{B^2}{A}\frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} + \frac{x_n}{A}. \quad (8)$$

# Kalman's filter

Let us consider a partially observable Markov sequence $(S_n, X_n)_{n\geq 1}$ defined by recursive linear equations

$$\begin{aligned} S_n &= aS_{n-1} + b\xi_n, \\ X_n &= AS_n + B\eta_n \end{aligned} \tag{9}$$

where $S_n, X_n \in \mathbb{R}$ for all $n$; $\xi_n$ and $\eta_n$ are mutually independent random variables with the standard gaussian distribution,

$$S_0 \in \mathcal{N}(0, \widetilde{\sigma}^2), \quad \widetilde{\sigma}^2 = \frac{b^2}{1 - a^2}$$
$$S_n \in \mathcal{N}(0, 1), \quad n = 1, 2, 3 \ldots$$

Coefficients $A, B, a, b$ are given real numbers and $|a| < 1$.

# Kalman's filter (Continuation)

Kalman's filter for the linear system (9) is defined by following recursive equations

$$\mathrm{E}(S_{n+1}|x_1^{n+1}) = a\mathrm{E}(S_n|x_1^n) + \frac{Ab^2 + a^2A\gamma_n}{B^2 + A^2b^2 + A^2a^2\gamma_n}(x_{n+1} - Aa\mathrm{E}(S_n|x_1^n))$$

$$\gamma_n = \frac{B^2(a^2\gamma_{n-1} + b^2)}{A^2(a^2\gamma_{n-1} + b^2) + B^2} \qquad (10)$$

with the initial conditions

$$\mathrm{E}(S_1|x_1) = \frac{A\tilde{\sigma}^2}{A^2\tilde{\sigma}^2 + B^2}x_1$$

$$\gamma_1 = \frac{B^2\tilde{\sigma}^2}{A^2\tilde{\sigma}^2 + B^2}.$$

# Auxiliary result 1:

The explicit form of the conditional density $f(x_n|x_1^{n-1})$.

# Theorem: The explicit form of the conditional density is

$$f(x_n|x_1^{n-1}) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left(-\frac{1}{2\sigma_n}\left(x_n - A\mathcal{L}_{n-1}\right)^2\right), \quad n \geq 2, \tag{11}$$

where

$$\mathcal{L}_n = \frac{Aa}{\sigma_{n-1}}\left(x_{n-1}æ_{n-1} + \frac{aB^2}{\sigma_{n-2}}\left(x_{n-2}æ_{n-2} + \frac{aB^2}{\sigma_{n-3}}\left(x_{n-3}æ_{n-3} + \dots\right.\right.\right.$$

$$\left.+ \quad \frac{aB^2}{\sigma_2}\left(x_2æ_2 + x_1\frac{aB^2æ_1}{\sigma_1}\underbrace{\big)\dots\big)\big)\big)^2\Big)}_{n}\right), \quad n = 2, 3, \dots. \tag{12}$$

with the following notations

$$æ_1 = \tilde{\sigma}^2, \quad \sigma_1 = B^2 + A^2æ_1, \tag{13}$$

$$æ_n = \frac{B^2a^2æ_{n-1} + \sigma_{n-1}b^2}{\sigma_{n-1}}, \quad \sigma_n = B^2 + A^2æ_n, \quad n \geq 2.$$

# The equation of the optimal filtering

Then the ratio of the density derivative to density itself is

$$\frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} = \frac{A\mathcal{L}_{n-1} - x_n}{\sigma_n}.$$

Then the equation of the optimal filtering (for the gaussian density)

$$\mathrm{E}(S_n|x_1^n) = \frac{B^2}{A}\frac{f'_{x_n}(x_n|x_1^{n-1})}{f(x_n|x_1^{n-1})} + \frac{x_n}{A}.$$

can be represented as

$$\mathrm{E}(S_n|x_1^n) = \frac{Ax_n æ_n}{\sigma_n} + \frac{B^2\mathcal{L}_{n-1}}{\sigma_n}. \tag{14}$$

