Approximate Recursive Identification of Autoregressive Systems with Skewed Innovations

Henri Nurminen
Dept. of Automation Science and Engineering
Tampere University of Technology
Tampere, Finland
henri.nurminen@tut.fi

Tohid Ardeshiri
Department of Engineering
University of Cambridge
Cambridge, UK
ta417@cam.ac.uk

Abstract

We propose a novel recursive system identification algorithm for linear autoregressive systems with skewed innovations. The algorithm is based on the variational Bayes approximation of the model with a multivariate normal prior for the model coefficients, multivariate skew-normally distributed innovations, and matrix-variate-normal–inverse-Wishart prior for the parameters of the innovation distribution. The proposed algorithm simultaneously estimates the model coefficients as well as the parameters of the innovation distribution, which are both allowed to be slowly time-varying. Through computer simulations, we compare the proposed method with a variational algorithm based on the normally-distributed innovations model, and show that modelling the skewness can provide improvement in identification accuracy.

1 Introduction

Many systems produce datasets with skewed noise distribution. Skewness means asymmetry. Positive skewness, for example, intuitively means producing large positive deviations from the median value more frequently than large negative deviations. For instance, some financial data sets show negative skewness because large drops tend to be more frequent than large upsurges [1, 2, 3, 4]. Wireless network based positioning often uses time delay measurement as a distance, but non-line-of-sight can produce large positive outliers, so the error distribution becomes positively skewed [5, 6].

One statistical model for skewed error distributions is the skew normal distribution and its multivariate generalisation [7, 8]. The posterior distribution of a normal prior and skew-normal measurement noise model is not analytically tractable. However, the distribution admits a hierarchical formulation whose favorable conjugacy properties enable efficient parameter estimation using the expectation–maximisation (EM) algorithm [9, 10, 11] and approximate Bayesian time-series filtering and smoothing based on the variational Bayes (VB) approximation [12, 13].

This paper studies autoregressive (AR) models, where the measurement is modelled to be a linear function of n_{AR} (the model order) previous measurements plus an independent random noise term referred to as the innovation. When the AR coefficients and/or the conditionally skew-normal innovation distribution's statistics are time-varying or they need to be identified online, recursive identification methods are used [14].

In this paper we propose a novel recursive system identification algorithm for AR models with skew-normally distributed measurement noise with unknown possibly slowly time-varying scale and skewness. The proposed approximation is based on a VB approximation.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

2 Problem formulation

Skew normal distribution is an asymmetric generalization of the normal distribution originally proposed by Azzalini [7]. Its multivariate version was later introduced by Azzalini and Dalla Valle [8]. The version that is used in this report is the canonical fundamental skew normal distribution (CFUSN) introduced by Arellano Valle and Genton [15]. However, we adopt a different parametrization following the guidelines of the canonical fundamental skew t-distribution's parametrization in [16] to obtain a suitable analytical tractability. The probability density function (PDF) of this skew normal distribution $z \sim \mathrm{SN}(\mu, R, \Delta)$ is

$$p(z) = 2^{n_z}\, \mathrm{N}(z; \mu - \Delta\sqrt{\tfrac{2}{\pi}}\mathbf{1}, \Omega)\, F_{\mathrm{N}}(\Delta^{\mathrm{T}}\Omega^{-1}(z - \mu + \Delta\sqrt{\tfrac{2}{\pi}}\mathbf{1}); 0, I - \Delta^{\mathrm{T}}\Omega^{-1}\,\Delta), \tag{1}$$

where $\mathbf{1}$ is a vector of ones, μ is a location parameter, $\Omega = R + \Delta\Delta^{\mathrm{T}}$, and F_{N} is the cumulative distribution function of the multivariate normal distribution. $R \in \mathbb{R}^{n_z \times n_z}$ (symmetric positive-definite, spd) and $\Delta \in \mathbb{R}^{n_z \times n_z}$ are shape matrices that determine the scale and skewness, and the sign and structure of Δ determine the direction of skewness as explained in [16]. Examples of the PDF in negatively skewed, symmetric, and positively skewed cases are given in Fig. 1. The moments of this multivariate skew normal distribution given the shape matrices are

$$\mathbb{E}[z] = \mu, \quad \mathbb{V}[z] = R + \tfrac{2}{\pi}\Delta\Delta^{\mathrm{T}}. \tag{2}$$

Compared to the formulation of [16], we shift the distribution with $\Delta\sqrt{2/\pi}\mathbf{1}$ so that the mean of the distribution does not depend on Δ. This ensures that the proposed algorithm identifies Δ as a measure of skewness, not as a measure of location.

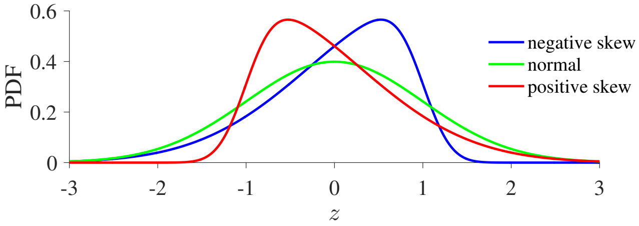

Figure 1: The PDFs of negatively-skewed, symmetric, and positively-skewed normal distributions. Each distribution has mean zero and variance one.

