Bayesian Cramér-Rao Lower Bounds for Magnetic Field-based Train Localization

Benjamin Siebler*, Stephan Sand*, and Uwe D. Hanebeck†
*German Aerospace Center (DLR), Institute of Communications and Navigation
†Karlsruhe Institute of Technology (KIT), Intelligent Sensor-Actuator-Systems Laboratory
benjamin.siebler@dlr.de

***Abstract*—In this paper, the theoretically achievable accuracy of magnetic field-based localization in railway environments is analyzed. The analysis is based on the Bayesian Cramér-Rao lower bound (BCRLB) that bounds the mean squared error of an estimator from below. The derivation of the BCRLB for magnetic field-based localization is not straightforward because the magnetic field cannot be described by an analytical equation but must be derived from measurements. In this paper we show how the BCRLB can be calculated by fitting a Gaussian process (GP) to magnetometer measurements to obtain an analytical expression of the magnetic field along a railway line. The proposed GP-based BCRLB is evaluated with the magnetic field of a 1 km long track segment. Furthermore, a comparison between the bound and the estimation error of a particle filter shows the sub-optimality of the particle filter for magnetic railway localization.**

I. Introduction

Localization using the magnetic field is based on distortions in the Earth magnetic field introduced by ferromagnetic material in the environment. In railway environments these distortions are caused by track-side infrastructure, e.g., signaling poles and rails but also by steel reinforced concrete in buildings close to the track. When the ferromagnetic material is at static positions also the distortions are static. A map of the distorted magnetic field therefore enables the estimation of the position of a magnetometer by comparing its measurements to the magnetic map. The feasibility of magnetic field-based localization has already been demonstrated in various environments such as interiors [1], roads [2] and the airspace [3]. For railways we showed the feasibility, e.g. in [4], where we used a particle filter and solely the measurements of a magnetometer. While the particle filter achieved meter-level accuracy, it is still unclear how well the filter performs w.r.t. an optimal filter. In this paper, therefore this question is addressed by deriving a Bayesian Cramér-Rao lower bound (BCRLB) for magnetic localization in railway environments. The presented bound is based on our prior work [5] in which we derived the BCRLB for magnetic localization of a wheeled robot. The derived BCRLB is a lower bound on the mean squared error (MSE) that a filter can achieve. The bound therefore can be utilized to check if a proposed filter is close to optimal with respect to the MSE. Another use case for the BCRLB, when applied to magnetic localization, is that it enables the investigation of the localization accuracy achievable for a given magnetic field without the need of actually implementing and evaluating a specific filter algorithm. Therefore, it is possible to analyze the potential localization accuracy of railway networks and label parts of a network where the accuracy is expected to be poor. Unfortunately, the derivation of the bound for magnetic localization comes with a challenge. In the derivation of the BCRLB it is assumed that an analytical expression of the measurement model and the associated likelihood exists and that derivatives w.r.t. the estimated variable can be calculated. For magnetic localization the measurement model is defined by the magnetic map. The map is a function that maps a position on the track to the corresponding magnetic vector field. Due to the high complexity of the interaction between the Earth magnetic field and the ferromagnetic material close to the track the map cannot be calculated from the laws of physics. Hence, in practice the map is obtained from measurements along the track. To derive the BCRLB it is therefore crucial to find an analytical representation of the magnetic map based on that measurements that can be incorporated into the bound. As proposed first in [6], using a Gaussian process (GP) to represent the measurement model nicely fits into the Bayesian framework of the BCRLB because the GP accounts for uncertainties in the measurements. In [5], we further extended the bound proposed in [6] with the option to incorporate a deterministic control input. Considering a control input in the bound has the benefit that the trajectory of the train on the track, for which the bound is calculated, becomes controllable. In contrast, the standard formulation of the bound assumes a pure stochastic motion model which potentially results in trajectories that are not of interest in practice.

This paper contains a detailed derivation of the equations required to implement the proposed BCRLB for magnetic train localization. After introducing the derivation, the bound is evaluated with a data set recorded in a real railway environment. In addition, the performance of a particle filter is compared to the bound.

II. Bayesian Cramér-Rao Lower Bound

In this section the BCRLB is introduced and it is shown how it can be applied to magnetic field-based train localization. Here only a brief overview is given, for a more detailed explanation we recommend to read [5] and [6].

A. BCRLB for Analytical Likelihood Models

The BCRLB is a lower bound on the MSE matrix $\mathbf{M}$ of an estimator $\hat{\mathbf{x}}(\mathbf{z})$ that estimate the state vector $\mathbf{x}$ based on some observations $\mathbf{z}$. The bound fulfills the inequality

$$\mathbf{M} = \mathrm{E}_{\mathbf{x},\mathbf{z}}\left[(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})(\hat{\mathbf{x}}(\mathbf{z}) - \mathbf{x})^T\right] \geq \mathcal{J}^{-1}. \quad (1)$$

In (1) the bound is the given by the inverse of the matrix $\mathcal{J}$ which is commonly called Bayesian information matrix (BIM) [7]. The BIM is obtained from the joined probability density function (pdf) of $\mathbf{x}$ and $\mathbf{z}$

$$\mathcal{J} = -\,\mathrm{E}_{\mathbf{x},\mathbf{z}}\left[\Delta_{\mathbf{x}}^{\mathbf{x}} \ln\left(p(\mathbf{z},\mathbf{x})\right)\right] \tag{2}$$

with the differential operator $\Delta_{\mathbf{b}}^{\mathbf{a}} = \nabla_{\mathbf{b}} \nabla_{\mathbf{a}}^{T}$. The BIM is closely related to the FIM $\mathcal{F}(\mathbf{x})$

$$\begin{aligned}\mathcal{J} &= \mathrm{E}_{\mathbf{x}}\left[\mathrm{E}_{\mathbf{z}|\mathbf{x}}\left[-\Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{z}|\mathbf{x})\right]\right] + \mathrm{E}_{\mathbf{x}}\left[-\Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{x})\right] \\ &= \mathrm{E}_{\mathbf{x}}\left[\mathcal{F}(\mathbf{x})\right] + \mathrm{E}_{\mathbf{x}}\left[-\Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{x})\right]. \end{aligned} \tag{3}$$

Relation (3) is easily obtained when the joint pdf in (2) is decomposed using the definition of conditional densities. Due to (3) the BIM is the sum of the expected FIM and the information about $\mathbf{x}$ obtained from its prior density $p(\mathbf{x})$. From (3) we also see that the calculation of the BIM requires us to calculate derivatives of the likelihood $p(\mathbf{z}|\mathbf{x})$ w.r.t. $\mathbf{x}$. Unfortunately, in the case of magnetic localization the likelihood contains the map of the magnetic field which is only known from measurements. The likelihood and its derivatives are therefore analytically unknown but in the next section an approach is introduced that enables the differentiation of the likelihood in such cases.

