presented at ICL-GNSS 2015, 22-24 June 2015, Gothenburg, Sweden

Statistical Trilateration with Skew-t Errors

Philipp Müller and Robert Piché
Department of Automation Science and Engineering (ASE)
Tampere University of Technology
P.O. Box 692, FI-33101 Tampere, Finland
Emails: philipp.muller@tut.fi, robert.piche@tut.fi

***Abstract*—In the problem of determining a target's location using radio signal time-of-flight to reference nodes with known locations, measurement errors can be skewed because of multipath effects. In this paper, range errors are modelled using the skew-t distribution. An Expectation-Maximisation (EM) algorithm for computing the unknown location is presented, and its accuracy is compared with a descending Gauss-Newton algorithm by simulations. The EM algorithm improves the positioning accuracy significantly. Furthermore, it is shown how to fit the parameters of a skew-t distribution to training data using a Gibbs sampler.**

I. Introduction

Trilateration is the process of determining the position of a target using measurements of distances (ranges) to reference nodes with known positions. The target's position can be computed as the solution of a nonlinear least squares (NLS) problem that is equivalent to nonlinear regression with an elliptical (e.g. normal or t-distributed) error model. Computational algorithms for NLS problems are well known, and in trilateration problems they outperform "closed-form" methods [1].

A drawback of using elliptical distributions is that they cannot model skewness in the distribution of measurement errors since they are symmetric. Skewness can arise in time-of-flight range measurements because of multipath effects. For example, the non-line-of-sight data from UWB network in [2] is clearly skewed (Fig. 4). There is therefore interest in extending trilateration methods to noise models that include skewness. The skew-t distribution, which is a parametric distribution family that includes the t-distribution as a special case, is well-suited for this purpose. There is a considerable body of research related to this distribution family in the statistical literature, and an extensive discussion can be found, for example, in Azzalini's recent monograph [3]. Much of the literature focuses on fitting the parameters of a skew-t distribution using methods such as the Expectation-Maximisation (EM) algorithm to given error data.

The object of this paper is to show how to apply skew-t statistical theory and methods to the trilateration problem, where the parameters of the skew-t distribution are known or have already been fitted to training data. For positioning the EM algorithm is used, which is an attractive choice for nonlinear regression because its monotone convergence ensures numerical stability. In addition, in nonlinear regression one can use a standard NLS solver inside the EM iteration loop.

This paper is organised as follows. The trilateration problem and basic properties of the skew-t distribution are reviewed in Section II. Section III presents an EM algorithm for solving the trilateration problem under the assumption of skew-t distributed measurement errors. The algorithm's performance is tested in Section IV and it is compared with a descending Gauss-Newton method. Section V explains briefly how to fit the parameters to given error data using a Gibbs sampler, and presents an example for the fitting. Some concluding remarks and an outlook are given in Section VI.

Notation : $\mathbf{x}$ and $x_{1:d}$ denote column vectors, $\mathbf{H}$ denotes a matrix, and underscores are used to denote random variables and random vectors in contexts where the distinction from deterministic variables is useful.

II. Model

A. Statistical trilateration

We use the following statistical formulation of the trilateration problem [4].

Let the unknown target location be represented by the d-dimensional random vector $\underline{x}_{1:d} = \underline{\mathbf{x}}$. The K scalar measurements are modelled as

$$\underline{y}_k | \, (\underline{\mathbf{x}} = \mathbf{x}) = h_k(\mathbf{x}) + \underline{v}_k \tag{1}$$

for $k \in 1, .., K$, where function $h_k : \mathbb{R}^d \to \mathbb{R}$ models the measurement geometry and $\underline{v}_1, .., \underline{v}_K$ are mutually independent random variables (additive errors). In addition, $\underline{v}_{1:K}$ and $\underline{\mathbf{x}}$ are independent.

The prior probability density function (pdf) of $\mathbf{x}$ is denoted as $p_{\underline{\mathbf{x}}}$, and the pdf of v_k as $p_{\underline{v}_k}$. The posterior distribution of $\mathbf{x}$ given the K-dimensional measurement vector $\mathbf{y}_{1:K}$ has the pdf

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_{1:K}}(\mathbf{x}|\mathbf{y}_{1:K}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} p_{\underline{v}_k}(y_k - h_k(\mathbf{x})) \tag{2}$$

A value of $\mathbf{x}$ that maximises (2) is called a maximum *a posteriori* (MAP) estimate. This MAP estimate coincides with the maximum likelihood (ML) estimate if the prior distribution is "flat", i.e. if $p_{\underline{\mathbf{x}}}(\mathbf{x}) \propto 1$.

