Real Time Analysis of Accelerometer Pair’s Observations Based on Maximum Relative Entropy Optimization Satisfying Model Constraints

R. Urniežius, S. Bartkevičius, S. Žebrauskas
Department of Electrical Engineering, Kaunas University of Technology, Studentų str. 48, Lt-51367 Kaunas, Lithuania, phone: +370 37 300253, e-mail: stanislovas.bartkevicius@ktu.lt

Introduction

Maximum relative Entropy (MrE) optimization principles were studied in previous paper [1] where convex optimization was performed incorporating model constraints resulting from one axis differential drive robot’s constraints such as the distance between wheels. MrE Lagrangian was constructed for just one discrete observation, so it had several things still to be considered.

First, uniform prior was selected when applying Bayesian filter for recorded data. While principle of entropy maximization states that the distribution closest to a priori knowledge has to be selected.

Second, resulting entropy multiple integral did not have analytical solution of its antiderivative. When solving this optimization problem had lead to the use of numerical integration.

Third, observation data arrived in high volumes so having numerical iterations influenced method’s performance as soon as the number of observation data became large.

Updating with observation data constraints

What is a priori knowledge when having observation data and knowing the model on how different observation channels relate to each other? The answer is inferred from the knowledge collected so far. First thing which is known states that discrete observation’s estimate must have its expectation at the exact value observed, i.e. its mean has to fall right at the observed value, and it is not being able to become “unobserved” (see work by A.Giffin and A.Caticha for wider discussion regarding this [2]). Dirac delta function was used in previous work, but expectation constraint is used for simplicity here. So we already have four constraints

\[
\int_{-\infty}^{\infty} a_x P(a_x) a_x = a_{x,\text{obs}} = c_{ax},
\]

\[
\int_{-\infty}^{\infty} a_y P(a_y) a_y = a_{y,\text{obs}} = c_{ay},
\]

\[
\int_{-\infty}^{\infty} b_x P(b_x) b_x = b_{x,\text{obs}} = c_{bx},
\]

\[
\int_{-\infty}^{\infty} b_y P(b_y) b_y = b_{y,\text{obs}} = c_{by},
\]

where a pair of two-axis accelerometers (z axis is not taken into account in this work) is represented by letters a and b. Thus \(a_{x,\text{obs}}, a_{y,\text{obs}}\) are observations of \(x, y\) axis of accelerometer a. Consequently, \(b_{x,\text{obs}}, b_{y,\text{obs}}\) are observations of \(x, y\) axis of accelerometer b. The very first entropy optimization is a simultaneous optimization taking into account the mathematical model between all four observation channels.

A priori distribution is calculated using the measure of entropy for the continuous-variate Probability Density Function (PDF):

\[
S[P] = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( P(a_x, a_y, b_x, b_y) \right) \ln(P(a_x, a_y, b_x, b_y)) da_x db_x da_y db_y,
\]

here \(P(a_x), P(a_y), P(b_x), P(b_y)\) as from formulas (1) – (4) are marginal PDFs respectively as follows

\[
P_{ax} = P(a_x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(a_x, a_y, b_x, b_y) da_y db_x db_y, \]

\[
P_{ay} = P(a_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(a_x, a_y, b_x, b_y) da_x db_y db_x, \]

\[
P_{bx} = P(b_x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(a_x, a_y, b_x, b_y) da_x db_x db_y, \]

\[
P_{by} = P(b_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(a_x, a_y, b_x, b_y) da_y db_y db_x.
\]
Then normal distribution’s normalization constraint is again incorporated into final Lagrangian as
\[
\int \int \int \int P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 = 1.
\] (10)

All constraints have been enumerated except those of mathematical model. The mathematical model is necessary to infer the shape of PDF for every channel. In other words, if we have a mathematical model which shows how all four channels relate we can infer which current observation of all four channels fits the model better and which fits worse. Second thing, which has to be considered, is that so far constraints contained the updating of the first probabilistic moment, i.e. the expectation. The second probabilistic moment can be taken from the mathematical model, so it has to be selected with caution.

**Updating with model and observation data constraints**

An experiment was performed with two accelerometers mounted on a rigid body (robot). The main aim was to infer both accelerometers’ x axis zero bias from dynamic real time observation data. Axis y had a very small zero bias so it was neglected in zero bias. So it was decided to perform an autocalibration experiment. During it rigid body was being rotated around static rotation center where all geometrical distances were known (see figure (1)).

