Application of Kalman Filter with alpha-stable distribution

Pavel Mozgunov, *National Research University Higher School of Economics*, pmozgunov@gmail.com[16]

Abstract. In this paper we consider the behavior of Kalman Filter state estimates in the case of distribution with heavy tails .The simulated linear state space models with Gaussian measurement noises were used. Gaussian noises in state equation are replaced by components with alpha-stable distribution with different parameters alpha and beta. We consider the case when "all parameters are known" and two methods of parameters estimation are compared: the maximum likelihood estimator (MLE) and the expectation- maximization algorithm (EM). It was shown that in cases of large deviation from Gaussian distribution the total error of states estimation rises dramatically. We conjecture that it can be explained by underestimation of the state equation noises covariance matrix that can be taken into account through the EM parameters estimation and ignored in the case of ML estimation.

Keywords. Kalman Filter, alpha-stable distribution, MLE, EM-algorithm

1 Introduction

State-space model (SSM) is convenient way for expressing dynamic systems that involve unobserved variables. If the model is linear and noises are Gaussian, the technique of Kalman Filter (KF) [1] can be applied. However, the condition of Gaussian noises in SSM is a strong enough and can significantly restrict the area of application of Kalman Filter. Some modification of Kalman Filter in cases of non-Gaussian noises was proposed in [7], but the case of non-Gaussian disturbances in measurement equation is considered, and the question of parameters estimation is still open. In this paper we used the KF technique in case of non-Gaussian state noises, and studied the behaviour of KF and methods of estimation and find some useful properties of EM-estimation that allows in process of parameters estimation (with out any additional computations) get acceptable results.

[16]This study (research grant No 14-05-0007) was supported by The National Research University Higher School of Economicsâ Academic Fund Program in 2014.

In Section 2 we consider linear SSM in general, and suppose a case of alpha-stable disturbances in state-equation. In Section 3, we overview two methods of SSM parameters estimation: MLE and EM. In section 4, we present the results of Kalman Filter estimation for different parameters of alpha-stable distribution and demonstrate useful properties of EM-algorithm that were found out in this research. We make conclusion in Section 5.

2 State-Space Model and Kalman Filter.

State-space model

Consider linear state-space model

$$X_{k+1} = A_k X_k + V_{k+1} \quad \in R^m, \; k = 0, 1, ... \tag{1}$$

$$Y_k = C_k X_k + W_k \quad \in R^d, \; k = 0, 1, ... \tag{2}$$

Where (1) is called *state equation* and (2) is *measurement equation.* A_k - matrix $m \times m$, V_k - state noise, C_k - matrix $d \times m$, W_k - measurement noise. To be the Kalman Filter optimal estimator in the least squares sense [1], it is necessary that noises and initial state vector should be Gaussian. It means that, $V_k \sim N(0, \delta_{kl} Q_k)$; $W_k \sim N(0, \delta_{kl} R_k)$ and $X_0 \sim N(\mu, \Sigma)$

Alpha -stable distribution.

We replace the Gaussian noises in equation (1) by disturbances with alpha-stable distribution and consider one-dimensional case. α - stable distribution is fully determined by its characteristic function [6]:

$$log\phi(t) = -\sigma^\alpha |t|^\alpha \{1 - i\beta sign(t) tan\frac{\pi\alpha}{2}\} + i\mu t; \alpha \neq 1 \tag{3}$$

$$log\phi(t) = -\sigma |t| \{1 + i\beta sign(t) \frac{2}{\pi} log|t|\} + i\mu t; \alpha = 1 \tag{4}$$

where $\alpha \in (0; 2], \beta \in [-1, 1], \sigma > 0, \mu \in R$, and α - characteristic exponent (refer to heavy tails of distribution); μ - location parameter; σ - scale parameter; β - skewness parameter. We denote the random variable with α - stable distribution in the following way: $X \sim S_\alpha(\sigma, \beta, \mu)$

Class of α -stable distributions was chosen because the case of $\alpha = 2$ and $\beta = 0$ ($S_2(\sigma, 0, \mu)$ corresponds to the Gaussian random variable $N(\mu, 2\sigma^2)$. It means that it can be studied, how deviations in α from 2, affects the sum of Kalman Filter prediction error.

Moreover, we put the distribution of initial state vector to be α-stable, with the same parameter α as in the state noise. Due to the property that the sum of two alpha-stable variables with the same parameter α is a alpha-stable variable again with parameter α, we get that the state vector has α stable distribution with the same α in all moments of time.