In trilateration, the measurement function is the Euclidean distance between the target and a reference node at a known location $\mathbf{c}_k$:

$$h_k(\mathbf{x}) = ||\mathbf{x} - \mathbf{c}_k||. \tag{3}$$

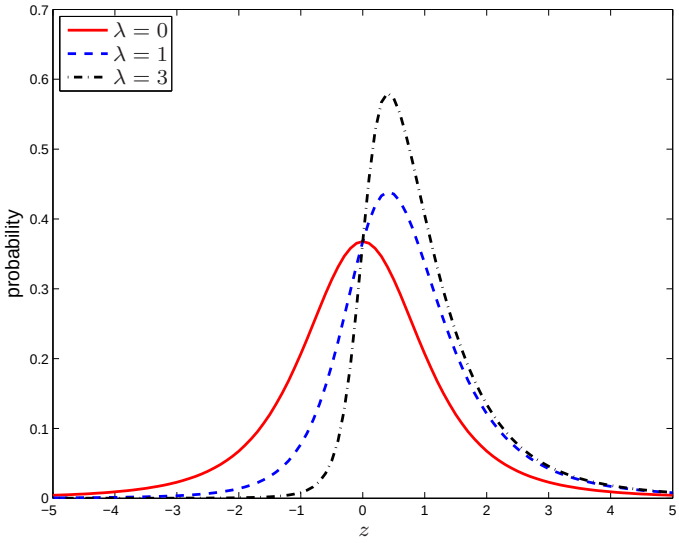

Fig. 1. Standardised skew-t distribution ($\nu = 3$).

The Jacobian of $\mathbf{h} = h_{1:K}$ is the $K \times d$ matrix $\mathbf{H}$ whose kth row is the transpose of a unit vector pointing from $\mathbf{c}_k$ to $\mathbf{x}$, that is,

$$\mathbf{H}_{k,1:d} = \nabla h_k(\mathbf{x}) = \frac{(\mathbf{x} - \mathbf{c_k})^T}{||\mathbf{x} - \mathbf{c}_k||}. \tag{4}$$

B. The skew-t distribution

In this subsection some general properties of the skew-t distribution are collected. For more extensive discussion see, for example, [3, pp. 101 ff.] and [5].

A random variable $\underline{z}$ is said to have a skew-t distribution with location ξ, scale σ^2, skewness λ and ν degrees-of-freedom (dof) if its pdf has the form

$$p_{\underline{z}}(z) = \frac{2}{\sigma} t_\nu \left(\frac{z-\xi}{\sigma} \right) T_{\nu+1} \left(\lambda \frac{z-\xi}{\sigma} \sqrt{\frac{\nu+1}{\nu + \frac{(z-\xi)^2}{\sigma^2}}} \right), \tag{5}$$

where t_ν and $T_{\nu+1}$ denote the pdf and the cumulative distribution function (cdf) of the standardised t-distribution. The skew-t distribution is denoted $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

The skew-t reduces to the t-distribution when $\lambda = 0$, that is, $\mathrm{ST}\left(\xi, \sigma^2, 0, \nu\right) = \mathrm{T}\left(\xi, \sigma^2, \nu\right)$, a t-distribution with location ξ, scale σ^2 and ν dof. The skew-t reduces to $\mathrm{N}\left(\xi, \sigma^2\right)$, a normal distribution with mean ξ and variance σ^2, when $\lambda = 0$ and $\nu \to \infty$. $\mathrm{ST}\left(0, 1, \lambda, \nu\right)$ is called a standardised skew-t distribution (Fig. 1) and if $\underline{u} \sim \mathrm{ST}\left(0, 1, \lambda, \nu\right)$ then $a + b\underline{z} \sim \mathrm{ST}\left(a, b^2, \lambda, \nu\right)$.

For $\nu > 2$ the mean and variance of $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ are

$$\mathrm{E}(\underline{z}) = \xi + \sigma g_\nu \delta_\lambda \quad , \quad \mathrm{var}(\underline{z}) = \sigma^2 \left(\frac{\nu}{\nu - 2} - (g_\nu \delta_\lambda)^2 \right) \tag{6}$$

where $g_\nu = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}} \in (-1, 1)$.

The skew-t distribution has the following hierarchical model. Let $\underline{\tau} \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and $\underline{w} \sim \mathrm{N}\left(0, 1\right)$, which denote the Gamma distribution with both shape and scale $\frac{\nu}{2}$ and the standard normal distribution. Then $\underline{t} = |\frac{\sigma}{\sqrt{\underline{\tau}}}\underline{w}|$ is a half-normal (HN) random variable with pdf

$$p_{\underline{t}}(t) \propto \Phi\left(t \frac{\sqrt{\tau}}{\sigma} \right) [t > 0], \tag{7}$$

where $[.]$ is the Iverson bracket and Φ denotes the cdf of the standard normal distribution. Samples from the distribution $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ can be drawn from the conditional distribution

$$\underline{z} | (\underline{t} = t, \underline{\tau} = \tau) \sim \mathrm{N}\left(\xi + \frac{\lambda t}{\sqrt{1+\lambda^2}}, \frac{1-\delta_\lambda^2}{\tau} \sigma^2 \right). \tag{8}$$

In the hierarchical representation (7 – 8), the conditional random variable $\underline{t} | (\underline{z} = z, \underline{\tau} = \tau)$ has the distribution $\mathrm{N}\left(\delta_\lambda (z - \xi), \frac{1-\delta_\lambda^2}{\tau} \sigma^2 \right)$ and the conditional random variable $\underline{\tau} | (\underline{z} = z)$ has the pdf

$$\underline{\tau} | (\underline{z} = z) \propto \tau^{(\nu-1)/2} \exp\left(-\frac{\tau}{2}(\eta^2 + \nu) \right) \Phi(\lambda \eta \sqrt{\tau}), \tag{9}$$

where $\eta = \frac{z-\xi}{\sigma}$.

