Author(s) Piché, Robert

Title Robust estimation of a reception region from location fingerprints

Citation Piché, Robert 2011. Robust estimation of a reception region from location fingerprints . 2011 International Conference on Localization and GNSS ICL-GNSS June 29-30, 2011 Tampere, Finland. International Conference on Localization and GNSS ICL-GNSS Piscataway, NJ, IEEE . 31-35.

Year 2011

DOI http://dx.doi.org/10.1109/ICL-GNSS.2011.5955261

Version Post-print

URN http://URN.fi/URN:NBN:fi:tty-201406191311 presented at ICL-GNSS, June 29-30, 2011, Tampere, Finland

Robust estimation of a reception region from location fingerprints

Robert Piché
Tampere University of Technology
robert.piche@tut.fi

***Abstract*— A method for fitting an ellipse-shaped reception region to a set of location-stamped radio signal reception reports, or location fingerprints, is presented. Reports are modelled as having a multivariate Student distribution. The method is less sensitive to outliers than existing smallest-enclosing ellipse and Normal-distribution based methods. A Gibbs sampling algorithm and an Expectation-Maximisation algorithm to compute ellipse parameters are presented.**

I. Introduction

Many positioning systems make use of a database of radio reception reports or "fingerprints". At its simplest, a fingerprint is a list of the radio signals (e.g. phone cell id's or WiFi access point addresses) that can be heard by a receiver, along with a timestamp and the receiver's location. Fingerprints can be collected in a systematic survey ("wardriving") or by (perhaps surreptitiously) polling GPS-equipped mobile devices.

A fingerprint database can be enormous, for example Skyhook Wireless claims to maintain a reference database of over 250 million WiFi and cellular phone access points. There is therefore interest in compressing fingerprint information in order to facilitate data storage and transmission as well as position calculations. One solution is to distill fingerprints into a database of signal sources' *reception regions* (coverage areas), the sets of locations where the signal strength exceeds some predefined threshold level. Actual reception regions are irregularly shaped (Fig. 1) and depend on the chosen signal strength threshold, receiver antenna orientation, and changing environmental factors (e.g. atmospheric humidity, number of people nearby), but reception region models for practical large-scale positioning systems need to be very simple.

A simple reception region model is a disk, centred for example at the mean or the median of the fingerprints. A slightly more elaborate reception region model is the smallest ellipse that contains all the fingerprints. Ellipses only need 5 real numbers to be specified, but provide a reasonable approximation of a convex region. (Some ellipse parametrisations are summarised in Appendix I.)

In [1], ellipse-shaped reception regions are fitted by modelling reception reports as having a Normal (Gaussian) distribution. Bayesian estimates of the model parameters can be rapidly computed using closed-form formulas available in the literature (e.g. [2]). Prior knowledge about typical radio signal range can be exploited to improve the reception region estimate.

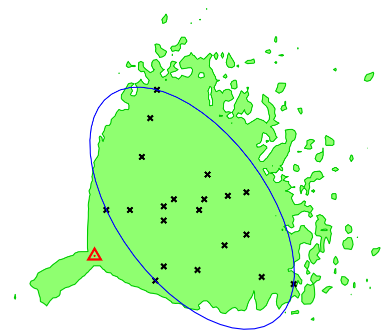

Fig. 1. A conceptual presentation of a mobile phone network cell id reception region (green), reception reports (x), and the smallest ellipse that encloses them. The base station location (triangle), which in this case is not inside the ellipse, is not needed for positioning.

The smallest ellipse containing all the fingerprints can also be given a Bayesian interpretation. It can be shown to be the maximum a-posteriori estimate when the probability of reception is modelled to be uniform inside an ellipse and zero outside, and the prior probability distribution of the ellipse parameters is assumed to be uniform.

Real fingerprint data can be expected to include some *outliers*, that is, locations that are not in the normal reception region are reported because of unusual reception conditions, software or hardware malfunctions in the GPS or radio signal reception, etc. The smallest enclosing ellipse region-fitting method is obviously very sensitive to even a single outlier, and the Normal regression model is well known to be overly sensitive to outliers. For both these methods, outliers produce reception regions that are too large. Various outlier-screening heuristics exist and these are usually adequate for dealing with a small number of gross outliers. "Moderate" outliers are more difficult to handle, especially if they are numerous.