B. BCRLB for Data-based Models

As proposed in [6] the FIM and the BIM can be also calculated when the likelihood contains parts which are known only from a set of observations. In the following measurement models of the form

$$z_k = \tilde{\mathrm{h}}\left(\mathbf{x}_k\right) + n_k \tag{4}$$

with additive white Gaussian measurement noise are considered. Where $\tilde{\mathrm{h}}\left(\mathbf{x}_k\right)$ is the true but unknown measurement function and $n_k \sim \mathcal{N}(0, \sigma_n^2)$ Gaussian noise. Note, for now the measurement z_k is a scalar, the extension to vector measurements depends on the application. For magnetic train localization the extension is straightforward and will be introduced in Section II-D. For models that can be described in the form of (4) the likelihood is

$$p(z_k|\mathbf{x}_k) = \mathcal{N}(z_k; \tilde{\mathrm{h}}\left(\mathbf{x}_k\right), \sigma_n^2). \tag{5}$$

Since the true function $\tilde{\mathrm{h}}\left(\mathbf{x}_k\right)$ is unknown and only (noisy) observations of its values are available it has to be replaced with an approximation $\mathrm{h}\left(\mathbf{x}_k\right)$. In principle many possible function approximation schemes exist. Here the approximation should be chosen such that the derivative of the likelihood exists. As we will see later, this implies that the function approximation $\mathrm{h}\left(\mathbf{x}_k\right)$ has to be smooth in the domain of interest. Furthermore, the approximation should account for uncertainty in the observations and should fit nicely into the Bayesian nature of the BCRLB. As first proposed in [6], GPs fulfill these requirements and are therefore used in the following.

Since the approximation of the measurement function depends on the observations, the BCRLB has to be modified to account for this dependency. In the following the observations z and the values of $\mathbf{x}$ for which they were observed are combined into a training data set $\mathcal{D} = \{\mathbf{x}_i, \mathbf{z}_i\}_i^{N_\mathcal{D}}$. Thus, the set contains $N_\mathcal{D}$ pairs of input values $\mathbf{x}_i$ and measured output values z_i of the true measurement function $\tilde{\mathrm{h}}\left(\mathbf{x}\right)$.

In addition, also a set of deterministic control inputs $\mathcal{U}$ is assumed to be given that forms together with the training data the set $\mathcal{D}_u = \{\mathcal{D}, \mathcal{U}\}$. As shown in [5] this results in the conditional BIM

$$\mathcal{J} = -\,\mathrm{E}_{\mathbf{x},\mathbf{z}|\mathcal{D}_u}\left[\Delta_{\mathbf{x}}^{\mathbf{x}} \ln\left(p(\mathbf{z},\mathbf{x}|\mathcal{D}_u)\right)\right]. \tag{6}$$

The comparison of (2) and (6) shows that accounting for this dependency is simply achieved by replacing the joint pdf in (2) with the joint pdf conditioned on the training and input set. As before in (3) the BIM can be decomposed into a data dependent part and a part that depends on the prior knowledge about $\mathbf{x}$

$$\mathcal{J} = \mathrm{E}_{\mathbf{x}|\mathcal{D}_u}\left[\mathcal{F}(\mathbf{x})\right] + \mathrm{E}_{\mathbf{x}|\mathcal{D}_u}\left[-\Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{x}|\mathcal{D}_u)\right]. \tag{7}$$

where the FIM is now

$$\mathcal{F}(\mathbf{x}) = \mathrm{E}_{\mathbf{z}|\mathcal{D}_u,\mathbf{x}}\left[-\Delta_{\mathbf{x}}^{\mathbf{x}} \ln p(\mathbf{z}|\mathbf{x},\mathcal{D}_u)\right]. \tag{8}$$

The FIM is also conditioned on the set $\mathcal{D}_u$ and therefore will vary depending on the given training data. Thus, changes in the training data would also change the value of the bound. Fortunately, if the training data contains a "sufficient" amount of observations to accurately approximate the true function in (4) it is to be expected that the bound is consistent across different data sets.

1) Short Introduction to Gaussian Processes: When the true measurement function $\tilde{\mathrm{h}}\left(\mathbf{x}_k\right)$ at a certain input $\mathbf{x}_k$ is approximated with an GP the output of the approximated function $\mathrm{h}\left(\mathbf{x}_k\right)$ itself is a Gaussian distributed random variable. The prior of the GP, before any observations are available, is fully specified by its mean function $\mathrm{m}\left(\mathbf{x}\right)$ and covariance function $\mathrm{k}\left(\mathbf{x},\mathbf{x}'\right)$ [8]

$$\begin{aligned}\mathrm{m}\left(\mathbf{x}\right) &= \mathrm{E}\left[\mathrm{h}\left(\mathbf{x}\right)\right] \\ \mathrm{k}\left(\mathbf{x},\mathbf{x}'\right) &= \mathrm{E}\left[\left(\mathrm{h}\left(\mathbf{x}\right) - \mathrm{m}\left(\mathbf{x}\right)\right)\left(\mathrm{h}\left(\mathbf{x}'\right) - \mathrm{m}\left(\mathbf{x}'\right)\right)\right]. \end{aligned} \tag{9}$$