III. Estimating the target position

A. Positioning using Expectation Maximisation

This subsection presents an expectation maximisation (EM) algorithm to compute the MAP estimate, i.e. the mode of (2), for additive skew-t measurement errors. In the maximisation step (M-step) the descending Gauss-Newton (GN) algorithm is used. A more detailed derivation of the equations used in the algorithm can be found in Subsection III.B.

The posterior pdf (2) for the measurement model $\underline{y}_k | (\underline{\mathbf{x}} = \mathbf{x}) \sim \mathrm{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right)$ is

$$\begin{aligned} &p_{\underline{\mathbf{x}} | \underline{\mathbf{y}}_{1:K}}(\mathbf{x} | \mathbf{y}_{1:K}) \\ &\quad \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} \frac{2}{\sigma} t_\nu \left(\frac{\bar{y}_k}{\sigma} \right) T_{\nu+1} \left(\lambda \frac{\bar{y}_k}{\sigma} \sqrt{\frac{\nu+1}{\nu + \frac{\bar{y}_k^2}{\sigma^2}}} \right), \end{aligned} \tag{10}$$

where $\bar{y}_k = y_k - h_k(\mathbf{x}) - \xi$. Using (8), a hierarchical version of the measurement model is

$$\begin{aligned} &\underline{y}_k | (\underline{\mathbf{x}} = \mathbf{x}, \underline{t}_k = t_k, \underline{\tau}_k = \tau_k) \\ &\quad \sim \mathrm{N}\left(\xi + h_k(\mathbf{x}) + \delta_\lambda t_k, \frac{1-\delta_\lambda^2}{\tau_k} \sigma^2 \right), \end{aligned} \tag{11}$$

where the hyperparameters are $\underline{t}_k = |\frac{\sigma}{\sqrt{\underline{\tau}_k}}\underline{w}_k|$ with $\underline{w}_k \sim \mathrm{N}\left(0, 1\right)$ and $\underline{\tau}_k \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

In the EM algorithm's expectation step (E-step) the hyperparameters are updated by setting them to the mean values of their conditional distribution, i.e.

$$\hat{\tau}_k \leftarrow \frac{\sigma^2 \left(1 - \delta_\lambda^2\right) \nu^2}{\left(y_k - h_k(\hat{\mathbf{x}}) - \xi - \delta_\lambda t_k\right)^2 + 4\sigma^2 \left(1 - \delta_\lambda^2\right)} \tag{12a}$$

$$\hat{t}_k \leftarrow \mu_k + \frac{\phi(-\mu_k/\sigma_k)}{1 - \Phi(-\mu_k/\sigma_k)} \sigma_k \tag{12b}$$

for $k = \{ 1, .., K \}$, where $\hat{\mathbf{x}}$ is the current MAP estimate, and ϕ and Φ denote the pdf and the cdf of the standard normal

distribution. In (12b) $\mu_k = \frac{1}{2\delta_\lambda}\left(y_k - h_k(\hat{\mathbf{x}}) - \xi\right)$ and $\sigma_k = \sqrt{\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}$.

Assuming a multivariate-normal prior distribution for $\mathbf{x}$ with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$, the M-step is the optimisation of the conditional posterior

$$p(\mathbf{x}|\mathbf{y}_{1:K}, t_{1:K}, \tau_{1:K}) \propto \mathrm{e}^{-\frac{1}{2}\left((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)+\frac{\sum_{k=1}^K \tau_k\left(\mathbf{y}_{1:K}-\mathbf{h}_{1:K}(\mathbf{x})-\xi-\delta_\lambda t_k\right)^2}{\sigma^2\left(1-\delta_\lambda^2\right)}\right)} \tag{13}$$

with $\tau_k = \hat{\tau}_k$ and $t_k = \hat{t}_k$, that is

$$\hat{\mathbf{x}} \leftarrow \arg\min_{\mathbf{x}} \left((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0) + \frac{\sum_{k=1}^K \hat{\tau}_k\left(y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda \hat{t}_k\right)^2}{\sigma^2\left(1-\delta_\lambda^2\right)}\right) \tag{14}$$