The Student t distribution, of which the Cauchy distribution is a special case, is known as an alternative to the Normal distribution that, because of its heavy tails, is better suited as a model of data that may contain outliers [3]. Student regression can be computed numerically by standard Gibbs sampling or Expectation Maximisation (EM) algorithms.

The objective of this work is to present in detail the fitting of ellipse-shaped reception regions using a Student model of reception reports. The paper is organised as follows. Section II reviews the Normal distribution-based fitting method. The

Student data model is presented in section III-A. EM and Gibbs sampling algorithms for Student regression are detailed in sections III-B and III-C. The use of prior information about reception region sizes to improve the region estimate is discussed in section IV.

II. Normal Model

A. Distributions

Here, each observation is modelled as a d-variate Normal random vector $\mathbf{x}_n$ with mean $\mathbf{m}$ and covariance $\mathbf{Q}^{-1}$,

$$\mathbf{x}_n \,|\, \mathbf{m}, \mathbf{Q} \sim \text{Normal}(\mathbf{m}, \mathbf{Q}^{-1})$$
$$p(\mathbf{x}_n \,|\, \mathbf{m}, \mathbf{Q}) = \frac{|\mathbf{Q}|^{\frac{1}{2}}}{(2\pi)^{d/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}_n-\mathbf{m})'\mathbf{Q}(\mathbf{x}_n-\mathbf{m})}$$

Assuming that N observations are conditionally independent given the model parameters, the likelihood density for the $d \times N$ observation array $\mathbf{x}_{1:N} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ is

$$\begin{aligned} p(\mathbf{x}_{1:N} \,|\, \mathbf{m}, \mathbf{Q}) &= \prod_{n=1}^{N} p(\mathbf{x}_n \,|\, \mathbf{m}, \mathbf{Q}) \\ &\propto |\mathbf{Q}|^{\frac{N}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\mathbf{m})'\mathbf{Q}(\mathbf{x}_n-\mathbf{m})} \\ &= |\mathbf{Q}|^{\frac{N}{2}} \mathrm{e}^{-\frac{N}{2}(\mathbf{m}-\mathbf{c})'\mathbf{Q}(\mathbf{m}-\mathbf{c})-\frac{1}{2}\operatorname{tr}\mathbf{QS}} \end{aligned}$$

where the empirical mean and the residual sum of squares are

$$\mathbf{c} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n, \quad \mathbf{S} = \sum_{n=1}^{N} (\mathbf{x}_n - \mathbf{c})(\mathbf{x}_n - \mathbf{c})' \tag{1}$$

respectively.

Assuming the uninformative scale and rotation invariant improper prior distribution for the model parameters

$$p(\mathbf{m}, \mathbf{Q}) \propto |\mathbf{Q}|^{-(d+1)/2} \tag{2}$$

leads to the posterior distribution with density

$$\begin{aligned} p(\mathbf{m}, \mathbf{Q} \,|\, \mathbf{x}_{1:N}) &\propto p(\mathbf{x}_{1:N} \,|\, \mathbf{m}, \mathbf{Q}) p(\mathbf{m}, \mathbf{Q}) \\ &\propto |\mathbf{Q}|^{\frac{N-d-1}{2}} \mathrm{e}^{-\frac{N}{2}(\mathbf{m}-\mathbf{c})'\mathbf{Q}(\mathbf{m}-\mathbf{c})-\frac{1}{2}\operatorname{tr}\mathbf{QS}} \end{aligned}$$

This is called a Normal-Wishart distribution because the posterior conditional distribution of $\mathbf{m}$ given $\mathbf{Q}$ is Normal,

$$\mathbf{m} \,|\, \mathbf{x}_{1:N}, \mathbf{Q} \sim \text{Normal}(\mathbf{c}, (N\mathbf{Q})^{-1})$$

and the posterior marginal distribution of $\mathbf{Q}$ is a Wishart distribution with $N-1$ degrees of freedom,

$$\mathbf{Q} \,|\, \mathbf{x}_{1:N} \sim \text{Wishart}(\mathbf{S}^{-1}, N-1)$$
$$p(\mathbf{Q} \,|\, \mathbf{x}_{1:N}) \propto |\mathbf{Q}|^{\frac{N-d-2}{2}} \cdot \mathrm{e}^{-\frac{1}{2}\operatorname{tr}\mathbf{QS}}$$