To simulate one-dimension alpha-stable distribution we used the same method as in Weron(1996)[5].

For $\alpha \neq 1$

$$X = S_{\alpha,\beta}(\frac{sin\alpha(V + B_{\alpha,\beta})}{(cosV)^{1/\alpha}})(\frac{cos(V - \alpha(V + B_{\alpha,\beta}))}{W})^{(1-\alpha)/\alpha} \tag{5}$$

$$S_{\alpha,\beta} = [1 + \beta^2 tan^2 \frac{\pi\alpha}{2}]^{1/(2\alpha)}$$
$$B_{\alpha,\beta} = \frac{arctan(\beta tan\frac{\pi\alpha}{2})}{\alpha}$$

And for $\alpha = 1$

$$X = \frac{2}{\pi}[(\pi/2 + \beta V)tanV - \beta log(\frac{\frac{\pi}{2}W cosV}{(\pi/2) + \beta V})] \tag{6}$$

Where V is uniformly distributed in interval $[-\pi/2; \pi/2]$ ($V \sim U[-\pi/2; \pi/2]$) and W exponentially distributed with parameter 1 ($W \sim Exp(1)$).

So the model that will be considered and simulated in this paper can be expressed in the following way:

$$x_{k+1} = Ax_k + \varepsilon_{k+1} \; ; \quad \varepsilon_k \sim S_\alpha(\sigma, \beta, \mu); \quad x_0 \sim S_\alpha(\sigma_2, \beta, \mu)$$
$$y_k = Cx_k + \mu_k \; ; \quad \mu_k \sim \mathcal{N}(0, R)$$

Kalman Filter

To get the Kalman Filter equations first, assuming that vector of parameters θ_k= $[\mu, \Sigma, A_k, C_k, Q_k, R_k]$ is known for all k. The aim is to estimate state variable at time k, based on available information at time k and the error of this estimation. It can be formulated the following way:

$\hat{X}_{k|k}$=$E[X_k|\mathcal{Y}_k]$;

$\Sigma_{k|k}$= $E[(X_k - \hat{X}_{k|k})(X_k - \hat{X}_{k|k})^*|\mathcal{Y}_k]$, *k=0,1,2,...,N*, where

$\mathcal{Y}_k = \sigma\{Y_0, ..., Y_k\}$ -sigma-algebra generated by Y_0,...,Y_k.,*k=0,1,2,...,N.*

Suppose that on the iteration k one has $\hat{X}_{k|k}$ and $\Sigma_{k|k}$, and it is necessary to find $\hat{X}_{k+1|k+1}$ and $\Sigma_{k+1|k+1}$. Before receiving a new observation Y_{k+1} one makes a prediction based on (1). Then when new observation received, the correction is started. G_k (Kalman Gain) is coefficient that shows the measure of uncertainly in new observation. All equations of Kalman Filter are received in assumption of Gaussian noises.

Prediction:

$\hat{X}_{k+1|k}$=$A_k\hat{X}_{k|k}$

$\Sigma_{k+1|k}$=$A_k\Sigma_{k|k}{A_k}^* + Q_{k+1}$

Innovations:

$\nu_{k+1} = Y_{k+1} - C_{k+1}\hat{X}_{k+1|k}$

$H_{k+1|k} = C_{k+1}\Sigma_{k+1|k}{C_{k+1}}^* + R_{k+1}$

Kalman Gain:

$G_{k+1} = \Sigma_{k+1|k}{C_{k+1}}^*{H_{k+1|k}}^{-1}$

Correction:

$\hat{X}_{k+1|k+1} = \hat{X}_{k+1|k} + G_{k+1}\nu_{k+1}$

$\Sigma_{k+1|k+1}$=$(I - G_{k+1}C_{k+1})\Sigma_{k+1|k}$

Given $\hat{X}_{k|k}$ and $\Sigma_{k|k}$ for all k, results can be improved by and find smoothed estimates of states: $\hat{X}_{k|N}$ and $\Sigma_{k|N}$ for k=0,1,2,...,N (all formulas can be found in [4])

3 Methods of parameters estimation

All equations above make sense only when all parameters are known. Assume that parameters are independent on time: θ= $[\mu, \Sigma, A, C, Q, R]$, and the aim is to estimate this vector based only on realisation $y_1, ..., y_k$.