For a GP performing regression at a certain input $\mathbf{x}_k$ amounts to calculating the posterior Gaussian distribution of the function value conditioned on the input and the training data

$$p(\mathrm{h}\left(\mathbf{x}_k\right)|\mathbf{x}_k, \mathcal{D}_u) = \mathcal{N}\left(\mathrm{h}\left(\mathbf{x}_k\right); \mu(\mathbf{x}_k, \mathcal{D}_u), r(\mathbf{x}_k, \mathcal{D}_u)\right), \tag{10}$$

where the posterior mean and variance are given by

$$\begin{aligned}\mu(\mathbf{x}_k, \mathcal{D}) &= \mathrm{m}\left(\mathbf{x}_k\right) \\ &\quad + \mathbf{K}_{\mathbf{x}_k\mathcal{D}}\left[\mathbf{K}_{\mathcal{D}\mathcal{D}} + \sigma_\mathcal{D}^2\mathbf{I}\right]^{-1}\left(\mathbf{Z}_\mathcal{D} - \mathrm{m}\left(\mathbf{X}_\mathcal{D}\right)\right) \\ r(\mathbf{x}_k, \mathcal{D}) &= \mathbf{K}_{\mathbf{x}_k,\mathbf{x}_k} - \mathbf{K}_{\mathbf{x}_k\mathcal{D}}\left[\mathbf{K}_{\mathcal{D}\mathcal{D}} + \sigma_\mathcal{D}^2\mathbf{I}\right]^{-1}\mathbf{K}_{\mathcal{D}\mathbf{x}_k}. \end{aligned} \tag{11}$$

$\mathbf{X}_\mathcal{D}$ is here a column vector in which all training inputs are stacked on top of each other, $\mathrm{m}\left(\mathbf{X}_\mathcal{D}\right)$ is the mean function evaluated at all training inputs, and $\mathbf{Z}_\mathcal{D}$ is a vector of the corresponding stacked observations. The matrices $\mathbf{K}_{\mathcal{D}\mathcal{D}}$ and $\mathbf{K}_{\mathbf{x}_k\mathcal{D}}$ in (11) are the short form for $\mathbf{K}_{\mathcal{D}\mathcal{D}} = \mathbf{K}(\mathbf{X}_\mathcal{D}, \mathbf{X}_\mathcal{D})$

and $\mathbf{K}_{\mathbf{x}_k\mathcal{D}} = \mathbf{K}(\mathbf{x}_k, \mathbf{X}_\mathcal{D})$ where $\mathbf{K}(\mathbf{A}, \mathbf{B})$ for two sets $\mathbf{A}$ and $\mathbf{B}$ of input points has the form

$$\mathbf{K}(\mathbf{A},\mathbf{B})=\begin{bmatrix} \mathrm{k}\left(\{\mathbf{A}\}_1,\{\mathbf{B}\}_1\right) & \cdots & \mathrm{k}\left(\{\mathbf{A}\}_1,\{\mathbf{B}\}_{|\mathbf{B}|}\right) \\ \vdots & \ddots & \vdots \\ \mathrm{k}\left(\{\mathbf{A}\}_{|\mathbf{A}|},\{\mathbf{B}\}_1\right) & \cdots & \mathrm{k}\left(\{\mathbf{A}\}_{|\mathbf{A}|},\{\mathbf{B}\}_{|\mathbf{B}|}\right) \end{bmatrix}.$$

Since $\mathrm{k}(\cdot,\cdot)$ describes the covariance of two inputs, the matrix $\mathbf{K}_{\mathcal{D}\mathcal{D}}$ is the covariance between the training inputs with themselves and $\mathbf{K}_{\mathbf{x}_k\mathcal{D}}$ is a row vector of the covariances between the input $\mathbf{x}_k$, at which the GP posterior is evaluated, with the training inputs. Note, here the GP is evaluated for one input value and hence the output is scalar.

Throughout this paper the covariance function is the well known squared exponential kernel

$$\mathrm{k}\left(\mathbf{a},\mathbf{b}\right) = \sigma_\mathrm{k}^2 \exp\left(-\frac{(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b})}{2l^2}\right) \tag{12}$$

that was shown to be suitable for the approximation of the magnetic field in [9] and [10]. The mean function is set to a constant

$$\mathrm{m}\left(\mathbf{x}_k\right) = \bar{m}. \tag{13}$$

2) FIM for GP-based Measurement Models: From the posterior distribution of the GP and the measurement model (4) the likelihood $p(z_k|\mathbf{x}_k, \mathcal{D}_u)$ becomes

$$\begin{aligned} p(z_k|\mathcal{D}_u,\mathbf{x}_k) &= \mathcal{N}\left(z_k;\mu(\mathbf{x}_k,\mathcal{D}), r(\mathbf{x}_k,\mathcal{D})+\sigma_n^2\right) \\ &= \mathcal{N}\left(z_k;\mu(\mathbf{x}_k,\mathcal{D}), r_\mathrm{C}(\mathbf{x}_k,\mathcal{D})\right). \end{aligned} \tag{14}$$

With the FIM for general Gaussian likelihoods given in [11] and using the notation from (14) the element in the i-th row and j-th column of the FIM (8) becomes

$$\begin{aligned} [\mathcal{F}(\mathbf{x}_k)]_{ij} = &\frac{\partial\mu(\mathbf{x}_k,\mathcal{D})}{\partial[\mathbf{x}_k]_i} r_\mathrm{C}^{-1}(\mathbf{x}_k,\mathcal{D}) \frac{\partial\mu(\mathbf{x}_k,\mathcal{D})}{\partial[\mathbf{x}_k]_j} \\ &+ \frac{1}{2} r_\mathrm{C}^{-2}(\mathbf{x}_k,\mathcal{D}) \frac{\partial r_\mathrm{C}(\mathbf{x}_k,\mathcal{D})}{\partial[\mathbf{x}_k]_i} \frac{\partial r_\mathrm{C}(\mathbf{x}_k,\mathcal{D})}{\partial[\mathbf{x}_k]_j}. \end{aligned} \tag{15}$$

This shows that for the FIM and hence for the BCRLB the derivative of the posterior mean and covariance of the GP is required. Fortunately, these derivatives exist as long as the chosen mean and covariance functions are smooth.