The minimisation in (14) can be computed by any nonlinear least-squares optimisation method. In this paper the descending Gauss-Newton algorithm (see e.g. [4]) is used, which for (14) is the iteration of

$$\mathbf{H} \leftarrow \frac{\partial \mathbf{h}_{1:K}}{\partial \mathbf{x}}(\hat{\mathbf{x}}) \tag{15a}$$

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha\left(\mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)\right) \tag{15b}$$

with $\mathbf{K} = \mathbf{P}_0\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}_0\mathbf{H}^T\right)^{-1}$, where $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\mathrm{diag}\{1/\hat{\tau}_1, .., 1/\hat{\tau}_K\}$, and modified "data" $\tilde{\mathbf{y}} = \mathbf{y}_{1:K} - \xi - \delta_\lambda \hat{t}_{1:K}$. The scale factor α ensures that the cost function is decreasing, and is found by line search (see Subsection III.B).

In case of a flat prior, (15b) is replaced by

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} - \alpha\left(\left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}}\right)\right) \tag{16}$$

To summarise, given independent scalar measurements $\mathbf{y}_{1:K}$ with additive ST $\left(\xi, \sigma^2, \lambda, \nu\right)$ noise, and a multivariate normal prior with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$, the EM algorithm computes the MAP estimate $\hat{\mathbf{x}}$ as shown in Algorithm 1.

Algorithm 1 Computing position estimate by EM

Input: $\mathbf{y}_{1:K}$, $\mathbf{m}_0$, $\mathbf{P}_0$, n_{EM} (number of EM iterations) and n_{dGN} (number of descending GN iterations)
Initialise $\mathbf{x}^{(0)} \leftarrow \mathbf{m}_0$, $t_{1:K}^{(0)} \leftarrow -\xi/\delta_\lambda$ and $\tau_{1:K}^{(0)} \leftarrow 1$
for $j = 1$ to n_{EM} **do**
  Given $t_{1:K}^{(j-1)}$ and $\tau_{1:K}^{(j-1)}$, initialise $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(j-1)}$, and compute modified "data" $\tilde{\mathbf{y}} = \mathbf{y}_{1:K} - \xi - \delta_\lambda t_{1:K}^{(j-1)}$ and $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\mathrm{diag}\{1/\tau_1^{(j-1)}, .., 1/\tau_K^{(j-1)}\}$
  for $i = 1$ to n_{dGN} **do**
    Compute $\mathbf{H}$ using (15a)
    Compute $\mathbf{K} = \mathbf{P}_0\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}_0\mathbf{H}^T\right)^{-1}$ and $d_{\text{GN}} = \mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)$ (or $d_{\text{GN}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}}\right)$ if $\mathbf{P}_0^{-1} = \mathbf{0}$, i.e. if prior is flat)
    Compute $f(\hat{\mathbf{x}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}}$
    Set $\alpha \leftarrow 1$ and compute $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
    while $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}}) \geq f(\hat{\mathbf{x}})$ **do**
      Set $\alpha \leftarrow \alpha/2$, and compute $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
    end while
    Set $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
  end for
  Set $\mathbf{x}^{(j)} \leftarrow \hat{\mathbf{x}}$
  Given $\mathbf{x}^{(j)}$, compute $t_{1:K}^{(j)}$ and $\tau_{1:K}^{(j)}$ using (12)
end for

B. Details of the EM algorithm

In this subsection the equations of the EM presented above are derived and details of the descending Gauss-Newton algorithm are given. Assuming a multivariate-normal prior distribution for $\mathbf{x}$ with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$ and given the priors for the hyperparameters, the joint prior distribution is

$$\begin{aligned} p(\mathbf{x}, t_{1:K}, \tau_{1:K}) &\propto p_{\underline{\mathbf{x}}}(\mathbf{x}) p_{\underline{t}_{1:K}|\underline{\tau}_{1:K}}(t_{1:K}|\tau_{1:K}) p_{\underline{\tau}_{1:K}}(\tau_{1:K}) \\ &= \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)} \prod_{k=1}^K [t_k \geq 0] \\ &\quad \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \tau_k t_k^2} \prod_{k=1}^K \tau_k^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{2}{\nu}\tau_k}. \end{aligned} \tag{17}$$

Thus, the joint posterior is

$$\begin{aligned} p(\mathbf{x}, t_{1:K}, \tau_{1:K}|\mathbf{y}_{1:K}) &\propto \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \frac{\tau_k}{1-\delta_\lambda^2}(\bar{y}_k - \delta_\lambda t_k)^2} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)} \\ &\quad \prod_{k=1}^K [t_k \geq 0] \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \tau_k t_k^2} \prod_{k=1}^K \tau_k^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{2}{\nu}\tau_k}. \end{aligned} \tag{18}$$