# Representation in the recursive form

The equation of the optimal filtering is recursive

$$\mathsf{E}(S_{n+1}|x_1^{n+1}) = \frac{Ax_{n+1}}{\sigma_{n+1}}æ_{n+1} + \frac{B^2a}{\sigma_{n+1}}\mathsf{E}(S_n|x_1^n) \quad (15)$$

$$\begin{aligned} æ_1 &= \widetilde{\sigma}^2, \quad \sigma_1 = B^2 + A^2æ_1, \\ æ_n &= \frac{B^2a^2æ_{n-1} + \sigma_{n-1}b^2}{\sigma_{n-1}}, \quad \sigma_n = B^2 + A^2æ_n, \quad n \geq 2. \end{aligned} \quad (16)$$

# Results: equivalence of equation of optimal filtering and Kalman's filter

**Lemma**

Parameters $\gamma_n$ (10) of the Kalman's filter are connected with $æ_n$ (16) as

$$\gamma_n = \frac{B^2 æ_n}{\sigma_n},$$

where $B$ is determined in (9).

**Theorem**

In case a partially observable Markov sequence $(S_n, X_n)_{n\geq 1}$ is defined by

$$S_n = aS_{n-1} + b\xi_n, \quad X_n = AS_n + B\eta_n,$$

the equation of optimal filtering is equivalent to the Kalman's filter.

# Part 2:

Equivalence of equation of optimal filtering (7) and Theorem on normal correlation.

# Theorem on normal correlation

Liptser R. S. , Shiryaev A. N. (2001) (Theorem 3.1, p.61):

For the gaussian vector $(\theta, \nu)$ the optimal estimate $\mathsf{E}(\theta|\nu)$ of the vector $\theta$ by $\nu$ is

$$\mathsf{E}(\theta|\nu) = \mathsf{E}(\theta) + D_{\theta\nu} D_{\nu\nu}^{-1} (\nu - \mathsf{E}(\nu)), \tag{17}$$

where $\mathsf{E}(\theta)$ and $\mathsf{E}(\nu)$ are expectations and

$$\begin{aligned} D_{\theta\nu} &= cov(\theta, \nu) = \|cov(\theta_i, \nu_j)\|, \quad 1 \le i \le k, 1 \le j \le l \\ D_{\nu\nu} &= cov(\nu, \nu) = \|cov(\nu_i, \nu_j)\|, \quad 1 \le i, j \le l \end{aligned} \tag{18}$$

are covariance matrices.

# Theorem on normal correlation in our terms

For Kalman's filter the following conditions on $(S_n, X_n)_{n\geq 1}$ holds

$$\begin{aligned} E(S_0) &= 0, \quad E(S_0^2) = \frac{b^2}{1-a^2}, \\ E(\xi_n) &= 0, \quad E(\eta_n) = 0, \quad E(X_n) = 0, \quad n \geq 1, \\ E(\xi_n^2) &= 1, \quad E(\eta_n^2) = 1, \quad n \geq 1. \end{aligned}$$

**Theorem on normal correlation**

$$E(S_n | x_1^n) = D_{S_n, \vec{X}_n} D^{-1}_{\vec{X}_n, \vec{X}_n} \vec{x}_n$$

Here $\vec{X}_n = (X_1, \ldots, X_n)^T$, $\vec{x}_n = x_1^n = (x_1, \ldots, x_n)^T$.

# Covariance matrices: general view

$$D_{\vec{X}_n,\vec{X}_n} = \begin{pmatrix} var(X_1) & cov(X_1,X_2) & \dots & cov(X_1,X_n) \\ cov(X_2,X_1) & var(X_2) & \dots & cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_n,X_1) & cov(X_n,X_2) & \dots & var(X_n) \end{pmatrix}$$

$$= \begin{pmatrix} A^2æ_1 + B^2 & A^2aæ_1 & \dots & A^2a^{n-1}æ_1 \\ A^2aæ_1 & A^2æ_1 + B^2 & \dots & A^2a^{n-2}æ_1 \\ \vdots & \vdots & \ddots & \vdots \\ A^2a^{n-1}æ_1 & A^2a^{n-2}æ_1 & \dots & A^2æ_1 + B^2 \end{pmatrix} \quad (19)$$

$$D_{S_n,\vec{X}_n} = \begin{pmatrix} cov(S_n,X_1) & cov(S_n,X_2) & \dots & cov(S_n,X_n) \end{pmatrix}$$

$$= Aæ_1 \begin{pmatrix} a^{n-1} & a^{n-2} & \dots & 1 \end{pmatrix}.$$

# Auxiliary result 2:

Explicit matrix inversion $D^{-1}_{\vec{x}_n,\vec{x}_n}$.