![Autocalibration experiment](image)

**Fig. 1.** Autocalibration experiment with two accelerometers mounted on rigid body (robot) and rotating about geometrically known rotation center

In other words, instant center of rotation during the whole autocalibration experiment was kept constant and distances OA, OB and AB geometrical layout were known. Moreover, angular velocity \( w \) was being tried to be kept as constant as possible, so it was known that angular acceleration fluctuations were small and there was an initial honest knowledge that \( w \) did not exceed values \( w_{\text{max}} \) and \( w_{\text{min}} \). Angular velocity and acceleration were derived from rigid body kinematics relations. One can derive two main relationships between all four observation channels
\[
a_1^2 + a_2^2 = k_1 (b_1^2 + b_2^2),
\] (11)
\[
a_1 b_1 + a_2 b_2 = k_2 (b_1^2 + b_2^2).
\] (12)

And this gives two more constraints which represented model constraints, i.e.
\[
\int \int \int (a_1^2 + a_2^2) P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 = k_1 \int \int \int (b_1^2 + b_2^2) P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 = \]
\[
= k_1 \int \int \int (b_1^2 + b_2^2) P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 =
\] (13)
\[
\int \int \int (a_1 b_1 + a_2 b_2) P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 = k_2 \int \int \int (b_1^2 + b_2^2) P(a_1, a_2, b_1, b_2) da_1 da_2 db_1 db_2 =
\] (14)

Constructing Lagrangian using (5) and the constraints (1) – (4), (10), (13), (14) require the definition of Lagrange multipliers \( k_{a_1}, k_{a_2}, k_{b_1}, k_{b_2} \) for constraints (1) – (4), \( \beta \) for constraint (13) and \( \rho \) for constraint (14) respectively. Then it can be proved that the marginal PDFs are
\[
P_{a_1} = \frac{2\pi^2 e^{-\frac{3}{4}c_1^2c_1^2\rho^2}}{c_{\text{norm}}},
\] (15)
\[
P_{a_2} = \frac{2\pi^2 e^{-\frac{3}{4}c_2^2c_2^2\rho^2}}{c_{\text{norm}}},
\] (16)
\[
P_{b_1} = \frac{2\pi^2 e^{-\frac{3}{4}c_3^2c_3^2\rho^2}}{c_{\text{norm}}},
\] (17)
\[
P_{b_2} = \frac{2\pi^2 e^{-\frac{3}{4}c_4^2c_4^2\rho^2}}{c_{\text{norm}}},
\] (18)

where normalization constant \( c_{\text{norm}} \). Lagrange multipliers, and constant \( c_i \) are
\[
c_1 = 4 \beta^2 k_1 + 4 \beta k_1 \rho + \rho^2,
\] (19)
\[
c_2 = c_3^2 + c_4^2,
\] (20)
\[
c_3 = c_1^2 + c_2^2,
\] (21)
\[
c_4 = c_1^2 + c_2^2,
\] (22)
\[
c_{\text{norm}} = \frac{4\pi^2 e^{-\frac{3}{4}c_1^2c_1^2\rho^2}}{c_1},
\] (23)
\[
k_{a_1} = c_{a_1} \left( \begin{array}{c}
-c_2 c_{c_a} c_{c_b} k_1 + c_{b_2} (c_{c_a} - c_{b_2}) k_2 \\
8 c_{a_1} c_{c_b} + 4 k_1 (2 c_{c_a} c_{c_b} + c_{c_a} c_{c_b})
\end{array} \right)/c_2
\] (24)

and Lagrange multipliers \( k_{a_2}, k_{b_1}, k_{b_2} \) are calculated similarly. Coefficients \( k_1 \) and \( k_2 \) depend on geometrical layout of accelerometers on the rigid body, i.e. they are calculated from OA, OB and AB distances.