Denoting $\tau_{-k} = \{\tau_1, .., \tau_{k-1}, \tau_{k+1}, .., \tau_K\}$ the conditional distribution of τ_k is

$$\begin{aligned} p(\tau_k|\mathbf{x}, \mathbf{y}_{1:K}, t_{1:K}, \tau_{-k}) &\propto \mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}(\bar{y}_k - \delta_\lambda t_k)^2} \tau_k^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{2}{\nu}\tau_k} \\ &= \tau_k^{\frac{\nu}{2}-1} \mathrm{e}^{-\tau_k \frac{(\bar{y}_k - \delta_\lambda t_k)^2 + 4\sigma^2\left(1-\delta_\lambda^2\right)}{2\sigma^2\left(1-\delta_\lambda^2\right)\nu}}, \end{aligned} \tag{19}$$

which is a Gamma distribution with shape parameter $\nu/2$ and scale parameter $\frac{2\sigma^2\left(1-\delta_\lambda^2\right)\nu}{(\bar{y}_k - \delta_\lambda t_k)^2 + 4\sigma^2\left(1-\delta_\lambda^2\right)}$. The mean value of this Gamma distribution, which is the product of shape and scale, is used in (12a) to update τ_k.

Denoting $t_{-k} = \{t_1, .., t_{k-1}, t_{k+1}, .., t_K\}$ the conditional

TABLE I
Number of operations from different classes for descending GN and EM algorithm, dependent on number of measurements K, number of EM iterations n_{EM} and number of descending GN iterations n_{dGN}.

Class	dGN	EM
addition	$\mathcal{O}(n_{\text{dGN}}K^3)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K^3)$
subtraction	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$
multiplication	$\mathcal{O}(n_{\text{dGN}}K^3)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K^3)$
division	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$
other	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$

distribution of t_k is

$$\begin{aligned} p(t_k|\mathbf{x},\mathbf{y}_{1:K},t_{-k},&\tau_{1:K}) \\ &\propto \mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}\left(\bar{y}_k-\delta_\lambda t_k\right)^2}[t_k\geq 0]\mathrm{e}^{-\frac{\tau_k}{2\sigma^2}t_k^2} \\ &= [t_k\geq 0]\mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}\left(\left(\frac{\bar{y}_k}{\delta_\lambda}-t_k\right)^2+t_k^2\right)} \\ &\propto [t_k\geq 0]\mathrm{e}^{-\frac{1}{2\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}\left(t_k-\frac{\bar{y}_k}{2\delta_\lambda}^2\right)}, \end{aligned} \tag{20}$$

which is a truncated normal distribution, i.e. values smaller zero are not allowed (due to $[t_k \geq 0]$), with center $\mu_k = \frac{1}{2\delta_\lambda}\left(y_k - h_k(\mathbf{x}) - \xi\right)$ and scale $\sigma_k = \sqrt{\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}$. The mean value of this distribution is used in (12b) to update t_k.

The α used in (15b) and (16) ensures that the cost function

$$\begin{aligned} f(\mathbf{x}) \quad = \quad & \frac{1}{2}\Big((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0) \\ & + \left(\mathbf{h}_{1:K}(\mathbf{x})-\tilde{\mathbf{y}}\right)^T\mathbf{R}^{-1}\left(\mathbf{h}_{1:K}(\mathbf{x})-\tilde{\mathbf{y}}\right)\Big) \end{aligned} \tag{21}$$

does not increase, which is possible in the standard GN algorithm which uses the step $d_{\text{GN}} = \mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)$. After initialising $\alpha \leftarrow 1$ its value is repeatedly halved as long as $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}}) \geq f(\hat{\mathbf{x}})$.

Table I shows a detailed complexity analysis of the EM method and compares it with the complexity of the descending GN, which will be used in Section IV as reference. The EM algorithm has for each operation class a n_{EM} times higher computational complexity. The number of operations in the E-step, for updating the hyperparameters, is negligible small in comparison to the number of operations in the M-step (the descending GN).

IV. Simulation experiment

In this section the positioning performance of using the skew-t likelihood model instead of a normal distribution likelihood model when the measurement errors are skew-t distributed is analysed. The MAP estimate for the normal model is computed using the descending GN algorithm. The source code and the full test suite are available at `https://PMullerTUT@bitbucket.org/PMullerTUT/trilaterationskewterrors.git`.

Here $\underline{\mathbf{x}}$ is a two-dimensional position and is assumed to have the prior distribution

$$\mathbf{x} \sim \text{MVN}\left(\mathbf{m}_0, \mathbf{P}_0\right) = \text{MVN}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], 100\,\mathbf{I}_{2\times 2}\right), \tag{22}$$

TABLE II
Positioning error statistics for simulations with additive skew-t noises. Column *Time* gives the relative computation time using a specific MATLAB implementation, scaled so that computation time for the descending GN is 1.

Method	Time	Mean	Median	95 perc.
dGN	1	1.53	1.39	2.97
EM	5	1.15	1.05	2.59

where MVN$(.,.)$ denotes a bivariate Gaussian distribution with given mean and covariance matrix. Four reference nodes are located at the corners of a 40-by-40 square centred at $\mathbf{m}_0$.