We formulate the AR coefficient estimation problem as the linear state-space model with the measurement noise being skew-normally distributed conditional on the unknown slowly-varying noise parameters R_k and Δ_k

$$p(x_1) = \mathrm{N}(x_1; x_{1|0}, P_{1|0}) \tag{3a}$$

$$x_k = x_{k-1} + w_{k-1}, \quad w_{k-1} \overset{iid}{\sim} \mathrm{N}(0, Q_{k-1}) \tag{3b}$$

$$z_k = C_k x_k + e_k, \quad e_k \overset{iid}{\sim} \mathrm{SN}(\mu, R_k, \Delta_k), \tag{3c}$$

where $x_k \in \mathbb{R}^{n_{\mathrm{AR}}}$ is the vector of AR coefficients, $Q_k \in \mathbb{R}^{n_{\mathrm{AR}} \times n_{\mathrm{AR}}}$ (spd) is the process noise covariance matrix that is assumed known and is thus an algorithm parameter, $z_k \in \mathbb{R}^{n_z}$ is the measurement, $C_k \in \mathbb{R}^{n_z \times n_{\mathrm{AR}}} = \begin{bmatrix} z_{k-1} & z_{k-2} & \cdots & z_{k-n_{\mathrm{AR}}} \end{bmatrix}$ is the measurement model matrix given by n_{AR} previous measurements, and $\{w_k \in \mathbb{R}^{n_{\mathrm{AR}}}\}_{k=1}^{K}$ and $\{e_k \in \mathbb{R}^{n_z}\}_{k=1}^{K}$ are mutually independent process and measurement noise sequences.

3 Proposed algorithm

3.1 Measurement update

Conditional on the parameters R_k and Δ_k, the skew-normal random variable $e_k | R_k, \Delta_k \sim \mathrm{SN}(\mu, R_k, \Delta_k)$ has the hierarchical formulation [9]

$$e_k | u_k, R_k, \Delta_k \sim \mathrm{N}(\mu + \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}), R_k) \tag{4a}$$

$$u_k \sim \mathrm{N}_+(0, I), \tag{4b}$$

where N_+ is the multivariate normal distribution truncated into positive orthant. To obtain the necessary conjugacy properties, let us assign the matrix-variate-normal–inverse-Wishart (MVNIW) prior distribution to the joint random variable (R_k, Δ_k):

$$p(R_k, \Delta_k) = \mathrm{N}(\Delta_k; \Delta_{k|k-1}, R_k \otimes V_{k|k-1})\,\mathrm{IW}(R_k; \Psi_{k|k-1}, \nu_{k|k-1}), \tag{5}$$

where $\Delta_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$, $V_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$ (spd), $\Psi_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$ (spd), and $\nu_{k|k-1} \in (2n_z, \infty)$ are parameters of the prior distribution. $\mathrm{N}(X; M, U \otimes V)$ is the PDF of the matrix-variate normal distribution with mean M, and variance parameters U (among-row) and V (among-column) [17, Ch. 2], and $\mathrm{IW}(X; \Psi, \nu)$ is PDF of the inverse-Wishart distribution with scale-matrix Ψ and ν degrees of freedom [17, Ch. 3].

The filtering posterior distribution $p(x_k, u_k, R_k, \Delta_k | z_{1:k})$ of the model defined by (3) and (5) is not analytically tractable. Our solution is to use a variational Bayesian approximation, where we find the functions $q_{x,u}(x_k, u_k)$ and $q_{R,\Delta}(R_k, \Delta_k)$ such that the reversed Kullback–Leibler divergence (KLD)

$$\mathrm{D}_{\mathrm{KL}}\big(q_{x,u}(x_k, u_k)\, q_{R,\Delta}(R_k, \Delta_k) \| p(x_k, u_k, R_k, \Delta_k | z_{1:k})\big) \tag{6}$$

is minimised, where $\mathrm{D}_{\mathrm{KL}}(q\|p) = \int q(x) \log(\frac{q(x)}{p(x)})\, \mathrm{d}x$. In general there is no exact analytical solution for $(q_{x,u}, q_{R,\Delta})$, but the iteration of

$$\log q_{x,u}(x_k, u_k) \leftarrow \mathbb{E}_{q_{R,\Delta}}[\log p(z_k, x_k, u_k, R_k, \Delta_k | z_{1:k-1})] + c_{x,u} \tag{7a}$$

$$\log q_{R,\Delta}(R_k, \Delta_k) \leftarrow \mathbb{E}_{q_{x,u}}[\log p(z_k, x_k, u_k, R_k, \Delta_k | z_{1:k-1})] + c_{R,\Delta} \tag{7b}$$

always reduces the KLD (6) and for many models gives a sequence that converges towards the optimal functions $(q_{x,u}, q_{R,\Delta})$ [18, Chapter 10][19]. The expected values on the right hand sides of (7) are taken with respect to the current $q_{x,u}$ and $q_{R,\Delta}$, and $c_{x,u}$ and $c_{R,\Delta}$ are constants with respect to the variables (x_k, u_k), and (R, Δ), respectively.

Thanks to the chosen prior distribution structure (5), the update (7b) has a closed form solution that preserves the functional form of the prior, and the moments of the distribution required by other computations are also analytically tractable. The analytical solution of the update (7a) is a multivariate normal distribution truncated by multiple linear constraints. The mean and covariance matrix of this distribution can be approximated using the sequential truncation algorithm [20, 21, 13]. The distribution $q_{x,u}(x_k, u_k)$ is then approximated by the unconstrained multivariate normal distribution with the obtained moments

$$q_{x,u}(x_k, u_k) \approx \mathrm{N}\left(\left[\begin{smallmatrix} x_k \\ u_k \end{smallmatrix}\right]; \xi_{k|k}, \Xi_{k|k}\right), \tag{8}$$

where $\xi_{k|k}$ and $\Xi_{k|k}$ are the mean and covariance matrix given by the sequential truncation algorithm. Normal marginal posterior approximation for x_k guarantees that we get a recursive algorithm. The approximative filtering posterior of (R_k, Δ_k) is the MVNIW distribution

$$q_{R,\Delta}(R_k, \Delta_k) = \mathrm{N}(\Delta_k; \Delta_{k|k}, R_k \otimes V_{k|k})\,\mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k}), \tag{9}$$

whose required moments are analytically tractable when $\nu_{k|k} > 2n_z$ as shown in Appendix A.

