Robust Inference for State-Space Models with Skewed Measurement Noise

Henri Nurminen*, Tohid Ardeshiri†, Robert Piché*, and Fredrik Gustafsson†
henri.nurminen@tut.fi, tohid@isy.liu.se, robert.piche@tut.fi, fredrik@isy.liu.se
*Dept. of Automation Science and Engineering, Tampere University of Technology, Finland
†Division of Automatic Control, Linköping University, Sweden

Motivation

Heavy-tailed and skewed distributions arise e.g. in radio positioning, economics, biostatistics, and psychiatry.

Figure 1: Non-line-of-sight causes skewness and outliers to TOA ranging error.

Figure 2: Student’s t (middle) and skew-t (right) models accommodate an outlier, while Gaussian (left) gives a large estimation error.

Figure 3: Skew t (right) uses the information that large negative outliers are improbable unlike Gaussian (left) and Student’s t (middle).

Skew t-distribution

The skew t-distribution [5] is an extension of Student’s t-distribution, \( z \sim \text{ST}(\mu, R, \Delta, \nu) \) has a hierarchical formulation as a Gaussian with random scaling and random bias with known sign:

\[
\begin{align*}
  z & \sim N(\mu + \Delta u, \frac{R}{\sqrt{\nu+1}}) \\
u & \sim N_+(0, \frac{1}{2}) \quad & \text{if } \Delta = 0 \\
\lambda & \sim \text{Gamma}(\Delta, \frac{1}{2}) \quad & \text{if } \Delta \neq 0
\end{align*}
\]

The parameters are:
- \( \mu \): location
- \( \Delta \): skewness
- \( R \): spread
- \( \nu \): degrees of freedom

Figure 4: Skew-t densities with different \( \Delta s \)

Linear state-space model with skew-t measurement noise:

\[
\begin{align*}
x_k &= Ax_{k-1} + w_{k-1}, \\
w_{k-1} &\sim N(0, Q) \\
y_k &= Cx_k + e_k, \\
e_k &\sim \text{ST}(\mu, R, \Delta, \nu).
\end{align*}
\]

Skew t variational Bayes filter [1]

The filtering posterior \( p(x_k|y_1:k) \) is not analytical, so we seek to approximate the posterior by

\[
p(x_k, u_k, \lambda_k | y_{1:k}) = q(x_k)q(u_k)q(\lambda_k).
\]

Variational Bayes (VB) gives optimal \( q \) functions in Kullback–Leibler sense. VB is an EM-type algorithm: update one variable at a time.

for \( k = 1 \) to \( K \) do
  Initialize \( q(u_k) \) and \( q(\lambda_k) \)
  repeat
    Update \( q(x_k) = N(\hat{x}_k; \cdot, \cdot) \) given \( q(u_k) \) and \( q(\lambda_k) \)
    Update \( q(u_k) = N_u(\mu_u; \cdot, \cdot) \) given \( q(\lambda_k) \) and \( q(x_k) \)
    Update \( q(\lambda_k) = \text{Gamma}(\lambda_k; \cdot, \cdot) \) given \( q(x_k) \) and \( q(u_k) \)
  until Converged
  Predict \( p(x_{k+1} | y_{1:k}) \approx f(p(x_{k+1} | x_k) q(x_k) dx_k) \)
end for

VB with recursive truncation [3]

The VB approximation (1) can show serious variance underestimation. We relax (1) so that \( x_k \) and \( u_k \) are not approximated as independent:

\[
p(x_k, u_k, \lambda_k | y_{1:k}) \approx q(x_k, u_k)q(\lambda_k).
\]

\( q(x_k, u_k) \) is a truncated multivariate normal distribution, whose mean and covariance matrix can be approximated with the computationally light recursive truncation algorithm.

for \( k = 1 \) to \( K \) do
  Initialize \( q(\lambda_k) \)
  repeat
    Approximate \( q(x_k, u_k) \approx N([\tilde{x}_k^T; \cdot, \cdot]) \) given \( q(\lambda_k) \)
    Update \( q(\lambda_k) = \text{Gamma}(\lambda_k; \cdot, \cdot) \) given \( q(x_k, u_k) \)
  until Converged
  Predict \( p(x_{k+1} | y_{1:k}) \approx f(p(x_{k+1} | x_k) q(x_k) dx_k) \)
end for

References