For the experiment 100 target positions are drawn from the prior distribution (22). Using (3) for computing the true distance between the target position and the reference node, $K = 12$ independent distance measurements (three to each reference node) are drawn from

$$y_k|\mathbf{x} \sim \text{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right) = \text{ST}\left(2 + h_k(\mathbf{x}), 3^2, 3, 3\right). \tag{23}$$

The EM algorithm uses the hierarchical version (11) of the measurement model with initial values $\tau_k = 1$ and $t_k = -\xi/\delta_\lambda$, which ensures that the descending GN finds the minimiser of the likelihood for Gaussian noise. For the EM algorithm 4 iterations are performed, and in each M-step 4 iterations of the descending GN algorithm are performed.

For comparison a descending GN that assumes measurement errors to be distributed as

$$y_k|\mathbf{x} \sim \text{N}\left(\xi + \sigma g_\nu \delta_\lambda + h_k(\mathbf{x}), \sigma^2\left(\frac{\nu}{\nu-2} - (g_\nu\delta_\lambda)^2\right)\right) \tag{24}$$

with 4 iterations is used. The parameters given in (24) ensure that the normal distribution has the same mean and variance as the skew-t distribution used by the EM algorithm.

In both the EM and the descending GN algorithm the number of repetitions to find a suitable α for (15b) and (16) is limited to 5 [4].

Table II presents the error statistics for the algorithms. *Mean* is the empirical mean, *Median* is the empirical median and *95% err* is the 95th percentile of all two-dimensional positioning errors, which are defined as the Euclidean distance (compare (3)) between the true position $\mathbf{x}$ and the position estimate $\hat{\mathbf{x}}$. *Time* gives the relative computation time using a specific MATLAB implementation, scaled so that computation time for the descending GN is 1. Fig. 1 shows the first 50 simulated positions and the corresponding estimates by EM and descending GN algorithm.

The EM algorithm clearly outperforms the descending GN algorithm in all three accuracy measures. That improvement in precision comes at the cost of an approximately five times higher computation time, which was expected since the EM uses four times more descending GN iterations and uses in addition the E-step for updating the hyperparameters. However, the computing cost is still reasonable, and could be brought down by tweaking the algorithm code and parameters.

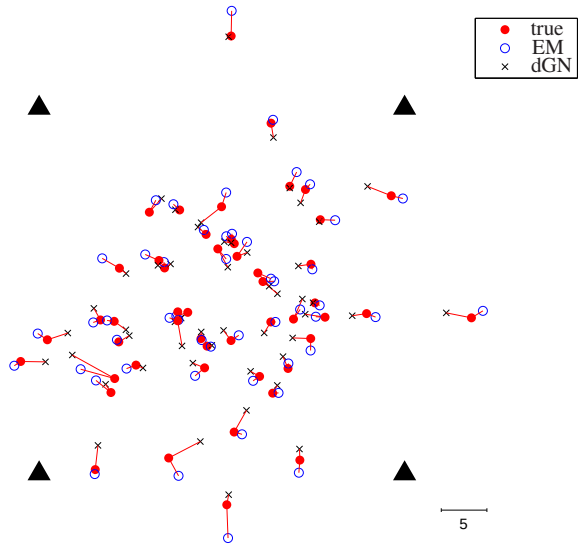

Fig. 2. First 50 simulated positions and their corresponding estimates computed by the EM and the descending GN algorithms.

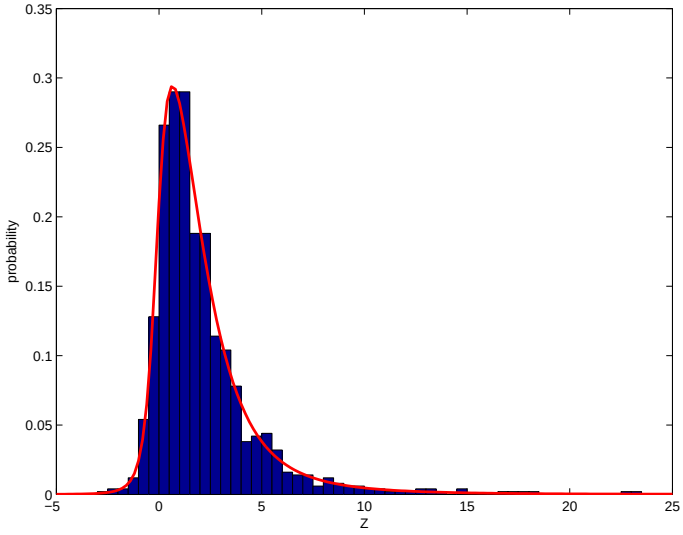

Fig. 3. Histogram of training data drawn from $v_j \sim \mathrm{ST}\left(0, 3^2, 3, 3\right)$ and pdf of a skew-t distribution that uses the median values given in Table III for $T_0 = 1\,000$ and $T = 10\,000$.