3.2 Time update

The marginal distribution of the AR coefficient vector x_k in the posterior approximation $\mathrm{N}\left(\left[\begin{smallmatrix} x_k \\ u_k \end{smallmatrix}\right]; \xi_{k|k}, \Xi_{k|k}\right) \cdot q_{R,\Delta}(R_k, \Delta_k)$ is a normal distribution and the state transition (3b) is linear and Gaussian. Thus, the filter prediction becomes the standard Kalman filter prediction and the prediction distribution is normal.

The dynamical model of the model parameters $p(R_k, \Delta_k | R_{k-1}, \Delta_{k-1})$ is typically unknown and/or intractable. Therefore, we adopt the forgetting factor update, which provides the maximum-entropy solution given the KLD from the previous posterior [22, 23]. Thus, the used prediction density given the MVNIW approximation of the previous posterior and the forgetting factor $\gamma \in (0, 1]$ is

$$\begin{aligned} \hat{p}(R_k, \Delta_k | y_{1:k-1}) \propto\; & \mathrm{N}(\Delta_k; \Delta_{k-1|k-1}, R_k \otimes \tfrac{1}{\gamma} V_{k-1|k-1}) \\ & \times \mathrm{IW}(R_k; \gamma \Psi_{k-1|k-1}, \gamma\, \nu_{k-1|k-1} + (1-\gamma) \cdot 2n_z). \end{aligned} \tag{10}$$

where the term $(1-\gamma) \cdot 2n_z$ guarantees that the resulting inverse-Wishart distribution is well-defined and has an expectation value.

The details of the proposed recursive identification algorithm including the prediction equations implied by the time update (10) are given in Appendix B.

4 Simulated example

We simulated 1000 Monte Carlo replications of the AR model with 25 AR coefficients with $n_z = 2$ dimensional skew-normally distributed innovations with parameters $R = 0.1^2 \cdot I$ and $\Delta = [\begin{smallmatrix} 2 & 0 \\ 1 & 2 \end{smallmatrix}]$. Thus, the true distribution has high positive skewness. The true coefficients were simulated by generating the zeros of the characteristic polynomial from the uniform distribution $\mathrm{unif}(-1, 1)$. The number of AR coefficients was assumed known. The initial prior covariance matrix for the AR coefficient vector was given by the 1st order stable spline kernel $[P_{1|0}]_{i,j} = \frac{30-1}{3} 0.5^{\max(i-1,j-1)}$, and the process noise covariance was chosen as $[Q_{k-1}]_{i,j} = (\frac{1}{\gamma} - 1) \max(\mathrm{diag}(P_{k-1|k-1})) \cdot 0.5^{\max(i-1,j-1)}$ to preserve the stable kernel form of the prior [24, 25].

The proposed method is compared with the Gaussian variational Bayes filter for slowly-drifting noise proposed by Agamennoni et al in [26]. The skew-normal based identification method was given the positive direction of the skewness by using the initial prior

$$p(R_1, \Delta_1) = \mathrm{N}(\Delta_1; \sqrt{\tfrac{\pi}{2} \tfrac{1}{2}} I, R_1 \otimes I) \, \mathrm{IW}(R_1; \tfrac{\nu_{1|0} - 3}{2} I, \nu_{1|0}), \tag{11}$$

where $\nu_{1|0} = 4 + 10^{-10}$. That is, the variance is divided equally between the symmetric and skewed component in the sense that $\mathbb{E}[R_1^{-1}]^{-1} = \mathbb{E}[\Delta_1]^2 \frac{2}{\pi} = \frac{1}{2} I$. The normal distribution based method was given the initial prior

$$p(R_1) = \mathrm{IW}(R_1; (\nu_{1|0} - 3) I, \nu_{1|0}). \tag{12}$$

The forgetting factor value used with both the methods was $\gamma = 0.975$, and the number of VB iterations was 10. Fig. 2 shows the relative difference of the identification error

$$\epsilon_k = \sqrt{\textstyle\sum_{i=1}^{n_{\text{AR}}} (x_{k|k} - (x_k)_{\text{true}})^2} \tag{13}$$

as a function of the fed number of measurements. The figure shows that the skew-normal based identification method gives a lower median error than the normal distribution based, and the relative differences increase as the number of measurements increase. Fig. 2 shows that after 10.000 measurements, the skew-normal based method is more accurate in about 95 % of the cases and gives at least 25 % lower identification error in most of the simulations.

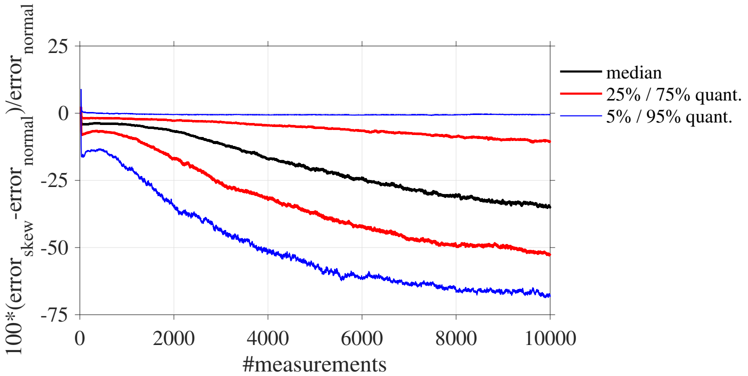

Figure 2: The proposed algorithm is more accurate than the normal distribution based algorithm in 95 % of the simulations, and in most of the simulations the error (13) is reduced by more than 25 %.

5 Conclusions

We proposed a novel recursive estimation algorithm for identifying the model coefficients and innovation distribution parameters of autoregressive models with skew-normally distributed innovations. Both model coefficients and innovation distribution parameters can be slowly time-varying. Our computer simulation showed that modelling skewness can improve the accuracy of identification.

Acknowledgments

H. Nurminen receives funding from Tampere University of Technology Graduate School, Nokia Technologies Oy, the Foundation of Nokia Corporation, and Tekniikan edistämissäätiö.

T. Ardeshiri receives funding from Swedish research council's (VR) project Scalable Kalman filters, and from Jaguar Land Rover (JLR), Whitley, UK.