The mode of the joint posterior $\mathbf{m}, \mathbf{Q} \,|\, \mathbf{x}_{1:N}$ (the maximum a-posteriori estimate, MAP estimate) is $(\hat{\mathbf{m}}, \hat{\mathbf{Q}})$ with

$$\hat{\mathbf{m}} = \mathbf{c}, \quad \hat{\mathbf{Q}} = (N-d-1)\mathbf{S}^{-1} \tag{3}$$

The region inside the density contour that contains α of the probability (the highest density region, α-HDR) of the Normal data model with the MAP estimate is

$$(\mathbf{x}-\mathbf{c})'\hat{\mathbf{Q}}(\mathbf{x}-\mathbf{c}) \le C(\alpha; d) \tag{4}$$

where $C(\,\cdot\,; d)$ is the inverse cumulative distribution function of a chi-square distribution with d degrees of freedom.

The ellipse defined by (1–4) with $d = 2$ can be used as a model of the reception region in the plane. The computational load is very small; most of the work is the calculation of the empirical mean and the 2×2 residual sum of squares matrix in (1). However, as the following examples illustrate, outliers in the data usually produce an ellipse that is too large and often with its major axis incorrectly oriented with respect to the inliers.

B. Examples

Fig. 2 illustrates the estimation of a region reception for actual location fingerpring data that was obtained by logging a heterogenous set of GPS-equipped mobile phones in normal use. Evidently, the Normal data model's ellipse is strongly influenced by the three outliers.

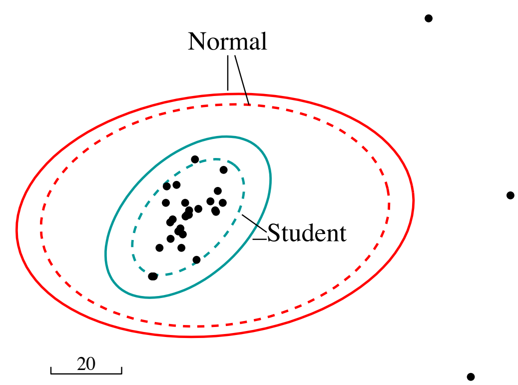

Fig. 2. Reception regions fitted as 90% HDR ellipses for a set of 30 location fingerprints that includes 3 obvious outliers. The solid lines are ellipses fitted using the uninformative prior (2), the dashed lines are fitted using the informative prior (13) with $\sigma = 1000$ and $\tau = 10$.

The ellipse-fitting methods presented here are applicable not only to location fingerprints; they can be applied to any set of multivariate real-valued data for which location and scale are to be estimated. Fig. 3 shows ellipses fitted to an astronomy data set from [4] that is widely used to test clustering and robust regression algorithms. Here also, the Normal model produces an ellipse that is too large and incorrectly oriented with respect to the inliers.

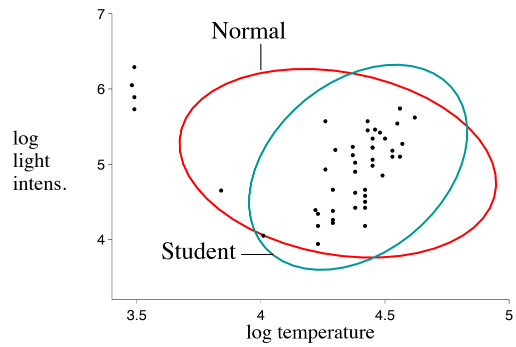

Fig. 3. 90% HDR ellipses fitted to astronomy data

III. Student model

A. Distributions

Here, each observation is modelled as a d-variate Student random vector $\mathbf{x}_n$ with location $\mathbf{m}$, shape $\mathbf{Q}$, and ν degrees of freedom:

$$\mathbf{x}_n \mid \mathbf{m}, \mathbf{Q} \sim \text{Student}(\mathbf{m}, \mathbf{Q}, \nu)$$
$$p(\mathbf{x}_n \mid \mathbf{m}, \mathbf{Q}) \propto |\mathbf{Q}|^{\frac{1}{2}} \left(1 + \tfrac{1}{\nu}(\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m})\right)^{-\frac{\nu+d}{2}}$$