C. BCRLB for Nonlinear Filtering

In the previous sections the BCRLB was introduced and it was shown how the FIM can be calculated when the true measurement model is approximated with a GP. In this section the bound is now extended to dynamic estimation problems in which a motion model is available to describe how the state vector $\mathbf{x}$ that is to be estimated evolves over time. In particular the BCRLB for nonlinear filtering is investigated. The filtering bound is of interest because it is the lower bound for all localization filters and therefore constitutes a benchmark for them.

To obtain the filtering BCRLB the naive approach would be to calculate the BIM over the complete state sequence $\mathbf{X}_{0:k+1} = \left[\mathbf{x}_0^T \cdots \mathbf{x}_{k+1}^T\right]^T$. For short sequences and low dimensional state spaces this might work but becomes quickly computational complex since the dimensions of the involved matrices keep growing with the length of the sequence. This increase in the complexity is avoided with the recursive formulation of the bound introduced in [12]. In the next paragraph we only present the recursive equations without derivations. A detailed derivation of the generic recursive bound can be found in [12]. Additional information about the derivation for GP-based measurement models can be found in [5] and [6].

In the recursive formulation the BIM $\mathcal{J}_{k+1}$ for the state $\mathbf{x}_{k+1}$ at time step $k+1$ is calculated based on the BIM of time step k

$$\mathcal{J}_{k+1} = \mathbf{D}_k^{22} - \left(\mathbf{D}_k^{12}\right)^T \left(\mathbf{D}_k^{11} + \mathcal{J}_k\right)^{-1} \mathbf{D}_k^{12} \tag{16}$$

where $\mathcal{J}_{k+1}$ is a square matrix with the dimension of the state vector $\mathbf{x}$. The remaining matrices in (16) are

$$\begin{aligned} \mathbf{D}_k^{11} &= -\mathrm{E}_{\mathbf{x}_{k:k+1}|\mathcal{D}_u}\left[\Delta_{\mathbf{x}_k}^{\mathbf{x}_k} \ln p(\mathbf{x}_{k+1}|\mathcal{D}_u,\mathbf{x}_k)\right] \\ \mathbf{D}_k^{12} &= -\mathrm{E}_{\mathbf{x}_{k:k+1}|\mathcal{D}_u}\left[\Delta_{\mathbf{x}_k}^{\mathbf{x}_{k+1}} \ln p(\mathbf{x}_{k+1}|\mathcal{D}_u,\mathbf{x}_k)\right] \\ \mathbf{D}_k^{22} &= -\mathrm{E}_{\mathbf{x}_{k+1},\mathbf{z}_{k+1}|\mathcal{D}_u}\left[\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{z}_{k+1}|\mathcal{D}_u,\mathbf{x}_{k+1})\right] \\ &\quad -\mathrm{E}_{\mathbf{x}_{k:k+1}|\mathcal{D}_u}\left[\Delta_{\mathbf{x}_{k+1}}^{\mathbf{x}_{k+1}} \ln p(\mathbf{x}_{k+1}|\mathcal{D}_u,\mathbf{x}_k)\right]. \end{aligned} \tag{17}$$

D. BCRLB Implementation for Train Localization

With the BCRLB equations (16) and (17) for nonlinear filtering all parts are in place to implement the bound for magnetic field-based localization systems. For the implementation we have to define the state space model consisting of a motion model that describes the evolution of the state over time and a measurement model that links the observations to the state variables (or at least some of the variables).

1) Motion Model: For train localization the state vector is

$$\mathbf{x}_k = \begin{bmatrix} s_k & \dot{s}_k \end{bmatrix}^T \tag{18}$$

where s is the 1D along-track position of the train and $\dot{s}$ its time derivative, i.e, the train speed. The motion model is a continuous white noise acceleration model [13] with an additional deterministic acceleration input a_k

$$\begin{aligned} \mathbf{x}_k &= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} \frac{1}{2}T^2 \\ T \end{bmatrix} a_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{x}_k &= \mathbf{F}\mathbf{x}_{k-1} + \mathbf{B}a_{k-1} + \mathbf{w}_{k-1} \end{aligned} \tag{19}$$

where T is the sampling period. The process noise is Gaussian $\mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and the covariance matrix is

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T \end{bmatrix} \sigma_\mathbf{w}^2. \tag{20}$$

The process noise accounts for imperfections in the actuators, e.g., the drive chain of the train, that cannot perfectly translate the deterministic input into motion.

2) Measurement Model: The measurement model for the magnetometer used in the bound is given by

$$\mathbf{z}_k = \mathbf{h}\left(s_k,\mathcal{D}\right) + \mathbf{n}_k, \tag{21}$$

with the magnetic map

$$\mathbf{h}\left(s_k,\mathcal{D}\right) = \begin{bmatrix} \mathrm{h}_x(s_k,\mathcal{D}) & \mathrm{h}_y(s_k,\mathcal{D}) & \mathrm{h}_z(s_k,\mathcal{D}) \end{bmatrix}^T. \tag{22}$$

The map has three components $\mathrm{h}_i(s_k,\mathcal{D})$, $i \in \{x,y,z\}$ which are the GPs describing the magnetic vector field

along the track. The measurement noise is considered Gaussian $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_n})$ with the diagonal convariance matrix $\mathbf{\Sigma_n} = \mathrm{diag}(\begin{bmatrix}\sigma_{n_x}^2 & \sigma_{n_y}^2 & \sigma_{n_z}^2\end{bmatrix})$. Note, the measurements depend only on the position state s_k and the training data set but not the train speed. In the measurement model we assume that the magnetometer is calibrated and its axes are aligned with the map. For details about sensor calibration and the sensor alignment for train mounted magnetometers we recommend reading our prior work [14]. From (21) we see that the likelihood of the measurement $\mathbf{z}_k$ is

$$p(\mathbf{z}_k|\mathcal{D}_u, \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \boldsymbol{\mu}_\mathbf{z}(s_k, \mathcal{D}), \mathbf{\Sigma}_\mathbf{z}(s_k, \mathcal{D}) + \mathbf{\Sigma_n}).$$