References

[1] C. R. Harvey and A. Siddique, "Autoregressive conditional skewness," *The Journal of Financial and Quantitative Analysis*, vol. 34, pp. 465–487, December 1999.

[2] E. Jondeau and M. Rockinger, "Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements," *Journal of Economic Dynamics and Control*, vol. 27, pp. 1699–1737, 2003.

[3] P. Christofferssen, S. Heston, and K. Jacobs, "Option valuation with conditional skewness," *Journal of Econometrics*, vol. 131, pp. 253–284, 2006.

[4] G. Tsiotas, "On generalised asymmetric stochastic volatility models," *Computational Statistics and Data Analysis*, vol. 56, pp. 151–172, 2012.

[5] K. Kaemarungsi and P. Krishnamurthy, "Analysis of WLAN's received signal strength indication for indoor location fingerprinting," *Pervasive and Mobile Computing*, vol. 8, no. 2, pp. 292–316, 2012. Special Issue: Wide-Scale Vehicular Sensor Networks and Mobile Sensing.

[6] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "A NLOS-robust TOA positioning filter based on a skew-t measurement noise model," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–7, October 2015.

[7] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, no. 2, pp. 171–178, 1985.

[8] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[9] T. I. Lin, "Maximum likelihood estimation for multivariate skew normal mixture models," *Journal of Multivariate Analysis*, vol. 100, pp. 257–265, 2009.

[10] S. Lee and G. J. McLachlan, "Finite mixtures of multivariate skew t-distributions: some recent and new results," *Statistics and Computing*, vol. 24, no. 2, pp. 181–202, 2014.

[11] H. J. Ho, S. Pyne, and T. I. Lin, "Maximum likelihood inference for mixtures of skew student-t-normal distributions through practical EM-type algorithms," *Statistics and Computing*, vol. 22, pp. 287–299, 2012.

[12] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Processing Letters*, vol. 22, pp. 1898–1902, November 2015.

[13] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Skew-t filter and smoother with improved covariance matrix approximation." Available online at `http://arxiv.org/abs/1608.07435`, August 2016.

[14] L. Ljung, "Recursive identification algorithms," *Circuits, Systems, and Signal Processing*, vol. 21, no. 1, pp. 57–68, 2002.

[15] R. B. Arellano-Valle and M. G. Genton, "On fundamental skew distributions," *Journal of Multivariate Analysis*, no. 96, pp. 93–116, 2005.

[16] S. X. Lee and G. J. McLachlan, "Finite mixtures of canonical fundamental skew t-distributions – the unification of the restricted and unrestricted skew t-mixture models," *Statistics and Computing*, no. 26, pp. 573–589, 2016.

[17] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. Boca Raton, FL: Chapman & Hall/CRC, 2000.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[19] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, pp. 131–146, Nov. 2008.

[20] T. Perälä and S. Ali-Löytty, “Kalman-type positioning filters with floor plan information,” in *6th International Conference on Advances in Mobile Computing and Multimedia (MoMM)*, pp. 350–355, 2008.

[21] D. J. Simon and D. L. Simon, “Constrained Kalman filtering via density function truncation for turbofan engine health estimation,” *International Journal of Systems Science*, vol. 41, no. 2, pp. 159–171, 2010.

[22] M. Kárný and K. Dedecius, “Approximate Bayesian recursive estimation: On approximation errors,” tech. rep., ÚTIA AV ČR, January 2012.

[23] E. Özkan, V. Šmídl, S. Saha, C. Lundquist, and F. Gustafsson, “Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters,” *Automatica*, vol. 49, 2013.

[24] G. Pillonetto and G. De Nicolao, “A new kernel-based approach for linear system identification,” *Automatica*, vol. 46, pp. 81–93, 2010.

[25] T. Chen, H. Ohlsson, and L. Ljung, “On the estimation of transfer functions, regularizations and Gaussian processes–revisited,” *Automatica*, vol. 48, pp. 1525–1535, 2012.

[26] G. Agamennoni, J. Nieto, and E. Nebot, “Approximate inference in state-space models with heavy-tailed noise,” *IEEE Transactions on Signal Processing*, vol. 60, pp. 5024–5037, October 2012.

Appendices

A Variational solution of the measurement update

Our variational solution uses this hierarchical formulation of the measurement noise model:

$$\begin{aligned} z_k|x_k, u_k, R_k, \Delta_k &\sim \mathrm{N}(C_k x_k + \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}), R_k), && (14a)\\ u_k &\sim \mathrm{N}_+(0, I), && (14b)\\ \Delta_k|R_k &\sim \mathrm{N}(\Delta_{k|k-1}, R_k \otimes V_{k|k-1}), && (14c)\\ R_k &\sim \mathrm{IW}(\Psi_{k|k-1}, \nu_{k|k-1}), && (14d) \end{aligned}$$

where $z_k \in \mathbb{R}^{n_z}$ is the measurement, $u_k \in \mathbb{R}^{n_z}$ is the skewness variable vector, and $\Delta_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$, $V_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$ (spd), $\Psi_{k|k-1} \in \mathbb{R}^{n_z \times n_z}$ (spd), and $\nu_{k|k-1} > 2n_z$ are the parameters of the joint prior distribution of Δ_k and R_k. The prior of Δ_k and R_k is implied by the previous filtering posterior and the time update step (filter prediction) that is explained in section 3.2.