V. Fitting parameters of a skew-t distribution

A. Fitting parameters using a Gibbs sampler

In Section IV the error distribution's parameters are assumed to be known. For real-world applications it will be necessary to estimate the parameters by fitting them to a set of training data. In the trilateration setting (Section IV), training data are typically obtained in a measurement campaign in which ranges observed from known locations are collected, which allows determination of ranging errors. This section briefly describes the Gibbs sampling algorithm [6], which can be used to compute statistics of the posterior distributions for the parameters from n independent ranging errors $v_j|\xi, \sigma^2, \lambda, \nu \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

The Gibbs sampling method could also be used to compute position, but can be expected to be much slower than EM.

The idea of the Gibbs sampler (GS) is to sample from the conditional posterior distributions for each parameter separately when sampling from the (multivariate) posterior is not feasible. For the parameter estimation problem at hand generating samples from the joint posterior $p\left(\xi, \sigma^2, \lambda, \nu|v_{1:n}\right)$ is unfeasible, but sampling from the conditional posterior distributions $p\left(\xi|\sigma^2, \lambda, \nu, v_{1:n}\right)$, $p\left(\sigma^2|\xi, \lambda, \nu, v_{1:n}\right)$, $p\left(\lambda|\xi, \sigma^2, \nu, v_{1:n}\right)$ and $p\left(\nu|\xi, \sigma^2, \lambda, v_{1:n}\right)$ is possible.

The algorithm works as follows. First, initial values $\xi_{(0)}, \sigma^2_{(0)}, \lambda_{(0)}, \nu_{(0)}$ are assigned to the unknown parameters. Then the parameters are ordered and samples from the conditional distribution of each parameter given the error data $v_{1:n}$ and the current estimates of the remaining parameters are drawn. This updating process is repeated $T_0 + T$ times. The posterior means of the estimates are estimated by the empirical sample means of the last T samples; the first T_0 "burn-in" samples are discarded. For example,

$$\mathrm{E}\left(\hat{\xi}|v_{1:n}\right) \approx \frac{1}{T}\sum_{t=1}^{T} \xi_{(t+T_0)}. \tag{25}$$

Other statistics, such as the posterior median, can be estimated in a similar way.

Since in the problem under consideration four parameters have to be estimated a large set of training data will be required to obtain estimates that are close to the true values of the parameters. If priors are vague, a large training set is needed to obtain a posterior with small dispersion. A smaller training set can be used if the priors are more informative.

B. An example for fitting parameters by the GS

In this subsection a small demo for fitting parameters of a skew-t distribution using the GS algorithm is presented. The training data consists of $n = 1\,000$ observations that are drawn from

$$v_j|\xi = 0, \sigma^2 = 9, \lambda = 3, \nu = 3 \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right). \tag{26}$$

The source code and the full test suite are available at `https://PMullerTUT@bitbucket.org/PMullerTUT/trilaterationskewterrors.git`.

For estimating the parameters a hierarchical model similar to (11) and JAGS [7], which is a program that allows analysing Bayesian hierarchical models by Gibbs samplers, are used. Algorithm 2 shows the pseudo-code for the model used in JAGS.

The GS is run twice with different $\xi_{(0)}, \sigma^2_{(0)}, \lambda_{(0)}, \nu_{(0)}$. For each parameter a diffuse prior is used; only for ν a slightly informative prior, namely a uniform distribution over the interval $(2, 100)$ is used because $\nu > 2$ is required in (6).

Table III contains the summary statistics of the posterior distribution for different numbers of "burn-in" and retained

Algorithm 2 Model for JAGS used for fitting parameters

Input: number of measurements n
Draw parameters from priors: $\xi \sim \mathrm{N}\left(0, 100^2\right)$, $\sigma = 1/p$ with $p \sim \Gamma\left(10\,000, 10\,000\right)$, $\lambda \sim \mathrm{N}\left(0, 100^2\right)$ and $\nu \sim \mathrm{Uniform}\left(2, 100\right)$
Compute $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}}$
for $j = 1$ to n **do**
  draw $\tau_j \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$
  draw $t_j \sim \mathrm{HN}\left(0, \frac{\sigma^2}{\tau_j}\right)$ (see (7) for its pdf)
  draw $v_j \sim \mathrm{N}\left(\xi + \delta_\lambda t_j, \frac{1-\delta_\lambda^2}{\tau_j}\sigma^2\right)$
end for
draw predicted latent variable $\tau^{(\mathrm{pr})} \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$
draw predicted latent variable $t^{(\mathrm{pr})} \sim \mathrm{HN}\left(0, \frac{\sigma^2}{\tau^{(\mathrm{pr})}}\right)$
draw predicted observation $v^{(\mathrm{pr})} \sim \mathrm{N}\left(\xi + \delta_\lambda t^{(\mathrm{pr})}, \frac{1-\delta_\lambda^2}{\tau^{(\mathrm{pr})}}\sigma^2\right)$

TABLE III
Statistics of posterior distributions estimated from 1 000 observations, drawn from $v_j|\xi = 0, \sigma^2 = 9, \lambda = 3, \nu = 3 \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$, by two Gibbs samplers with each T_0 "burn-in" and T retained samples.