# Toeplitz matrix

The covariance matrix (19) is a Toeplitz matrix

$$D_{\vec{x}_n,\vec{x}_n} = \begin{pmatrix} c_1 + B^2 & ac_1 & a^2c_1 & \dots & a^{n-1}c_1 \\ ac_1 & c_1 + B^2 & ac_1 & \dots & a^{n-2}c_1 \\ a^2c_1 & ac_1 & c_1 + B^2 & \dots & a^{n-3}c_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a^{n-1}c_1 & a^{n-2}c_1 & a^{n-3}c_1 & \dots & c_1 + B^2 \end{pmatrix}.$$

$$D_{\vec{x}_n,\vec{x}_n} = (D_{\vec{x}_n,\vec{x}_n})_{B=0} + B^2\mathbf{I}$$

where $\mathbf{I}$ is the identity matrix, $c_1 = A^2æ_1$.

$$(\mathbf{P} + \mathbf{I})^{-1} = \mathbf{P}^{-1} - \mathbf{P}^{-1}\left(\mathbf{I} + \mathbf{P}^{-1}\right)^{-1}\mathbf{P}^{-1},$$

where $\mathbf{P}$ is a squared invertible matrix.

# Toeplitz and Tridiagonal symmetric matrices

Using algorithm from (Trench: 2001) we can find the inverted matrix

$$(D^{-1}_{\vec{X}_n,\vec{X}_n})_{B=0} = \frac{1}{c_1(1-a^2)} \begin{pmatrix} 1 & -a & 0 & \dots & 0 \\ -a & 1+a^2 & -a & \dots & 0 \\ 0 & -a & 1+a^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

We get the tridiagonal symmetric matrix

$$\mathbf{I} + B^2 (D_{\vec{X}_n,\vec{X}_n})^{-1}_{B=0} = \frac{a}{d_0-1} \begin{pmatrix} \frac{d_0}{a} & -1 & 0 & \dots & 0 \\ -1 & \frac{d_0+a^2}{a} & -1 & \dots & 0 \\ 0 & -1 & \frac{d_0+a^2}{a} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{d_0}{a} \end{pmatrix},$$

where $d_0 - 1 = \frac{c_1(1-a^2)}{B^2}$.

# Inverted tridiagonal symmetric matrix

Using theory developed in (Forseca:2007, Usmani:1994) we get

$$(\mathbf{I} + B^2 (D_{\vec{X}_n,\vec{X}_n})^{-1}_{B=0})^{-1} = \frac{d_0 - 1}{a\psi_n} \cdot \begin{cases} \psi_{i-1}\psi_{n-j}, & \text{if} \quad i \le j \\ \psi_{j-1}\psi_{n-i}, & \text{if} \quad i > j. \end{cases}$$

where

$$\psi_m \quad = \quad \left(\frac{d_0 + a^2}{a}\right) \psi_{m-1} - \psi_{m-2}, \quad \text{for} \quad m = (2, \ldots, n-1)$$

$$\psi_n \quad = \quad \frac{d_0}{a} \psi_{n-1} - \psi_{n-2}, \quad \text{with initial conditions} \quad \psi_0 = 1, \psi_1 = \frac{d_0}{a}.$$

## Explicit matrix inversion $D^{-1}_{\vec{x}_n,\vec{x}_n}$

$$D^{-1}_{\vec{x}_n,\vec{x}_n} = \frac{1}{c_1(1-a^2)} \begin{pmatrix} 1 & -a & \dots & 0 \\ -a & 1+a^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

$$- \quad \frac{1}{c_1(1-a^2)a\psi_n} \begin{pmatrix} 1 & -a & \dots & 0 \\ -a & 1+a^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \cdot$$

$$\cdot \quad \begin{pmatrix} \psi_{n-1} & \psi_{n-2} & \dots & \psi_1 & 1 \\ \psi_{n-2} & \psi_1\psi_{n-2} & \dots & \psi_1\psi_1 & \psi_1 \\ \psi_{n-3} & \psi_1\psi_{n-3} & \dots & \psi_2\psi_1 & \psi_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \psi_1 & \dots & \psi_{n-2} & \psi_{n-1} \end{pmatrix} \begin{pmatrix} 1 & -a & \dots & 0 \\ -a & 1+a^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