The mean of the likelihood is a vector containing the posterior GP means

$$\boldsymbol{\mu}_\mathbf{z}(s_k, \mathcal{D}) = \begin{bmatrix}\mu_x(s_k, \mathcal{D}) & \mu_y(s_k, \mathcal{D}) & \mu_z(s_k, \mathcal{D})\end{bmatrix}^T. \quad (23)$$

Thus, the covariance matrix $\mathbf{\Sigma_z}$ is a diagonal matrix where the elements are the posterior GP variances

$$\mathbf{\Sigma}_\mathbf{z}(s_k, \mathcal{D}) = \begin{bmatrix} r_x(s_k, \mathcal{D}) & 0 & 0 \\ 0 & r_y(s_k, \mathcal{D}) & 0 \\ 0 & 0 & r_z(s_k, \mathcal{D}) \end{bmatrix}. \quad (24)$$

Due to the diagonal structure of the matrices $\mathbf{\Sigma_z}$ and $\mathbf{\Sigma}_c$ the likelihood can be written as the product of the likelihoods of the single magnetometer axes

$$p(\mathbf{z}_k|\mathcal{D}_u, \mathbf{x}_k) = \prod_{i\in\{x,y,z\}} \mathcal{N}(z_{i,k}; \mu_i(s_k, \mathcal{D}), \underbrace{r_i(s_k, \mathcal{D}) + \sigma_{n_i}^2}_{r_{i,\mathrm{C}}(s_k,\mathcal{D})}). \quad (25)$$

With the likelihood (25) the concrete equations for the FIM can be found by observing that due to the logarithm in (8) the FIM is just the sum over the FIM of the individual sensor axes. From (15) we get the FIM of sensor axis i

$$\begin{aligned}[\mathcal{F}_i(\mathbf{x}_k)]_{11} &= \frac{\partial \mu_i(s_k, \mathcal{D})}{\partial s_k} r_{i,\mathrm{C}}^{-1}(s_k, \mathcal{D}) \frac{\partial \mu_i(s_k, \mathcal{D})}{\partial s_k} \\ &\quad + \frac{1}{2} r_{i,\mathrm{C}}^{-2}(s_k, \mathcal{D}) \frac{\partial r_{i,\mathrm{C}}(s_k, \mathcal{D})}{\partial s_k} \frac{\partial r_{i,\mathrm{C}}(s_k, \mathcal{D})}{\partial s_k} \\ [\mathcal{F}_i(\mathbf{x}_k)]_{12} &= [\mathcal{F}_i(\mathbf{x}_k)]_{21} = [\mathcal{F}_i(\mathbf{x}_k)]_{22} = 0. \end{aligned} \quad (26)$$

Therefore, the overall FIM becomes

$$\mathcal{F}(\mathbf{x}_k) = \sum_{i\in\{x,y,z\}} \mathcal{F}_i(\mathbf{x}_k). \quad (27)$$

To obtain the concrete equations for the implementation the posterior mean and variance from (11) has to be differentiated w.r.t. the components $[\mathbf{x}_k]_i$ of the state vector $\mathbf{x}$

$$\begin{aligned}\frac{\partial \mu(\mathbf{x}_k, \mathcal{D})}{\partial [\mathbf{x}_k]_i} &= \frac{\partial}{\partial [\mathbf{x}_k]_i} \mathrm{m}(\mathbf{x}_k) + \sum_{l=1}^{M} [\mathbf{S}^{-1}\tilde{\mathbf{z}}]_l \frac{\partial \mathrm{k}(\mathbf{x}_k, \{\mathbf{X}_\mathcal{D}\}_l)}{\partial [\mathbf{x}_k]_i} \\ \frac{\partial r_\mathrm{C}(\mathbf{x}_k, \mathcal{D})}{\partial [\mathbf{x}_k]_i} &= \frac{\partial}{\partial [\mathbf{x}_k]_i} \mathrm{k}(\mathbf{x}_k, \mathbf{x}_k) - \\ &\quad 2\left(\frac{\partial}{\partial [\mathbf{x}_k]_i} \mathbf{K}_{\mathbf{x}_k\mathcal{D}}\right) \mathbf{S}^{-1} \mathbf{K}_{\mathcal{D}\mathbf{x}_k}, \end{aligned} \quad (28)$$

where we used in the derivative of the posterior variance that the covariance matrix $\mathbf{S}$ is symmetric and that the squared exponential kernel is a radial basis function for which $\mathrm{k}(\mathbf{a}, \mathbf{b}) = \mathrm{k}(\mathbf{b}, \mathbf{a})$. The derivative of the squared exponential kernel w.r.t. the i-th component $[\mathbf{x}_k]_i$ of the state is

$$\begin{aligned}\frac{\partial \mathrm{k}(\mathbf{x}_k, \mathbf{a})}{\partial [\mathbf{x}_k]_i} &= \frac{\partial}{\partial [\mathbf{x}_k]_i} \sigma_\mathrm{k}^2 \exp\left(-\frac{(\mathbf{x}_k - \mathbf{a})^T(\mathbf{x}_k - \mathbf{a})}{2l^2}\right) \\ &= -\mathrm{k}(\mathbf{x}_k, \mathbf{a}) \frac{([\mathbf{x}_k]_i - [\mathbf{a}]_i)}{l^2} = \frac{\partial \mathrm{k}(\mathbf{a}, \mathbf{x}_k)}{\partial [\mathbf{x}_k]_i}. \end{aligned} \quad (29)$$