The derivations for the variational solution (7) are given in Sections A.1 and A.2. For brevity all constant values are denoted by c in the derivation. The logarithm of the full filtering distribution which is needed for the derivations is

$$\begin{aligned} &\log p(z_k, x_k, u_k, R_k, \Delta_k|z_{1:k-1})\\ =&-\frac{1}{2}(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))^{\mathrm{T}} R_k^{-1}(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))\\ &-\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) - \frac{1}{2}u_k^{\mathrm{T}} u_k\\ &-\frac{1}{2}\mathrm{Tr}\{(\Delta_k - \Delta_{k|k-1})V_{k|k-1}^{-1}(\Delta_k - \Delta_{k|k-1})^{\mathrm{T}} R_k^{-1}\}\\ &-\frac{\nu_{k|k-1}+1}{2}\log\det(R_k) - \frac{1}{2}\mathrm{Tr}\{\Psi_{k|k-1}R_k^{-1}\} + c, \quad u \geq 0, \qquad (15) \end{aligned}$$

where $x_{k|k-1}$ and $P_{k|k-1}$ are the mean and covariance matrix of the current predictive distribution, and $\mathrm{Tr}\{\cdot\}$ is the matrix trace.

A.1 Derivations for $q_{x,u}$

Using equation (7a) we obtain

$$\begin{aligned} &\log q_{x,u}(x_k, u_k)\\ =&-\frac{1}{2}\mathbb{E}_{q_{R,\Delta}}\left[(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))^{\mathrm{T}} R_k^{-1}(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))\right]\\ &-\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) - \frac{1}{2}u_k^{\mathrm{T}} u_k + c \qquad (16)\\ =&-\frac{1}{2}(z_k - C_k x_k - \Delta_{k|k}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))^{\mathrm{T}} R_{k|k}^{-1}(z_k - C_k x_k - \Delta_{k|k}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))\\ &-\frac{1}{2}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}}(\mathbb{E}_{q_{R,\Delta}}[\Delta_k^{\mathrm{T}} R_k^{-1}\Delta_k] - \Delta_{k|k}^{\mathrm{T}} R_{k|k}^{-1}\Delta_{k|k})(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1})\\ &-\frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) - \frac{1}{2}u_k^{\mathrm{T}} u_k + c \quad u_k \geq 0, \qquad (17) \end{aligned}$$

where $(R_{k|k}, \Delta_{k|k}) \triangleq (\mathbb{E}_{q_{R,\Delta}}[R_k^{-1}]^{-1}, \mathbb{E}_{q_{R,\Delta}}[\Delta_k])$ as well as the identity $\mathbb{E}_{q_{R,\Delta}}[R_k^{-1}\Delta_k] = R_{k|k}^{-1}\Delta_{k|k}$ are derived in Section A.2. The inequality $u_k \geq 0$ denotes that each element of the vector u_k is required to be greater or equal than zero. Further, in Section A.2 it is proved that

$\mathbb{E}_{q_{R,\Delta}}[\Delta_k^{\mathrm{T}} R_k^{-1} \Delta_k] = n_z V_{k|k} + \Delta_{k|k}^{\mathrm{T}} R_{k|k}^{-1} \Delta_{k|k}$, so Eq. (17) becomes

$$\begin{aligned}
&\log q_{x,u}(x_k, u_k)\\
&= -\frac{1}{2}(z_k - C_k x_k - \Delta_{k|k}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))^{\mathrm{T}} R_{k|k}^{-1}(z_k - C_k x_k - \Delta_{k|k}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))\\
&\quad - \frac{n_z}{2}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} V_{k|k}(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1})\\
&\quad - \frac{1}{2}(x_k - x_{k|k-1})^{\mathrm{T}} P_{k|k-1}^{-1}(x_k - x_{k|k-1}) - \frac{1}{2}u_k^{\mathrm{T}} u_k + c \qquad (18)\\
&= -\frac{1}{2}(z_k + \Delta_{k|k}\sqrt{\tfrac{2}{\pi}}\mathbf{1} - \begin{bmatrix} C_k & \Delta_{k|k}\end{bmatrix}\begin{bmatrix} x_k \\ u_k\end{bmatrix})^{\mathrm{T}} R_{k|k}^{-1}(z_k + \Delta_{k|k}\sqrt{\tfrac{2}{\pi}}\mathbf{1} - \begin{bmatrix} C_k & \Delta_{k|k}\end{bmatrix}\begin{bmatrix} x_k \\ u_k\end{bmatrix})\\
&\quad - \frac{1}{2}(\begin{bmatrix} x_k \\ u_k\end{bmatrix} - \xi_{k|k-1})^{\mathrm{T}} \Xi_{k|k-1}^{-1}(\begin{bmatrix} x_k \\ u_k\end{bmatrix} - \xi_{k|k-1}) + c, \quad u_k \geq 0, \qquad (19)
\end{aligned}$$

where

$$\Xi_{k|k-1} = \begin{bmatrix} P_{k|k-1} & \mathrm{O} \\ \mathrm{O} & (I + n_z V_{k|k})^{-1}\end{bmatrix}, \qquad (20)$$

$$\xi_{k|k-1} = \begin{bmatrix} x_{k|k-1} \\ n_z\sqrt{\tfrac{2}{\pi}}(I + n_z V_{k|k})^{-1} V_{k|k}\mathbf{1}\end{bmatrix}. \qquad (21)$$

Hence,

$$\begin{aligned}
q_{x,u}(x_k, u_k) &\propto \mathrm{N}(z_k + \Delta_{k|k}\sqrt{\tfrac{2}{\pi}}\mathbf{1}; \begin{bmatrix} C_k & \Delta_{k|k}\end{bmatrix}\begin{bmatrix} x_k \\ u_k\end{bmatrix}, R_{k|k})\,\mathrm{N}(\begin{bmatrix} x_k \\ u_k\end{bmatrix}; \xi_{k|k-1}, \Xi_{k|k-1}) \cdot [\![u_k \geq 0]\!] \qquad (22)\\
&\propto \mathrm{N}(\begin{bmatrix} x_k \\ u_k\end{bmatrix}; \widehat{\xi}_{k|k}, \widehat{\Xi}_{k|k}) \cdot [\![u_k \geq 0]\!], \qquad (23)
\end{aligned}$$

where $[\![\cdot]\!]$ is the Iverson bracket, and $\widehat{\xi}_{k|k}$ and $\widehat{\Xi}_{k|k}$ are the outputs of the Kalman filter update