It can be seen from formulas (15) – (24) that antiderivative function for MrE PDF calculation is found.
So MrE optimization when finding a priori PDF is an O(1) operation, which makes this method practical. An attentive reader would notice that

$$\text{arg max } P_{a_{i}} = \int_{-\infty}^{+\infty} \hat{a}_{i} P(a_{i}) da_{i},$$

$$\text{arg max } P_{a_{i}} = \int_{-\infty}^{+\infty} \hat{a}_{i} P(a_{i}) da_{i},$$

$$\text{arg max } P_{b_{i}} = \int_{-\infty}^{+\infty} \hat{b}_{i} P(b_{i}) db_{i},$$

$$\text{arg max } P_{b_{i}} = \int_{-\infty}^{+\infty} \hat{b}_{i} P(b_{i}) db_{i},$$

and it means that Maximum Likelihood (ML) estimate falls at the observation reading just like expectation does, and it confirms that the approach to treat likelihood included in the a priori PDF was correct indeed, see work [2]. In other words,

$$P_{old}(a_{i}, c_{a_{i}}) = P_{old}(c_{a_{i}}) \cdot P_{old}(a_{i} | c_{a_{i}}),$$

where model constraints (11) and (12) are updated with constraints moments.

**Updating time series distributions with model constraints**

Time series observation data can contain not only a static noise which can be filtered with regular autoregression efforts, but also a dynamic noise which could be a result of external forces like a hit to the wall when accelerometers had an instantaneous peak. Then averaging does not help even if a huge number of observation data is collected. There is one more property of accelerometers which has to be considered. Every axis has its zero bias. It can be observed by turning the axis to geometrical layout distances) are taken into the account:

Important question is whether first or second probabilistic moment will be used for updating with model constraints. After multiple cases analysis it was found that updating with second probabilistic moment gives little benefit when finding an analytical posteriori. And updating with first probabilistic moment would help avoid time consuming antiderivative numerical calculations. After some manipulation with (11) and (12) it is found that these constraints could be rewritten as

$$a_{i} = b_{i} k_{c} \pm b_{i} \sqrt{k_{c} - k_{c}^{2}},$$

$$a_{i} = b_{i} k_{c} \pm b_{i} \sqrt{k_{c} - k_{c}^{2}},$$

which makes them acceptable as model constraints and preserves relatively easy calculation of antiderivative function of the entropy multiple integral (30). However, with n observations there are 4n unknowns and only 2n equations were at hand so far. Then two more model constraints are used which bring more constraints on time series over time domain, i.e.

$$a_{i,j} - a_{i,j+1} = c_{i} + b_{i,j} - b_{i,j+1},$$

$$a_{i,j} - b_{i,j} = c_{i},$$

where constant $c_{i}$ and $c_{j}$ are calculated using cubic spline derivations when three main inputs (in addition to rough geometrical layout distances) are taken into the account: $\Delta t$ which was a time period between two samples, $V_{\text{init}}$ and $V_{\text{final}}$ were maximum and minimum angular velocities during the whole autocalibration experiment according to a priori knowledge.

A raw MrE calculation has a performance of O(n), but the problem is that the signs in formulas (32) and (33) are not known in advance. Because of that local optimization over a certain window has to be run first. The local processing window (LPW) is selected to contain 6 samples, which give an asymptotic notation as O($\log n$).

Unlike particle filters methods [3], this method has the analytical representation of normalization constant. Also it shows how to extract more information from dynamic measurements (see work [4]).

After the baseline of observation data was found satisfying all time-domain model constraints, the unbiased estimator for calculating zero bias can be approximated and found from Normal distribution as

$$a_{i,j} = \sum_{j=1}^{n} (a_{i,j}, c_{i,j}) \cdot n \approx 0.01 g,$$

$$b_{i,j} = \sum_{j=1}^{n} (b_{i,j}, c_{i,j}) \cdot n \approx 0.065 g,$$

where g is Earth’s gravity.
Conclusions

Analytical and iterations-free antiderivative function has been derived for MrE optimization with model constraints. The estimated zero bias values are compared to the values observed from static bias when x axis is being pointed to the direction of Earth’s gravity direction, and it is found to be consistent. Multiple dependent cheap sensors can be used in inference with the same benefit which would be given by a single expensive sensor, when the sufficient number of observation data is available. Aging of a sensor can be monitored from dynamic observation data.

Acknowledgement

We would like to thank A. Giffin from Princeton Institute for the Science and Technology of Materials (PRISM) for constructive discussions.

References