To obtain the MLE [2], it is necessary to maximize the likelihood function of innovations by numerical technique (e.g. quasi-Newton-Raphson). The likelihood function of innovations:

$$L_Y(\theta) = \prod_{k=1}^{N} \frac{1}{(2\pi)^{n/2}}|H_{k|k-1}(\theta)|^{-1/2} exp(-\frac{1}{2}\nu_k(\theta)^* H^{-1}_{k|k-1}\nu_k(\theta)) \tag{7}$$

The idea to use EM-algorithm was proposed in [3]. At E-step it is necessary to find the conditional expectation of the following function:

$Q(\theta,\theta') = E_{\theta'}[log\frac{dP_\theta}{dP_{\theta'}}|\mathcal{Y}_N]$, $;log\frac{dP_\theta}{dP_{\theta'}}{=}\sum_{l=0}^{N} log\bar{\lambda}_l + constant$

where $\bar{\lambda}_k =$
$\frac{|Q|^{-1/2}\phi(Q^{-1/2}(X_k - AX_{k-1})}{\phi(X_k)} \cdot \frac{|R|^{-1/2}\psi(R^{-1/2}(Y_k - CX_k))}{\psi(Y_k)}$

and $X_k \sim \mathcal{N}(0, I_n)$; $Y_k \sim \mathcal{N}(0, I_m)$; $\psi(x)$ and $\psi(y)$ is probability density function of standard Gaussian random variable.

At the M-step, (at $j + 1$ step), one maximises $Q(\theta, \theta')$: $\theta_{j+1} = argmax_{(\theta)}\ Q(\theta,\theta')$. The update recursive equations for all parameters can be found in [4] except matrix C, so we give it here.

$C = \sum_{k=0}^{N} Y_k \hat{X}^*_{k|N} (\sum_{k=0}^{N}[\Sigma_{k|N} + \hat{X}_{k|N}\hat{X}^*_{k|N}])^{-1}$

4 Simulation results and findings

The following parameters were chosen for simulation:

$$\begin{gathered} x_{k+1} = x_k + \varepsilon_{k+1}\ ; \quad \varepsilon_k \sim S_\alpha(20,\beta,0) \\ y_k = 1.2x_k + \mu_k\ ; \quad \mu_k \sim \mathcal{N}(0,150) \\ x_0 \sim S_\alpha(50,\beta,100) \end{gathered}$$

Each sample contents 1000 observations. For all parameters α and β under consideration the mean values of Z simulations are used. To apply the Kalman Filter, we used that α - stable distribution with $\alpha = 2$ and $\beta = 0$ ($S_2(\sigma, 0, \mu)$ corresponds to Gaussian random variable $N(\mu, 2\sigma^2)$, and we simply replaced each α to $\alpha = 2$. It means that we ignore true heavy tails.

All parameters are known. Consider how α and β influence total error of prediction of unobserved states. Firstly, we studied the case of all parameters are known.

Figure 1 demonstrates the mean error of estimation that was calculated as

$$Error = \frac{1}{Z}\sum_{m=0}^{Z}\sum_{k=0}^{N}(X^{(m)} - X^{\hat{(m)}}{}_{k|k})^2 \tag{8}$$

It can be seen from the Figure 1, that in cases of large deviation from $\alpha = 2$ (refer to Gaussian distribution) the total error is rising dramatically, and by decreasing α error continues to rise, e.g. for $\alpha = 1.1$ the total error was 4000 times higher then in Gaussian case (error displayed as a constant on the first picture because of great mistake for small α, so "zoomed" graphs are provided). Moreover, the Kalman Smoother gives more accurate state estimates only for α in near [1.85;2]. Note that in the interval [1.85;2] the KF and KS are not so sensitive to deviation in α (for α=1.85, the total error increases by 20%) - then the slopes of both curves only rise.

The only parameter that was not set to true values is **distribution of state equation noise** and initial state vector. So it is logical to assume that such large total error of predictions is strongly connected with the non-observance of Gaussian assumption of state equation noise.To