This distribution can be obtained as a mixture of Normals (Fig. 4) by marginalising an auxiliary weight parameter w_n having the prior distribution

$$w_n \sim \text{Gamma}(\tfrac{\nu}{2}, \tfrac{\nu}{2}), \quad p(w_n) \propto w_n^{\frac{\nu}{2}-1} \mathrm{e}^{-\frac{\nu}{2} w_n}$$

out of

$$\mathbf{x}_n \mid \mathbf{m}, \mathbf{Q}, w_n \sim \text{Normal}(\mathbf{m}, (w_n \mathbf{Q})^{-1})$$

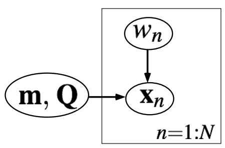

Fig. 4. Directed acyclic graph representation of the Student data model as a mixture of Normals

Assuming that N observations are conditionally independent given the model parameters, the likelihood density is

$$\begin{aligned} p(\mathbf{x}_{1:N} \mid \mathbf{m}, \mathbf{Q}, w_{1:N}) &= \prod_{n=1}^{N} p(\mathbf{x}_n \mid \mathbf{m}, \mathbf{Q}, w_n) \\ &\propto |\mathbf{Q}|^{\frac{N}{2}} \cdot \prod_{n=1}^{N} w_n^{d/2} \cdot \mathrm{e}^{-\frac{1}{2} \sum_{n=1}^{N} w_n (\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m})} \\ &= |\mathbf{Q}|^{\frac{N}{2}} \cdot \prod_{n=1}^{N} w_n^{d/2} \cdot \mathrm{e}^{-\frac{N\bar{w}}{2}(\mathbf{m}-\mathbf{c})'\mathbf{Q}(\mathbf{m}-\mathbf{c}) - \frac{1}{2} \operatorname{tr} \mathbf{QS}} \end{aligned} \tag{5}$$

where $\bar{w} = \frac{1}{N}\sum_{n=1}^{N} w_n$ and the *weighted* empirical mean and residual sum of squares are

$$\mathbf{c} = \frac{\sum_{n=1}^{N} w_n \mathbf{x}_n}{\sum_{n=1}^{N} w_n}, \quad \mathbf{S} = \sum_{n=1}^{N} w_n (\mathbf{x}_n - \mathbf{c})(\mathbf{x}_n - \mathbf{c})' \tag{6}$$

respectively. See Appendix II for a derivation.

Assuming that $w_1, \ldots, w_N, \mathbf{m}, \mathbf{Q}$ are a-priori jointly independent, with the uninformative prior (2), leads to the posterior density

$$\begin{aligned} &p(\mathbf{m}, \mathbf{Q}, w_{1:N} \mid \mathbf{x}_{1:N}) \\ &\quad \propto p(\mathbf{x}_{1:N} \mid \mathbf{m}, \mathbf{Q}, w_{1:N}) p(\mathbf{m}, \mathbf{Q}) p(w_{1:N}) \\ &\quad \propto |\mathbf{Q}|^{\frac{N-d-1}{2}} \cdot \prod_{n=1}^{N} w_n^{\frac{d+\nu}{2}-1} \cdot \mathrm{e}^{-\frac{1}{2} \sum_{n=1}^{N} w_n \left(\nu + (\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m})\right)} \end{aligned}$$