The derivative of the constant mean function is simply zero everywhere. For the BIM in addition to FIM also the remaining terms of the recursive bound in (16) have to be derived. The remaining terms all require the pdf $p(\mathbf{x}_{k+1}|\mathcal{D}_u, \mathbf{x}_k)$ which is obtained from the motion model (19)

$$p(\mathbf{x}_{k+1}|\mathcal{D}_u, \mathbf{x}_k) = \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{F}\mathbf{x}_k + \mathbf{B}a_k, \mathbf{Q}). \quad (30)$$

Inserting (30) into (17) results in

$$\begin{aligned}\mathbf{D}_k^{11} &= \mathbf{F}^T\mathbf{Q}^{-1}\mathbf{F} \\ \mathbf{D}_k^{12} &= -\mathbf{F}^T\mathbf{Q}^{-1} = (\mathbf{D}_k^{21})^T \\ \mathbf{D}_k^{22} &= \mathrm{E}_{\mathbf{x}_{k+1}|\mathcal{D}_u}[\mathcal{F}(\mathbf{x}_{k+1})] + \mathbf{Q}^{-1}. \end{aligned} \quad (31)$$

For the term $\mathbf{D}_k^{22}$ the expected FIM w.r.t. the pdf $p(\mathbf{x}_{k+1}|\mathcal{D}_u)$ is required. In most cases the involved integral cannot be solved analytically and requires a numerical approximation. Since it is easy to obtain samples from $p(\mathbf{x}_{k+1}|\mathcal{D}_u)$, by propagating the prior state pdf $p(\mathbf{x}_0)$ with the motion model, the expected FIM can be obtained by evaluating the FIM for the different samples and then take the mean. To start the recursion in (16) the prior distribution is set to $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0, \mathbf{\Sigma}_{\mathbf{x},0})$ which gives the initial BIM $\mathcal{J}_0 = \mathbf{\Sigma}_{\mathbf{x},0}^{-1}$.

III. Evaluation

A. Measurement Setup

The evaluation is based on the magnetic field along a $1\,\mathrm{km}$ long track segment near the city of Augsburg. The sensors were mounted inside the driver's cabin of a Alstom Coradia LINT 41 diesel train from the Bayerische Regiobahn. The measurements were performed during normal service hours.

The magnetic field was measured with the magnetometer inside a Xsens MTi IMU at a rate of $200\,\mathrm{Hz}$. During the measurements the train speed and position were recorded using a u-blox LEA-6T single frequency GPS receiver.

IV. GP Training

The data set used in the GP to approximate the magnetic map are the along track position s and the corresponding three components of the magnetic vector field of the $1\,\mathrm{km}$ long track segment. For the bound only the spatial relation between the magnetometer measurements is of interest, the absolute along track position therefore is not important. Hence, to generate the training data for the GP it was sufficient to integrate the train speed to obtain the along track position for each of the magnetometer measurements. Furthermore, as described in our prior work [15], the magnetometer data has to be filtered with a band-stop filter at $16.67\,\mathrm{Hz}$ to remove the alternating magnetic fields introduced by the overhead power line currents. In the GP the data of four runs on the same track is considered to avoid overfitting the GP to temporary disturbances present

TABLE I
HYPERPARAMETERS OF THE GP

Axis	Sensor noise σ_n / a.u.	Length Scale l / m	Kernel Variance σ_k / a.u.	Mean $\bar{m}/a.u.$
x	$2.98 \cdot 10^{-3}$	3.18	$1.04 \cdot 10^{-2}$	$1.58 \cdot 10^{-4}$
y	$3.73 \cdot 10^{-3}$	4.92	$1.67 \cdot 10^{-2}$	$1.77 \cdot 10^{-3}$
z	$4.26 \cdot 10^{-3}$	3.69	$1.99 \cdot 10^{-2}$	$1.77 \cdot 10^{-3}$

only in one run and to get a more realistic estimate of the measurement noise. The four data sets are slighlty biased relative to each other. The bias is removed by removing the mean from the measurements since a constant bias can be removed in practice by magnetometer calibration and will not bring additional information into the bound.

Using the whole data sets of the four runs would lead to large covariance matrices in the GP and unpractical high computational complexity. Therefore, a two-step data compression is performed. First, a equidistant grid is placed along the track segment with a spacing of $0.5\,\mathrm{m}$. At each grid point then the closest point in each of the data sets is stored. This already reduces the amount of data significantly and ensures a minimum spatial density of data points. In the second step, for each grid point one of the associated points of the four data sets is selected randomly such that in average the different runs are included equally in the GP.

Based on the compressed data set hyperparameter optimization is performed with the GPML toolbox [16] using the squared exponential covariance function and a constant mean function. The optimization is done for each magnetic sensor axis separately. The resulting hyperparameters are listed in Table I and the resulting GP and the training data are shown in Fig. 1.

A. Particle Filter Implementation

To perform a sanity check for the derived BCRLB we implemented a particle filter and performed Monte Carlo simulation to compare its MSE to the bound. The filter uses the same motion model and measurement model as the bound. Unfortunately, incorporating the GP directly into the filter would lead to a high computational complexity that would make the Monte Carlo simulations very time consuming. Furthermore, the filter should be able to run in real time when used as localization system in a real train. We therefore decided to not include the full GP into the filter but instead use a offline calculated discretized map derived from the GP. The discretized map contains the mean and variance of the GP evaluated with a resolution of the along track position of $\delta s = 5\,\mathrm{mm}$. The value for δs was chosen such that it is well below the localization accuracy of the bound to avoid that the map resolution limits the accuracy of the particle filter. With the discretized map the GP regression is replaced with a look up table which considerably speeds up the filters weight update. Besides the discretized map the implemented filter is a basic particle filter where resampling is triggered when the effective number of particles falls below the threshold $0.8 \cdot N_p$ where the number of particles is set to the value $N_p = 15000$.