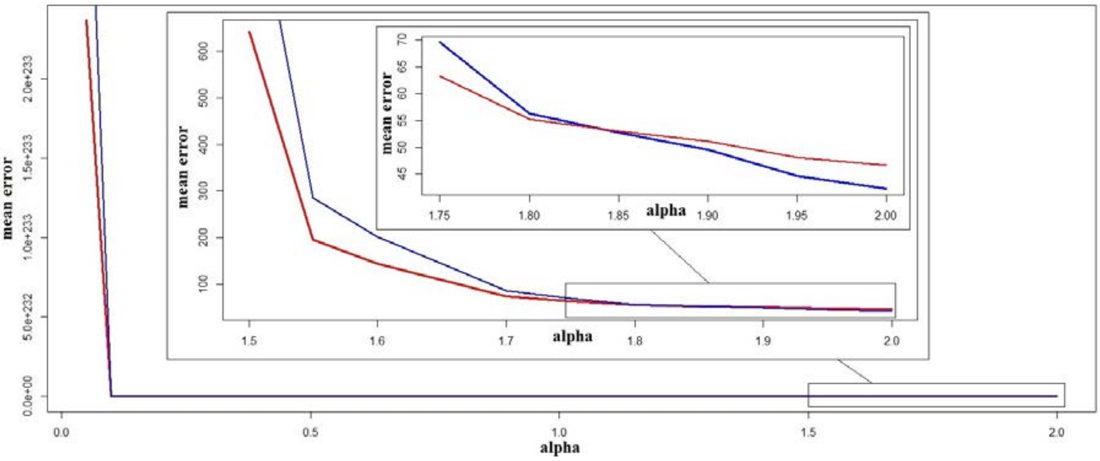

Figure 1: Mean error (10000 simulation) of KF (red) and KS (blue) estimation. $\alpha \in [0.05; 2]$, step=0.05, $\beta = 0$

be more precise, we conjecture that it can be explained by **underestimation of the state equation noises covariance matrix** (matrix Q $= 2\sigma^2$ in our model set up).

It is enough to compare α - stable distribution and Gaussian distribution with the same covariance matrix, to understand it[17]:

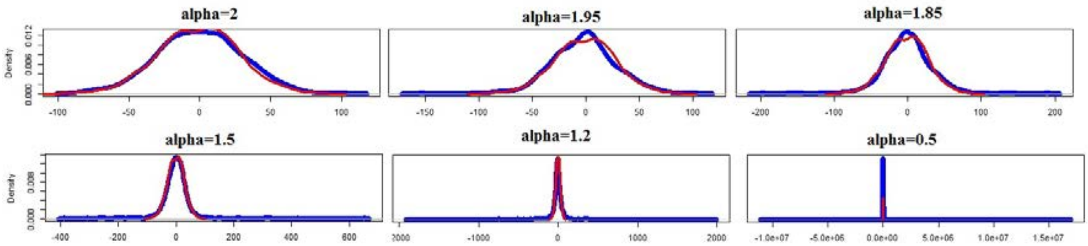

Figure 2: Kernel densities of α - stable distribution(blue)($\sigma = 20; \beta = 0$) and Gaussian(red) distribution with same covariance matrix.

The consequences of the assumption of Gaussian noise when it truly alpha-stable is shown on Figure 2. To obtain KF estimates we substitute true distribution (that corresponds to blue line) by Gaussian distribution with the same covariance matrix (that corresponds to red lines). However, it is seen that long tails of blue graphs are not covered by red ones, so as a result **algorithm can not cannot "detect" jumps** in a process that are appearing because of heavy tail distribution of disturbances. It is obvious that with α closer to zero, the tails become longer and the underestimation becomes larger and we met larger total error (when α is close to 2, the difference is not so dramatic). In terms of Kalman filter, algorithm consider state equation as not so noisy how it is truly is, and it assign small weight (Kalman Gain) to received observation.

The same mean total error as before for different β and same α is plotted in Figure 3.

[17]Gaussian variables were simulated by Box-Muller, and α-stable were as in [5]

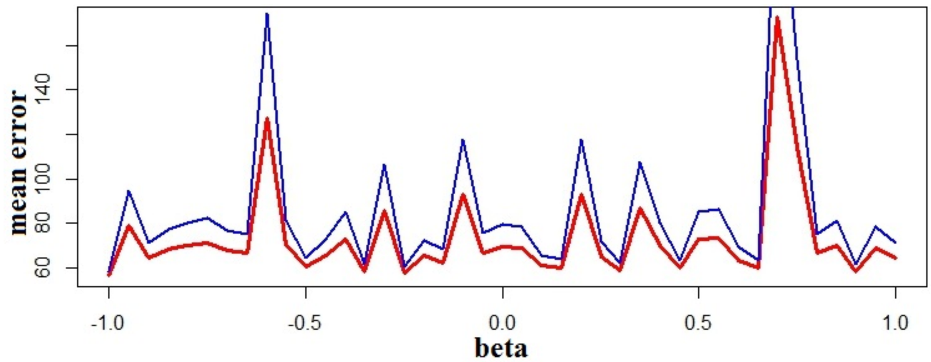

Figure 3: Mean error (10000 simulations) of KF (red) and KS (blue) estimation. $\beta \in [-1; 1]$ step=0.05 $\alpha = 1.75$

There is probably no dependence between total error and parameter β. The case of symmetric tails ($\beta = 0$) do not refer to the least total error. Moreover, the fluctuations of the total error are not so large. The deviation in heavy tails (α) is more significant for state estimates than in β, so we switch to different α only in estimation methods.