The posterior conditional weights are conditionally independent Gamma-distributed:

$$\begin{aligned} &w_n \mid \mathbf{x}_{1:N}, \mathbf{m}, \mathbf{Q} \sim \\ &\qquad \text{Gamma}\left(\frac{d+\nu}{2}, \frac{\nu + (\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m})}{2}\right) \end{aligned} \tag{7}$$
$$p(w_{1:N} \mid \mathbf{x}_{1:N}, \mathbf{m}, \mathbf{Q}) = \prod_{n=1}^{N} p(w_n \mid \mathbf{x}_{1:N}, \mathbf{m}, \mathbf{Q})$$

with posterior conditional means

$$\mathsf{E}(w_n \mid \mathbf{x}_{1:N}, \mathbf{m}, \mathbf{Q}) = \frac{d+\nu}{\nu + (\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m})} \tag{8}$$

The posterior conditional location and shape parameters are Normal-Wishart distributed:

$$\mathbf{m} \mid \mathbf{x}_{1:N}, \mathbf{Q}, w_{1:N} \sim \text{Normal}(\mathbf{c}, (N\bar{w}\mathbf{Q})^{-1}) \tag{9}$$
$$\mathbf{Q} \mid \mathbf{x}_{1:N}, w_{1:N} \sim \text{Wishart}(\mathbf{S}^{-1}, N-1) \tag{10}$$

The mode of the joint posterior $\mathbf{m}, \mathbf{Q} \mid \mathbf{x}_{1:N}$ is $(\hat{\mathbf{m}}, \hat{\mathbf{Q}})$ with

$$\hat{\mathbf{m}} = \mathbf{c}, \quad \hat{\mathbf{Q}} = (N-d-1)\mathbf{S}^{-1} \tag{11}$$

The α-HDR of the Student data model with the MAP estimate is

$$(\mathbf{x} - \hat{\mathbf{m}})'\hat{\mathbf{Q}}(\mathbf{x} - \hat{\mathbf{m}}) \le d \cdot F(\alpha; d, \nu) \tag{12}$$

where $F(\,\cdot\,; d, \nu)$ is the inverse cumulative distribution function of an F distribution with (d, ν) degrees of freedom.

B. EM algorithm

The elements of the MAP estimate can be computed using the Expectation-Maximisation (EM) algorithm [6]. For the Student data model presented here, the algorithm is

1. initialize $w_{1:N} \leftarrow$ ones
2. for t from 1 to T do
3. $\quad \mathbf{c} \leftarrow (\sum_n w_n \mathbf{x}_n)/(\sum_n w_n)$
4. $\quad \mathbf{S} \leftarrow \sum_n w_n (\mathbf{x}_n - \mathbf{c})(\mathbf{x}_n - \mathbf{c})'$
5. $\quad \hat{\mathbf{Q}} \leftarrow (N-d-1)\mathbf{S}^{-1}$
6. $\quad \hat{\mathbf{m}} \leftarrow \mathbf{c}$
7. $\quad$ for n from 1 to N do
8. $\qquad w_n \leftarrow \frac{d+\nu}{\nu + (\mathbf{x}_n - \hat{\mathbf{m}})'\hat{\mathbf{Q}}(\mathbf{x}_n - \hat{\mathbf{m}})}$
9. $\quad$ end do
10. end do

Here, the EM algorithm's E-step (expectation of the auxiliary parameters) in lines 7–9 uses (8). The M-step (maximisation of the parameters) in lines 5–6 uses (11).

C. Gibbs sampling algorithm

The posterior mean can be used in place of the MAP estimate. In [5] it is shown how the posterior means of the location and shape parameters of a univariate Student data model can be computed using a Gibbs sampling algorithm. Generalised to d dimensions, this algorithm is:

initialize $w_{1:N} \leftarrow$ ones
for t from 1 to $T + T_0$ do
$\quad \mathbf{c} \leftarrow (\sum_n w_n \mathbf{x}_n)/(\sum_n w_n)$