$$\begin{aligned}
\widetilde{C}_k &= \begin{bmatrix} C_k & \Delta_{k|k}\end{bmatrix}, \qquad (24)\\
K_k &= \Xi_{k|k-1}\widetilde{C}_k^{\mathrm{T}}(\widetilde{C}_k \Xi_{k|k-1}\widetilde{C}_k^{\mathrm{T}} + R_{k|k})^{-1}, \qquad (25)\\
\widehat{\xi}_{k|k} &= \xi_{k|k-1} + K_k(z_k + \Delta_{k|k}\sqrt{\tfrac{2}{\pi}}\mathbf{1} - \widetilde{C}_k \xi_{k|k-1}), \qquad (26)\\
\widehat{\Xi}_{k|k} &= (I - K_k\widetilde{C}_k)\Xi_{k|k-1}. \qquad (27)
\end{aligned}$$

To make the algorithm recursive, we approximate $q_{x,u}$ with a multivariate normal distribution

$$\begin{aligned}
q_{x,u}(x_k, u_k) &= \mathrm{N}(\begin{bmatrix} x_k \\ u_k\end{bmatrix}; \widehat{\xi}_{k|k}, \widehat{\Xi}_{k|k}) \cdot [\![u_k \geq 0]\!] \qquad (28)\\
&\approx \mathrm{N}(\begin{bmatrix} x_k \\ u_k\end{bmatrix}; \xi_{k|k}, \Xi_{k|k}), \qquad (29)
\end{aligned}$$

whose approximate mean and covariance matrix $\xi_{k|k}$ and $\Xi_{k|k}$ are obtained through approximate moment-matching. Our approach for approximating the moments is the sequential truncation algorithm [20, 21][13, Table I]. Let us denote the approximate distribution with $\widetilde{q}_{x,u}(x_k, u_k) \triangleq \mathrm{N}(\begin{bmatrix} x_k \\ u_k\end{bmatrix}; \xi_{k|k}, \Xi_{k|k})$.

In Section A.2, certain moments of $q_{x,u}$ are required. They are approximated as

$$\begin{aligned}
x_{k|k} &\triangleq \mathbb{E}_{\widetilde{q}_{xu}}[x_k] = [\xi_{k|k}]_{1:n_x}, \qquad (30)\\
P_{k|k} &\triangleq \mathbb{V}_{\widetilde{q}_{xu}}[x_k] = [\Xi_{k|k}]_{1:n_x,1:n_x}, \qquad (31)\\
u_{k|k} &\triangleq \mathbb{E}_{\widetilde{q}_{xu}}[u_k] = [\xi_{k|k}]_{n_x+(1:n_z)}, \qquad (32)\\
U_{k|k} &\triangleq \mathbb{V}_{\widetilde{q}_{xu}}[u_k] = [\Xi_{k|k}]_{n_x+(1:n_z),n_x+(1:n_z)}, \qquad (33)\\
\Upsilon_{k|k} &\triangleq \mathbb{E}_{\widetilde{q}_{xu}}[x_k u_k^{\mathrm{T}}] - x_{k|k}u_{k|k}^{\mathrm{T}} = [\Xi_{k|k}]_{1:n_x,n_x+(1:n_z)}, \qquad (34)
\end{aligned}$$

where $n_x{+}(1:n_z)$ denotes $(n_x{+}1):(n_x{+}n_z)$.

A.2 Derivations for $q_{R,\Delta}$

Using equation (7b) and the approximation (29) we obtain

$$
\begin{aligned}
\log q_{R,\Delta}(R_k, \Delta_k) = {} & \mathbb{E}_{\widetilde{q}_{x,u}}\left[\log \mathrm{N}(z_k; C_k x_k + \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}), R_k)\right] \\
& + \log \mathrm{N}(\Delta_k; \Delta_{k|k-1}, R_k \otimes V_{k|k-1}) + \log \mathrm{IW}(R_k; \Psi_{k|k-1}, \nu_{k|k-1}) + c
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
= & -\frac{1}{2}\log\det(R_k) \\
& - \frac{1}{2}\mathrm{Tr}\Big\{\mathbb{E}_{\widetilde{q}_{x,u}}\Big[(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))(z_k - C_k x_k - \Delta_k(u_k - \sqrt{\tfrac{2}{\pi}}\mathbf{1}))^{\mathrm{T}}\Big] R_k^{-1}\Big\} \\
& - \frac{n_z}{2}\log\det(R_k) - \frac{1}{2}\mathrm{Tr}\left\{(\Delta_k - \Delta_{k|k-1})V_{k|k-1}^{-1}(\Delta_k - \Delta_{k|k-1})^{\mathrm{T}} R_k^{-1}\right\} \\
& - \frac{\nu_{k|k-1}}{2}\log\det(R_k) - \frac{1}{2}\mathrm{Tr}\left\{\Psi_{k|k-1}R_k^{-1}\right\} + c
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
= & -\frac{\nu_{k|k-1} + n_z + 1}{2}\log\det(R_k) \\
& - \frac{1}{2}\mathrm{Tr}\Big\{\Big(\Delta_k\big(U_{k|k} + (u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}}\big)\Delta_k^{\mathrm{T}} \\
& - \big(z_k(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} - C_k(\Upsilon_{k|k} + x_{k|k}(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}})\big)\Delta_k^{\mathrm{T}} \\
& - \Delta_k\big(z_k(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} - C_k(\Upsilon_{k|k} + x_{k|k}(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}})\big)^{\mathrm{T}} \\
& + (z_k - C_k x_{k|k})(z_k - C_k x_{k|k})^{\mathrm{T}} + C_k P_{k|k} C_k^{\mathrm{T}}\Big) R_k^{-1}\Big\} \\
& - \frac{1}{2}\mathrm{Tr}\left\{\Big((\Delta_k - \Delta_{k|k-1})V_{k|k-1}^{-1}(\Delta_k - \Delta_{k|k-1})^{\mathrm{T}} + \Psi_{k|k-1}\Big) R_k^{-1}\right\} + c
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
= & -\frac{\nu_{k|k-1} + n_z + 1}{2}\log\det(R_k) \\
& - \frac{1}{2}\mathrm{Tr}\Big\{\Big(\Delta_k\big(U_{k|k} + (u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} + V_{k|k-1}^{-1}\big)\Delta_k^{\mathrm{T}} \\
& - \big(z_k(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} - C_k(\Upsilon_{k|k} + x_{k|k}(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}}) + \Delta_{k|k-1}V_{k|k-1}^{-1}\big)\Delta_k^{\mathrm{T}} \\
& - \Delta_k\big(z_k(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} - C_k(\Upsilon_{k|k} + x_{k|k}(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}}) + \Delta_{k|k-1}V_{k|k-1}^{-1}\big)^{\mathrm{T}} \\
& + (z_k - C_k x_{k|k})(z_k - C_k x_{k|k})^{\mathrm{T}} + C_k P_{k|k} C_k^{\mathrm{T}} + \Delta_{k|k-1}V_{k|k-1}^{-1}\Delta_{k|k-1}^{\mathrm{T}} + \Psi_{k|k-1}\Big) R_k^{-1}\Big\}
\end{aligned}
\tag{38}
$$