Parameters estimation. We consider the behaviour of two estimation procedures. Initial values were set to true ones, and the following results were found.[18]

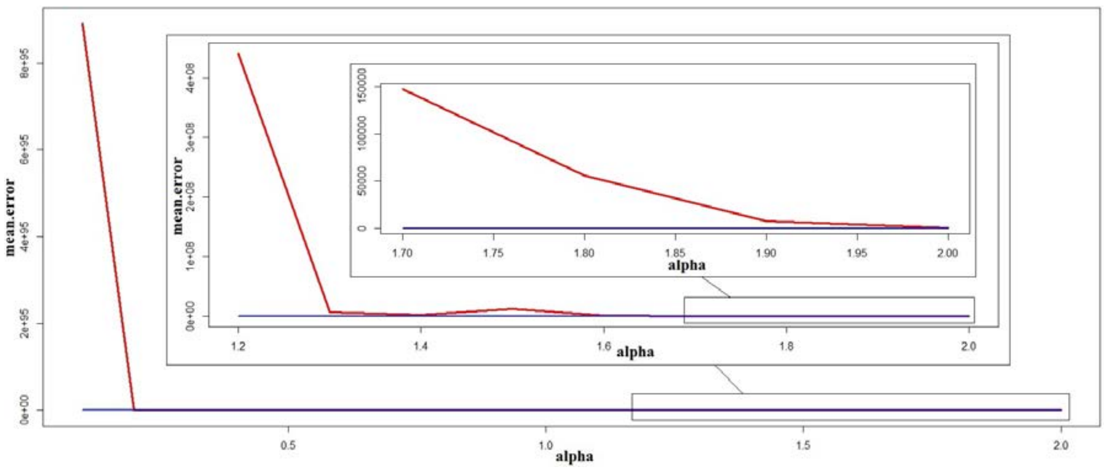

Figure 4: Mean error (1000 simulations). Parameters estimation by MLE (red) and by EM (blue).$\alpha \in [0.1; 2]$, step=0.1 $\beta = 0$

Figure 4 shows that the total error of estimation by MLE is increasing sharp, and as expected MLE is *extremely sensitive to deviation* from Gaussian distribution. Because of different ranges of scale, we plot the error of EM estimation separately (Figure 5).

[18]Because of numerical optimization of likelihood function in ML estimation, the number of simulations was reduced.

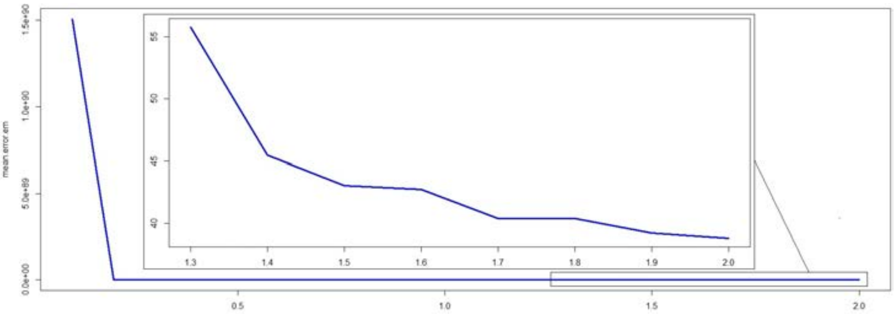

Figure 5: Mean error (1000 simulations). EM (blue) parameters estimation. $\alpha \in [0.1; 2]$, step=0.1 $\beta = 0$

Figure 5 shows that total error of KF prediction when parameters are estimated by EM increases slowly in the interval of [1.3;2],e.g. for $\alpha = 1.4$ error increases only by 14% (in average), against near 622% when "all parameters are known" and it is more satisfying result than before. But out of this interval the total error is large but less than in case of MLE estimation. To understand the nature of such good behaviour of EM we turn to figures of parameters estimation (Figure 6).