4. $\mathbf{S} \leftarrow \sum_n w_n (\mathbf{x}_n - \mathbf{c})(\mathbf{x}_n - \mathbf{c})'$
5. draw $\mathbf{Q}^{(t)}$ from Wishart$(\mathbf{S}^{-1}, N-1)$
6. draw $\mathbf{m}^{(t)}$ from Normal$(\mathbf{c}, (\sum_n w_n \mathbf{Q}^{(t)})^{-1})$
7. for n from 1 to N do
8. draw w_n from Gamma$\left(\frac{d+\nu}{2}, \frac{\nu + (\mathbf{x}_n - \mathbf{m}^{(t)})' \mathbf{Q}^{(t)} (\mathbf{x}_n - \mathbf{m}^{(t)})}{2}\right)$
9. end do
10. end do

where lines 5–6 make use of (9–10) and lines 7–9 make use of (7). The posterior means of the parameters are estimated by the empirical sample means

$$\mathsf{E}(\mathbf{m} \,|\, \mathbf{x}_{1:N}) \approx \frac{1}{T} \sum_{t=1}^{T} \mathbf{m}^{(t+T_0)}, \;\; \mathsf{E}(\mathbf{Q} \,|\, \mathbf{x}_{1:N}) \approx \frac{1}{T} \sum_{t=1}^{T} \mathbf{Q}^{(t+T_0)}$$

in which the first T_0 "burn-in" samples are discarded.

D. Examples

The 90% HDR ellipse of the Student data model with MAP estimate of location fingerprint data is shown in Fig. 2. This ellipse is evidently a much better fit to the inliers than the ellipse based on a Normal data model; in particular, the ellipse major axis is correctly aligned with respect to the inliers. I used a Student distribution with $\nu = 5$ degrees of freedom because this is a widely recommended choice for robust estimation. Ellipses fitted with different values $\nu \in \{3, \ldots, 10\}$ were similar in appearance.

The EM algorithm converged in about $T = 10$ steps for this example, and in my Matlab implementation the computation takes 3 ms, about 30 times slower than with the Normal data model.

The posterior mean estimates of location and shape for the data in Fig. 2 computed with the Gibbs sampling algorithm with $T = 30$ and $T_0 = 3$ gave essentially the same ellipse as the MAP estimate computed with the EM algorithm. My Matlab implementation of the Gibbs sampling algorithm is about 10 times slower than EM.

The Student-fitted ellipse is better than Normal-fitted ellipse for the astronomy data as well (Fig. 3).

IV. Informative prior

For simplicity, in the previous sections the reception region is estimated using the uninformative prior distribution (2) for the ellipse parameters. The reception region estimates could be improved by exploiting prior knowledge about reception region features. The improvement can be expected to be most noticeable in cases where the number of observations is small, say, $N < 10$.

In reception region estimation, one usually has some prior knowledge about the typical range of the radio signal. For example, WiFi reception range is typically at most 90 m outdoors. In the statistical models discussed here, such knowledge can be exploited by using a conjugate prior [1]. A convenient prior for $\mathbf{Q}$ is

$$\mathbf{Q} \sim \text{Wishart}(\frac{1}{\sigma}\mathbf{I}, \tau) \tag{13}$$

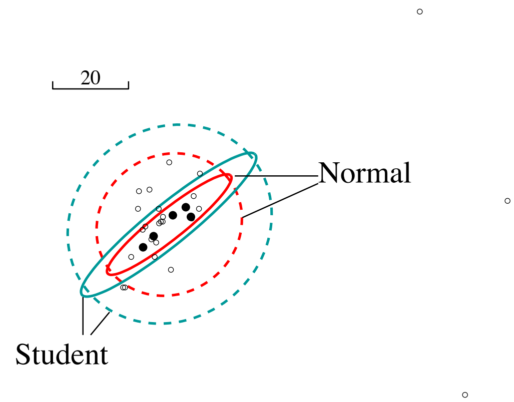

Fig. 5. Reception regions fitted to a set of 5 location fingerprints (filled dots) with no outliers. The open dots are location fingerprints that are not used in the region fitting. Solid lines are for uninformative prior, dashed lines are for informative prior.

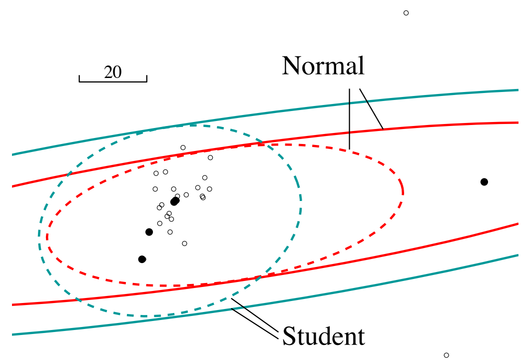

Fig. 6. Reception regions for a set of 5 location fingerprints (filled dots) with 1 obvious outlier. Solid lines are for uninformative prior, dashed lines are for informative prior.