$$
\begin{aligned}
= & -\frac{n_z}{2}\log\det(R_k) - \frac{1}{2}\mathrm{Tr}\{(\Delta_k - \Delta_{k|k})V_{k|k}^{-1}(\Delta_k - \Delta_{k|k})^{\mathrm{T}} R_k^{-1}\} \\
& - \frac{\nu_{k|k}}{2}\log\det(R_k) - \frac{1}{2}\mathrm{Tr}\{\Psi_{k|k}R_k^{-1}\}
\end{aligned}
\tag{39}
$$

where

$$
V_{k|k} = \big(U_{k|k} + (u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} + V_{k|k-1}^{-1}\big)^{-1}, \tag{40}
$$

$$
\Delta_{k|k} = \big((z_k - C_k x_{k|k})(u_{k|k} - \sqrt{\tfrac{2}{\pi}}\mathbf{1})^{\mathrm{T}} - C_k\Upsilon_{k|k} + \Delta_{k|k-1}V_{k|k-1}^{-1}\big)V_{k|k}, \tag{41}
$$

$$
\nu_{k|k} = \nu_{k|k-1} + 1, \tag{42}
$$

$$
\begin{aligned}
\Psi_{k|k} = {} & \Delta_{k|k-1}V_{k|k-1}^{-1}\Delta_{k|k-1}^{\mathrm{T}} - \Delta_{k|k}V_{k|k}^{-1}\Delta_{k|k}^{\mathrm{T}} + (z_k - C_k x_{k|k})(z_k - C_k x_{k|k})^{\mathrm{T}} \\
& + C_k P_{k|k} C_k^{\mathrm{T}} + \Psi_{k|k-1}.
\end{aligned}
\tag{43}
$$

Therefore,

$$
q_{R,\Delta}(R_k, \Delta_k) = \mathrm{N}(\Delta_k; \Delta_{k|k}, R_k \otimes V_{k|k})\,\mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k}). \tag{44}
$$

The following moments are required for the derivations of Section A.1:

$$\mathbb{E}_{q_{R,\Delta}}[\Delta_k] = \Delta_{k|k}, \tag{45}$$

$$R_{k|k} \triangleq \mathbb{E}_{q_{R,\Delta}}[R_k^{-1}]^{-1} = \frac{1}{\nu_{k|k} - n_z - 1} \Psi_{k|k}. \tag{46}$$

Eq. (46) follows from the fact that $R_k \sim \mathrm{IW}(\Psi_{k|k}, \nu_{k|k})$ implies that R_k^{-1} is Wishart-distributed with shape matrix $\Psi_{k|k}^{-1}$ and $\nu_{k|k} - n_z - 1$ degrees of freedom [17, Ch. 3.4]. Furthermore,

$$\begin{aligned}
&\mathbb{E}_{q_{R,\Delta}}[R_k^{-1}\Delta_k] \\
&= \iint R_k^{-1}\Delta_k \,\mathrm{N}(\Delta_k; \Delta_{k|k}, R_k \otimes V_{k|k})\, \mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k})\, \mathrm{d}\Delta_k \,\mathrm{d}R_k && (47)\\
&= \int R_k^{-1}\Delta_{k|k}\, \mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k})\, \mathrm{d}R_k && (48)\\
&= (\nu_{k|k} - n_z - 1)\Psi_{k|k}^{-1}\Delta_{k|k} && (49)\\
&= R_{k|k}^{-1}\Delta_{k|k} && (50)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{q_{R,\Delta}}[\Delta_k^{\mathrm{T}} R_k^{-1}\Delta_k] \\
&= \iint \Delta_k^{\mathrm{T}} R_k^{-1}\Delta_k \,\mathrm{N}(\Delta_k; \Delta_{k|k}, R_k \otimes V_{k|k})\, \mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k})\, \mathrm{d}\Delta_k \,\mathrm{d}R_k && (51)\\
&= \int (\mathrm{Tr}\{R_k R_k^{-1}\} V_{k|k} + \Delta_{k|k}^{\mathrm{T}} R_k^{-1}\Delta_{k|k})\, \mathrm{IW}(R_k; \Psi_{k|k}, \nu_{k|k})\, \mathrm{d}R_k && (52)\\
&= n_z V_{k|k} + (\nu_{k|k} - n_z - 1)\Delta_{k|k}^{\mathrm{T}} \Psi_{k|k}^{-1}\, \Delta_{k|k} && (53)\\
&= n_z V_{k|k} + \Delta_{k|k}^{\mathrm{T}} R_{k|k}^{-1}\Delta_{k|k}, && (54)
\end{aligned}$$

where (52) follows from the matrix-variate normal identity $\mathbb{E}[X^{\mathrm{T}}AX] = \mathrm{Tr}\{UA^{\mathrm{T}}\}V + M^{\mathrm{T}}AM$ for $X \sim \mathrm{N}(M, U \otimes V)$ [17, Ch. 2.3].