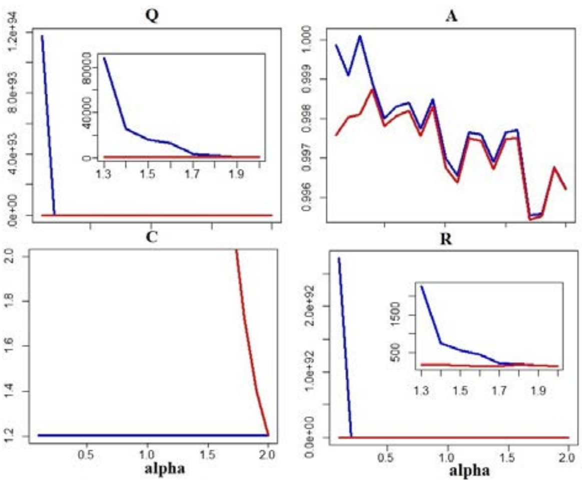

Figure 6: Mean (1000 simulation) MLE (red) and EM (blue) parameters estimation.$\alpha \in [0.1; 2]$, step=0.1 $\beta = 0$

It is obvious that results of *EM-estimation are more accurate* that MLE ones. The parameters of A and C are estimated correctly by EM-algorithm. But *EM-algorithm overestimate* the covariance matrix Q(compare to real value). **However this over estimation allows to take into account heavy tails**. The *overestimation of matrix Q* tends to larger Kalman Gain. It means that in a moment of correction we **assign larger weight to observation** we received than to our prediction (because of large covariance matrix). It is important to mention, that received results were confirmed when EM estimation was used only for Q, and all other parameters were true, so the cause of little increase in prediction error is overestimation of Q. Of course, we increase the confidence interval of our state prediction, but it seems to be quite acceptable cost for considerable error decreasing. In our simulation example we decrease (e.g. for $\alpha = 1.3$ the total error 140 times while $Sigma_{k|k}$ increase only by 1.28). Unfortunately, overestimation of Q is not enough in cases of larger deviation in α. But in cases when it is necessary to estimate parameters and $\alpha \in [1.3; 2]$ it is acceptable not to complicate KF, and use only EM.

5 Conclusion

By simulation it was shown that in spite of the heavy tails of state equation noise, EM and ML estimate parameter A properly. EM-algorithm can **overestimate** covariance matrix of state noise in such way that the total error of prediction increases a little (compared to the Gaussian case) in the interval $[1.3; 2]$. It was confirmed that exactly due to the overestimation of covariance matrix Q, it is possible to prevent large prediction error. Without any additional computations, only in process of parameters estimation by EM and if $\alpha \in [1.3; 2]$, we can get approximately true values of unobserved states. ML estimation does not demonstrate such good properties, and more likely gives wrong estimates of states. It is a possible evidence of unacceptable application MLE for KF estimation, because it can lead to incorrect results. Although it is evidence of useful properties of EM that can allow to apply standard KF procedure in cases of α-stable distribution. The detecting how proposed method can reduce the prediction error comparing to modification of KF in [7], and how to estimate all parameters in this modification are questions for our further research.

Bibliography

[1] Bucy, R.S., Kalman, R.E. (1961). *New Results in Linear Filtering and Prediction Theory.* Trans. ASME J. of Basic Engineering. 83, 95-108

[2] Gupta, N.K., R.K.Mehra. (1974). *Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculation.* IEEE Trans. Aut.Cont. AC-19 774-783

[3] Shumway , R.H., Stoffer, D.S. (1982). *An approach to time series smoothing and forecasting using the EM algorithm.* Journal of Time Series Analysis 3(4): 253-264.

[4] Shumway, R.H. and Stoffer, D.S. (2006). *Time Series Analysis and Its Applications (with R Examples).* Springer.

[5] Weron, R.(1996) . *On the Chambers-Mallows-Stuck method for simulating skewed stable random variables.*, Statist. Probab. Lett. 28, 165-171

[6] Zolotarev, V. (1986) *One dimensional Stable Distributions.* American Mathematical Society, Providence, RI. Russian original, 1983

[7] Xu Sun, Jinqiao Duan, Xiaofan Li, Xiangjun Wang (2013). *State estimation under non-Gaussian LÂ ´evy noise: A modified Kalman filtering method.* arXiv:1303.2395.v1

Acknowledgement. I would like to thank Prof. Valentin Konakov for the statement of problem and for stimulating discussions.