The prior's parameters can be chosen such that τ reflects the strength of the prior knowledge (the "equivalent" number of observations), and $\sigma = \tau R^2$, where R is a "typical" reception range. The uninformative prior (2) corresponds to (13) with $\tau = 0$ and $\sigma \to 0$.

When the prior (13) is used in place of the uninformative prior (2), the formulas presented in the previous sections need to be modified as follows:

(1), (4–9), (12): no change

(3): $\hat{\mathbf{Q}} = \mathsf{E}(\mathbf{Q} \,|\, \mathbf{x}_{1:N}) = (N + \tau - 1)(\mathbf{S} + \sigma \mathbf{I})^{-1}$

(10): $\mathbf{Q} \,|\, \mathbf{x}_{1:N}, w_{1:N} \sim \text{Wishart}((\mathbf{S} + \sigma\mathbf{I})^{-1}, N + \tau - 1)$

(11): $\hat{\mathbf{m}} = \mathbf{c}, \quad \hat{\mathbf{Q}} = (N + \tau - d - 1)(\mathbf{S} + \sigma\mathbf{I})^{-1}$

The only change that is needed in the EM algorithm is in line 5, which becomes

$$\hat{\mathbf{Q}} \leftarrow (N + \tau - d - 1)(\mathbf{S} + \sigma\mathbf{I})^{-1}$$

Similarly, the only change needed in the Gibbs sampler algorithm is in line 5, which becomes

$$\text{draw } \mathbf{Q}^{(t)} \text{ from Wishart}((\mathbf{S} + \sigma\mathbf{I})^{-1}, N + \tau - 1)$$

Using the informative prior (13) with $\sigma = 1000$ and $\tau = 10$ to the 30 location fingerprints in Fig. 2 has little effect, because the number of observations is fairly large.

When only 5 observations are used, the informative prior has a strong influence. If none of the observations are outliers, the Normal and Student ellipses are about the same (Fig. 5), and the informative prior helps to produce a rounder reception ellipse when the observations are nearly collinear. If one of the observations is an outlier, the informative prior serves to reduce the size of the fitted ellipses (Fig. 6).

V. Concluding Remarks

It has been shown how fitting reception region ellipses using a Student data model is robust in the sense of insensitivity to outliers in the location fingerprints. The ellipse parameters can be quickly estimated using the EM algorithm. Prior knowledge of "typical" reception ranges can be exploited to improve the estimate when the number of observations is small.

It is expected that using this method to convert fingerprint data into a reception region database, in combination with an appropriate positioning algorithm, will ultimately lead to better positioning accuracy.

In this work, the reception region is assumed to be fixed and outliers are treated as isolated reports with unusually large errors. In practice, large errors can also arise if the reception region actually moves or changes, for instance because of changes made in a mobile phone network's cell layout. Some preliminary work on the problem of inferring "movable" reception regions is presented in [7]; a study of the problem using robust fitting methods is a topic for further research.

Appendix I
Ellipse parametrisations

An ellipse can be characterised by a set of five real-valued parameters. Three possible parametrisations are:

foci and diameter: The interior of an ellipse in the plane is the set of points $\mathbf{x}$ such that

$$\|\mathbf{x}-\mathbf{f}\| + \|\mathbf{x}-\mathbf{g}\| \le s$$

where $\mathbf{f}$ and $\mathbf{g}$ are the foci and $s \ge \|\mathbf{f}-\mathbf{g}\|$ is the major diameter. When $s = \|\mathbf{f}-\mathbf{g}\|$ the point set is a line segment. When $\mathbf{f} = \mathbf{g}$ the point set is a circular disk with diameter s.