B Recursive Identification Algorithm for Linear Systems with Skewed Innovations

```
**Inputs:** $x_{1|0}, P_{1|0}, \Delta_{1|0}, V_{1|0}, \Psi_{1|0}, \nu_{1|0}, Q_{1:K}, C_{1:K}, z_{1:K}, \gamma$
**for** $k$ = 1 to $K$ **do**
  *Initialize*
    $x_{k|k} \leftarrow x_{k|k-1}$
    $u_{k|k} \leftarrow u_{k|k-1}$
    $\Delta_{k|k} \leftarrow \Delta_{k|k-1}$
    $V_{k|k} \leftarrow V_{k|k-1}$
    $\Psi_{k|k} \leftarrow \Psi_{k|k-1}$
    $\nu_{k|k} \leftarrow \nu_{k|k-1} + 1$
    **repeat**
        $R_{k|k} \leftarrow \frac{1}{\nu_{k|k}-n_z-1} \Psi_{k|k}$
      *Update* $q_{x,u}(x_k, u_k) \approx \mathrm{N}(\left[\begin{smallmatrix} x_k \\ u_k \end{smallmatrix}\right]; \xi_{k|k}, \Xi_{k|k})$
        $\xi_{k|k-1} \leftarrow \begin{bmatrix} x_{k|k-1} \\ n_z \sqrt{2/\pi} (I_{n_z} + n_z V_{k|k})^{-1} V_{k|k} \mathbf{1} \end{bmatrix}$
        $\Xi_{k|k-1} \leftarrow \mathrm{blockdiag}(P_{k|k-1}, (I + n_z V_{k|k})^{-1})$
        $\widetilde{C}_k \leftarrow \begin{bmatrix} C_k & \Delta_{k|k} \end{bmatrix}$
        $K_k \leftarrow \Xi_{k|k-1} \widetilde{C}_k^{\mathrm{T}} (\widetilde{C}_k \Xi_{k|k-1} \widetilde{C}_k^{\mathrm{T}} + R_{k|k})^{-1}$
        $\widehat{\xi}_{k|k} \leftarrow \xi_{k|k-1} + K_k (z_k - \widetilde{C}_k \xi_{k|k-1} + \Delta_{k|k} \sqrt{\frac{2}{\pi}} \mathbf{1})$
        $\widehat{\Xi}_{k|k} \leftarrow \Xi_{k|k-1} - K_k \widetilde{C}_k K_k^{\mathrm{T}}$
        $[\xi_{k|k}, \Xi_{k|k}] \leftarrow \texttt{seq\_trunc}(\widehat{\xi}_{k|k}, \widehat{\Xi}_{k|k}, \{n_{\mathrm{AR}}+1, \ldots, n_{\mathrm{AR}}+n_z\})$        ▷ See [13, Table I]
        $x_{k|k} \leftarrow [\xi_{k|k}]_{1:n_{\mathrm{AR}}}$
        $P_{k|k} \leftarrow [\Xi_{k|k}]_{1:n_{\mathrm{AR}}, 1:n_{\mathrm{AR}}}$
        $\widetilde{u}_{k|k} \leftarrow [\xi_{k|k}]_{n_{\mathrm{AR}}+(1:n_z)} - \sqrt{\frac{2}{\pi}} \mathbf{1}$
        $U_{k|k} \leftarrow [\Xi_{k|k}]_{n_{\mathrm{AR}}+(1:n_z), n_{\mathrm{AR}}+(1:n_z)}$
        $\Upsilon_{k|k} \leftarrow [\Xi_{k|k}]_{1:n_{\mathrm{AR}}, n_{\mathrm{AR}}+(1:n_z)}$
      *Update* $q_{R,\Delta}(R_k, \Delta_k) = \mathrm{N}(\Delta_k; \Delta_{k|k}, R_k \otimes V_{k|k}) \, \mathrm{IW}(R_k; R_{k|k}, \nu_{k|k})$
        $V_{k|k} \leftarrow \left(U_{k|k} + \widetilde{u}_{k|k} \widetilde{u}_{k|k}^{\mathrm{T}} + V_{k|k-1}^{-1}\right)^{-1}$
        $\Delta_{k|k} \leftarrow \left((z_k - C_k x_{k|k}) \widetilde{u}_{k|k}^{\mathrm{T}} - C_k \Upsilon_{k|k} + \Delta_{k|k-1} V_{k|k-1}^{-1}\right) V_{k|k}$
        $\Psi_{k|k} \leftarrow \Delta_{k|k-1} V_{k|k-1}^{-1} \Delta_{k|k-1}^{\mathrm{T}} - \Delta_{k|k} V_{k|k}^{-1} \Delta_{k|k}^{\mathrm{T}}$
            $+ (z_k - C_k x_{k|k})(z_k - C_k x_{k|k})^{\mathrm{T}} + C_k P_{k|k} C_k^{\mathrm{T}} + \Psi_{k|k-1}$
    **until converged**
  *Predict*
    $x_{k+1|k} \leftarrow x_{k|k}$
    $P_{k+1|k} \leftarrow P_{k|k} + Q_k$
    $\Delta_{k+1|k} \leftarrow \Delta_{k|k}$
    $V_{k+1|k} \leftarrow \frac{1}{\gamma} V_{k|k}$
    $\Psi_{k+1|k} \leftarrow \gamma \Psi_{k|k}$
    $\nu_{k+1|k} \leftarrow \gamma \, \nu_{k|k} + (1-\gamma) \cdot 2 n_z$
**end for**
**Outputs:** $x_{k|k}$ **and** $P_{k|k}$ **for** $k = 1, \ldots, K$
```