# The theorem on normal correlation

$$D_{S_n,\vec{X}_n} D^{-1}_{\vec{X}_n,\vec{X}_n} = -\frac{1}{Aa\psi_n}\Big(1 - a\psi_1 \quad -a + (1+a^2)\psi_1 - a\psi_2 \quad \dots$$

$$\dots \quad -a\psi_{n-3} + (1+a^2)\psi_{n-2} - a\psi_{n-1} \quad -a\psi_{n-2} + \psi_{n-1} - a\psi_n\Big)$$

The theorem on normal correlation (17) is

$$\mathsf{E}(S_n|x_1^n) = D_{S_n,\vec{X}_n} D^{-1}_{\vec{X}_n,\vec{X}_n}\vec{x}_n =$$

$$= \frac{a\psi_1 - 1}{Aa\psi_n}x_1 - \frac{a\psi_0 - (1+a^2)\psi_1 + a\psi_2}{Aa\psi_n}x_2 - \frac{a\psi_1 - (1+a^2)\psi_2 + a\psi_3}{Aa\psi_n}x_3$$

$$\dots \quad -\frac{a\psi_{n-3} - (1+a^2)\psi_{n-2} + a\psi_{n-1}}{Aa\psi_n}x_{n-1} - \frac{a\psi_{n-2} - \psi_{n-1} + a\psi_n}{Aa\psi_n}x_n$$

# Results: equivalence of equation of optimal filtering and Theorem on normal correlation

The theorem on normal correlation is

$$\mathsf{E}(S_n|x_1^n) = \frac{a\psi_1 - 1}{Aa\psi_n}x_1 - \frac{a\psi_0 - (1+a^2)\psi_1 + a\psi_2}{Aa\psi_n}x_2 -$$

$$\dots \quad - \frac{a\psi_{n-3} - (1+a^2)\psi_{n-2} + a\psi_{n-1}}{Aa\psi_n}x_{n-1} - \frac{a\psi_{n-2} - \psi_{n-1} + a\psi_n}{Aa\psi_n}x_n$$

The equation of optimal filtering is

$$\mathsf{E}(S_n|x_1^n) = \frac{Aa^{n-1}B^{2(n-1)}æ_1}{\sigma_1 \cdot \ldots \cdot \sigma_{n-2}\sigma_{n-1}\sigma_n}x_1 + \frac{Aa^{n-2}B^{2(n-2)}æ_2\sigma_1}{\sigma_1\sigma_2 \cdot \ldots \cdot \sigma_{n-2}\sigma_{n-1}\sigma_n}x_2 +$$

$$\dots \quad + \frac{AaB^2æ_{n-1}\sigma_1\sigma_2 \cdot \ldots \cdot \sigma_{n-2}}{\sigma_1\sigma_2 \cdot \ldots \cdot \sigma_{n-3}\sigma_{n-2}\sigma_{n-1}\sigma_n}x_{n-1} + \frac{Aæ_n\sigma_1\sigma_2 \cdot \ldots \cdot \sigma_{n-2}\sigma_{n-1}}{\sigma_1\sigma_2 \cdot \ldots \cdot \sigma_{n-3}\sigma_{n-2}\sigma_{n-1}\sigma_n}x_n$$

Their equivalence follows from coinciding of coefficients by $x_1, \dots, x_n$.

# References


- Dobrovidov A. V. (1983) Nonparametric methods of nonlinear filtering of stationary random sequences. Automat. and Remote Control, 44 (6), pp. 757–768.
- Liptser R. S. , Shiryaev A. N. (2001) Statistics of Random Processes: II. Applications, Springer.
- Stratonovich R.L. (1966) Conditional Markovian processes and their application to the optimal control theory, Moscow: Moscow Univ. Press, (in Russian).
- Dobrovidov, A. V, and G.M. Koshkin and V. A. Vasiliev (2012) Non-parametric state space models, Kendrick press, USA
- Gantmacher, F. R. (1990) The Theory of Matrices.
- Trench W. F. (2001) Properties of some generalizations of Kac-Murdock-Szegö matrices.
- Fonseca da C.M. (2007) On the eigenvalues of some tridiagonal matrices, Journal of Computational and Applied Mathematics 200, pp. 283 – 286.