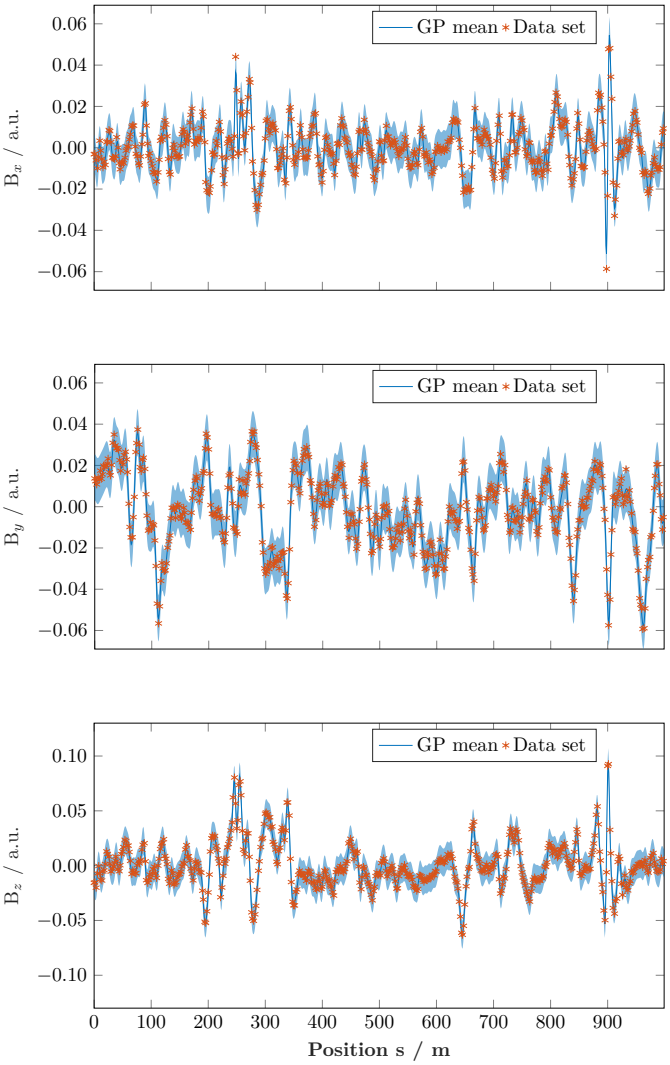

Fig. 1. The three components of the magnetic flux density B. The blue solid line is the mean of the GP and the colored area around it is three times the sum of the GP posterior standard deviation and measurement noise standard deviation $\pm 3 \cdot r_{i,c}(s_k, \mathcal{D})$. The shown flux density is dimension less (arbitrary units a.u.) because the sensor used during the measurements performs an internal normalization. According to the sensor manufacturer 1 a.u. is $\approx 40\,\mu\mathrm{T}$. The red stars show the data used in the GP regression.

B. Parameter Setup for Monte Carlo Simulation

The numerical approximation of $\mathbf{D}_k^{22}$ in (31) and the evaluation of the filter MSE requires a Monte Carlo simulation. In the simulation the expected FIM is obtained by averaging over 1000 state trajectories with a length of $30\,\mathrm{s}$ that are randomly generated from the motion model. For each of the generated trajectories then additionally 50 realizations of the measurements are generated to calculate the MSE of the filter. Hence, in total the filter is evaluated fifty thousand times.

The acceleration input for the generation of the trajectories is set to $a_k = 2\,\mathrm{m/s^2}$ for the first ten seconds, then to $a_k = 0\,\mathrm{m/s^2}$ for the next ten seconds, and finally the train decelerates with $a_k = -2\,\mathrm{m/s^2}$ for the last ten seconds. Overall 300 time steps are simulated with a sampling period of $T = 0.1\,\mathrm{s}$. The noise of the motion model is $0.25\,\mathrm{m}/\sqrt{\mathrm{s}^3}$. The

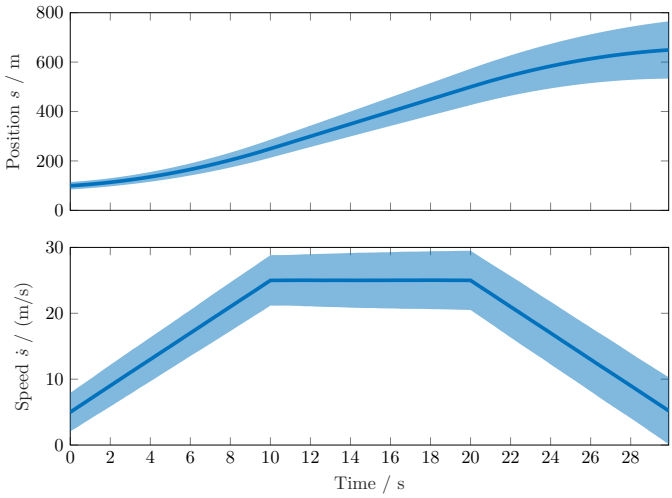

Fig. 2. Mean (solid line) and three times the standard deviation (colored area) of the predicted state pdf $p(\mathbf{x}_k|\mathcal{D}_u)$ over which the FIM is averaged.

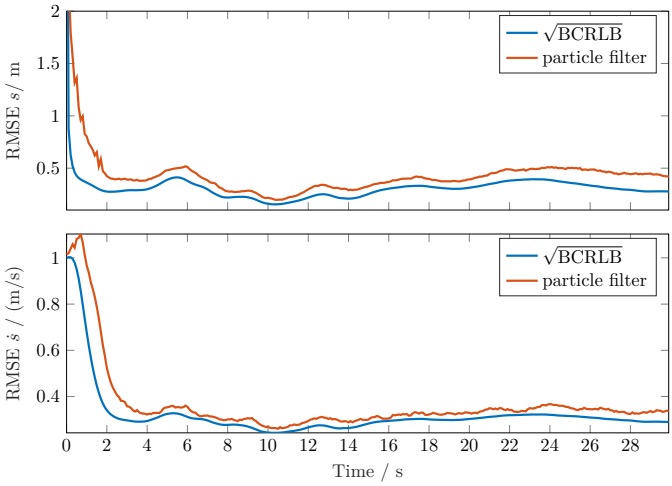

Fig. 3. Comparison between the square root of the BCRLB and the RMSE of the particle filter for position (top) and speed (bottom). The y-axis of the position RMSE is zoomed in to better see the difference between the bound and the filter. The initial value of the position error is $5\,\text{m}$ as defined by the value of $\mathbf{\Sigma}_{\mathbf{x},0}$.

initial state is set to $\mathbf{x}_0 = \begin{bmatrix}100\,\text{m} & 5\,\text{m/s}\end{bmatrix}^T$ and its covariance matrix to $\text{diag}(\begin{bmatrix}25\,\text{m}^2 & 1\,\text{m}^2/\text{s}^2\end{bmatrix})$. The distribution of the simulated trajectories are shown in Fig. 2.