Parameter	T_0	T	5%-ile	Median	95%-ile
ξ	200	500	-0.230	-0.114	-0.001
σ^2	200	500	3.474	4.129	5.065
λ	200	500	2.867	3.681	4.617
ν	200	500	2.390	2.818	3.410
ξ	1000	10000	-0.219	-0.115	-0.008
σ^2	1000	10000	3.455	4.159	5.001
λ	1000	10000	2.987	3.646	4.459
ν	1000	10000	2.390	2.813	3.389

samples. For each parameter the median, and 5%-ile and 95%-ile are given.

Already with $T_0 = 200$ "burn-in" and $T = 500$ retained samples the parameter estimates are quite good; only σ^2 is underestimated slightly. Using $T_0 = 1\,000$ and $T = 10\,000$ does not improve the quality of the estimates significantly. To further improve the estimates more training data (i.e. observations) would be required or could be fixed to reasonable values and the GS could be repeatedly run with various fixed ν for estimating ξ, σ^2 and λ. However, from Fig. 3, which shows the histogram of the training data and the pdf of a skew-t distribution that uses the median values given in Table III for $T_0 = 1\,000$ and $T = 10\,000$, these approaches seem to be unnecessary.

The computation times on a laptop where 15 seconds for $T_0 = 200$ and $T = 500$ and less than 4 minutes for $T_0 = 1\,000$ and $T = 10\,000$, which is tolerable for an algorithm that is intended for offline use.

As mentioned in the introduction, the non-line-of-sight data from UWB network in [2] is clearly skewed. Fig. 4 shows the density histogram of the data and a skew-t distribution fitted by JAGS using $T_0 = 1\,000$ and $T = 10\,000$ and two GS.

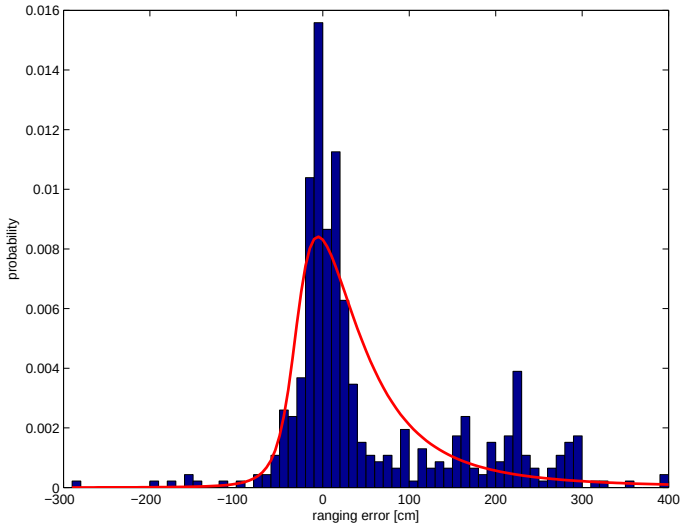

Fig. 4. Density histogram of non-line-of-sight errors from UWB data in [2] and fitted skew-t ST $(-30.73, 4450.05, 3.22, 2.05)$.

VI. Concluding Remarks

This paper explained how the trilateration problem can be solved using an Expectation-Maximisation algorithm when the measurements contain additive skew-t distributed errors. For such measurement data it was shown by simulations that the presented EM algorithm improves the positioning precision significantly compared with a descending Gauss-newton algorithm, which models the measurement noise as being additive normal. This improvement in precision comes at the cost of higher computational demand.

In addition, it is shown how the parameters of a skew-t can be fitted to training data by using a Gibbs sampler.

Future work includes the testing of the algorithm with real-world data that has skew measurement noise. Moreover, the algorithm can be modified for other measurement functions, e.g. pseudo-ranges, angle-of-arrival or time-of-arrival type measurements, and the algorithm could be implemented in a filter to further improve its positioning precision.

References

[1] N. Sirola, "Closed-form algorithms in mobile positioning: myths and mis- conceptions," in *Proceedings of the 7th Workshop on Positioning, Navigation and Communications*, 2010, pp. 38–44.

[2] P. Müller, H. Wymeersch, and R. Piché, "UWB positioning with generalized Gaussian mixture filters," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2406–2414, October 2014.

[3] A. Azzalini, *The skew-normal and related families*. Cambridge University Press, 2014.

[4] R. Piché, "Estimation of model parameters," in *Mathematical Modeling with Multidisciplinary Applications*, X.-S. Yang, Ed. John Wiley and Sons, 2013, pp. 169–190.

[5] T. I. Lin, J. C. Lee, and W. J. Hsieh, "Robust mixture modeling using the skew t distribution," *Statistics and Computing*, vol. 17, no. 2, pp. 81–92, June 2007.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[7] (2013, September). [Online]. Available: http://mcmc-jags.sourceforge.net