centre, radii and orientation: The ellipse centre is $\mathbf{c} = \frac{1}{2}(\mathbf{f}+\mathbf{g})$, the semimajor axis length (or major radius) is $a = \frac{1}{2}s$, the semiminor axis length (or minor radius) is $b = \frac{1}{2}\sqrt{s^2 - \|\mathbf{f}-\mathbf{g}\|^2}$, and the orientation ϕ is the angle between the x-axis and the major axis. The ellipse area is πab and the ellipse boundary is the point set

$$\mathbf{c} + \begin{bmatrix} a\cos\theta\cos\phi - b\sin\theta\sin\phi \\ a\cos\theta\sin\phi + b\sin\theta\cos\phi \end{bmatrix} \quad (0 \le \theta < 2\pi)$$

quadratic form: The ellipse interior is the set of points $\mathbf{x}$ such that

$$(\mathbf{x}-\mathbf{c})'\mathbf{Q}(\mathbf{x}-\mathbf{c}) \le 1$$

where $\mathbf{Q}$ is a symmetric non-negative definite matrix. When $\frac{1}{4}\mathbf{Q}$ is nonsingular, it is the inverse of the covariance matrix of a random variable having a uniform probability density inside the ellipse and zero outside [8]. Formulas for converting between (a, b, ϕ) and $\mathbf{Q}$ are obtained by noting that $\mathbf{Q}$ has the singular value decomposition

$$\begin{bmatrix} -\sin\phi & \cos\phi \\ \cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} b^{-2} & 0 \\ 0 & a^{-2} \end{bmatrix} \begin{bmatrix} -\sin\phi & \cos\phi \\ \cos\phi & \sin\phi \end{bmatrix}$$

Appendix II
Derivation of formula (5)

Denoting $\mathbf{W} = \mathrm{diag}(w_{1:N})$, $\mathbf{1}_N$ a vector of N ones, $\otimes$ the Kronecker product, and $\mathrm{vec}(\,\cdot\,)$ the stacking operator, the quadratic form can be written as

$$\begin{aligned}
&\sum_{n=1}^{N} w_n(\mathbf{x}_n - \mathbf{m})'\mathbf{Q}(\mathbf{x}_n - \mathbf{m}) \\
&\quad= \big(\mathrm{vec}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')\big)' \cdot (\mathbf{W}\otimes\mathbf{Q}) \cdot \mathrm{vec}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N') \\
&\quad= \big(\mathrm{vec}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')\big)' \cdot \mathrm{vec}(\mathbf{Q}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')\mathbf{W}) \\
&\quad= \mathrm{tr}\big((\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')'\mathbf{Q}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')\mathbf{W}\big) \\
&\quad= \mathrm{tr}(\mathbf{Q}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')\mathbf{W}(\mathbf{x}_{1:N} - \mathbf{m}\mathbf{1}_N')') \\
&\quad= \mathrm{tr}\Big(\mathbf{Q}\big(N\bar{w}(\mathbf{m}-\mathbf{c})(\mathbf{m}-\mathbf{c})' + \mathbf{S}\big)\Big) \\
&\quad= N\bar{w}(\mathbf{m}-\mathbf{c})'\mathbf{Q}(\mathbf{m}-\mathbf{c}) + \mathrm{tr}\,\mathbf{QS}
\end{aligned}$$

References

[1] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, Positioning with coverage area estimates generated from location fingerprints, 7th Workshop on Positioning, Navigation and Communication, Dresden, 2010.

[2] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, 1973, republished by Wiley in 1992.

[3] K. L. Lange, R. J. A. Little, J. M. G. Taylor, Robust statistical modeling using the t distribution, *J. American Statistical Assoc.*, **84** (408), 881–896, 1989.

[4] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 2003.

[5] I. Verdinelli and L. Wasserman, Bayesian analysis of outlier problems using the Gibbs sampler, *Statistics and Computing*, **1**, 105–117, 1991.

[6] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Statistical Society, Series B*, **39**, 1–38, 1977.

[7] L. Koski, Positioning with Bayesian coverage area estimates and location fingerprints, Masters thesis, University of Tampere, 2010. `http://tutkielmat.uta.fi/tutkielma.php?id=20461`

[8] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 1984.