V. Results

The results for the BCRLB and the root mean squared error (RMSE) of the particle filter are shown in Fig. 3. The bound starts at the value defined by covariance matrix $\mathbf{\Sigma}_{\mathbf{x},0}$ of the prior distribution of the state vector. The error is then quickly reduced to values in the range of roughly $\approx 0.15 - 0.41\,\text{m}$ due to the information obtained from the magnetic field measurements. From the bound no clear relation between the train speed and the achievable position error can be seen, indicating that magnetic localization can achieve high accuracies also at low speeds. The bound for the speed shows that also the speed error is quickly reduced from its initial value to values in the range of roughly $\approx 0.24 - 0.33\,\text{m/s}$. This shows that even though the FIM (26) for the speed is zero, the speed is observable due to the motion model that links the speed to the position. While in this particular example the observability of the speed is obvious, for more involved state space models the BCRLB is a good way to analyze the observability of state variables.

In comparison to the bound the particle filter has a larger RMSE for both state variables. The difference after $5\,\text{s}$ is roughly $\approx 0.1\,\text{m}$ for the position and $\approx 0.03\,\text{m/s}$ for the speed. Similar to the bound, after an initial larger position error the RMSE is quickly reduced to sub meter accuracy. In general, the filter RMSE nicely follows the shape of the bound but never attains it. A comparable behavior is observed for the estimated speed. This shows that even under the ideal conditions of a simulation where the motion and measurement model are exactly known to the filter the filter is not optimal in the MSE sense. But this does not necessarily mean that a better filter exists for this particular estimation problem because the bound might be too optimistic.

VI. Conclusion

In this paper the BCRLB for magnetic field-based train localization was derived. The issue in the derivation is that the measurement model that describes the magnetic field and that relates the magnetometer measurements to the train position is known only from measurements and does not have an analytical, differentiable form that is required by the bound. To address this issue, we showed how the unknown measurement model can be approximated with a GP based on observations of the magnetic field.

An evaluation of the derived bound with measurement data showed that in theory position errors below $1\,\text{m}$ are possible. Since in the bound some optimistic assumptions are used, e.g., perfectly calibrated sensors, this accuracy most likely will not be attainable in practice. Furthermore, a comparison of the bound with the RMSE of a particle filter resulted in a gap between the two, showing that even under ideal conditions the filter cannot attain the bound. Nevertheless, the bound gives an indication of what is possible and it serves as benchmark during filter development.

Acknowledgment

This work is carried out in the Ubiquitous Spatio-Temporal Learning for Future Mobility (ULearn4Mobility) project funded by the Helmholtz AI Cooperation Unit under the grant ZT-I-PF-5-49.

We also want to thank the railway transportation company BRB (”Bayerische Regiobahn”) for their cooperation with the measurement recordings.

References

[1] J. Haverinen and A. Kemppainen, “Global indoor self-localization based on the ambient magnetic field,” *Robotics and Autonomous Systems*, vol. 57, no. 10, pp. 1028–1035, Oct. 2009.

[2] J. A. Shockley and J. F. Raquet, “Navigation of Ground Vehicles Using Magnetic Field Variations,” *Navigation*, vol. 61, no. 4, pp. 237–252, 2014.

[3] A. Canciani and J. Raquet, “Airborne Magnetic Anomaly Navigation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 67–80, Feb. 2017.

[4] B. Siebler, O. Heirich, and S. Sand, "Train Localization with Particle Filter and Magnetic Field Measurements," in *2018 21st International Conference on Information Fusion (FUSION)*, Jul. 2018, pp. 1–5.

[5] B. Siebler, S. Sand, and U. D. Hanebeck, "Bayesian Cramér-Rao Lower Bound for Magnetic Field-Based Localization," *IEEE Access*, vol. 10, pp. 123 080–123 093, 2022.

[6] Y. Zhao, C. Fritsche, G. Hendeby, F. Yin, T. Chen, and F. Gunnarsson, "Cramér-Rao Bounds for Filtering Based on Gaussian Process State-Space Models," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 5936–5951, Dec. 2019.

[7] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. Hoboken: John Wiley & Sons, 2004.

[8] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge: MIT Press, 2006.

[9] N. Akai and K. Ozaki, "Gaussian processes for magnetic map-based localization in large-scale indoor environments," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015, pp. 4459–4464.

[10] I. Vallivaara, J. Haverinen, A. Kemppainen, and J. Röning, "Simultaneous localization and mapping using ambient magnetic field," in *2010 IEEE Conference on Multisensor Fusion and Integration*, Sep. 2010, pp. 14–19.

[11] S. M. Kay, *Fundamentals of statistical signal processing*. New Jersey: Prentice Hall PTR, 1993.

[12] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[13] B.-S. Yaakov, L. X.-Rong, and K. Thiagalingam, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. Hoboken: John Wiley & Sons, 2002.

[14] B. Siebler, A. Lehner, S. Sand, and U. D. Hanebeck, "Evaluation of Simultaneous Localization and Calibration of a Train Mounted Magnetometer," in *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, Sep. 2021, pp. 2285–2293.

[15] O. Heirich and B. Siebler, "Train-side passive magnetic measurements," in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2015, pp. 687–692.

[16] C. E. Rasmussen and H. Nickisch, "Gaussian Processes for Machine Learning (GPML) Toolbox," *Journal of Machine Learning Research*, vol. 11, no. 100, pp. 3011–3015, 2010.