

Research Article

Bayesian Estimation Applied to Stochastic Localization with Constraints due to Interfaces and Boundaries

Wolfgang Hoegele,^{1,2} Rainer Loeschel,² Barbara Dobler,¹
Oliver Koelbl,¹ and Piotr Zygmanski³

¹ Department of Radiation Oncology, Regensburg University Medical Center, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany

² Department of Computer Science and Mathematics, University of Applied Sciences, UniversitaetsstraÙe 31, 93053 Regensburg, Germany

³ Department of Radiation Oncology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

Correspondence should be addressed to Wolfgang Hoegele; wolfgang@hoegele.de

Received 21 December 2012; Accepted 13 March 2013

Academic Editor: Marcelo Moreira Cavalcanti

Copyright © 2013 Wolfgang Hoegele et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. We present a systematic Bayesian formulation of the stochastic localization/triangulation problem close to constraining interfaces. **Methods.** For this purpose, the terminology of Bayesian estimation is summarized suitably for applied researchers including the presentation of Maximum Likelihood (ML), Maximum A Posteriori (MAP), and Minimum Mean Square Error (MMSE) estimation. Explicit estimators for triangulation are presented for the linear 2D parallel beam and the nonlinear 3D cone beam model. The priors in MAP and MMSE optionally incorporate (A) the hard constraints about the interface and (B) knowledge about the probability of the object with respect to the interface. All presented estimators are compared in several simulation studies for live acquisition scenarios with 10,000 samples each. **Results.** First, the presented application shows that MAP and MMSE perform considerably better, leading to lower Root Mean Square Errors (RMSEs) in the simulation studies compared to the ML approach by typically introducing a bias. Second, utilizing priors including (A) and (B) is very beneficial compared to just including (A). Third, typically MMSE leads to better results than MAP, by the cost of significantly higher computational effort. **Conclusion.** Depending on the specific application and prior knowledge, MAP and MMSE estimators strongly increase the estimation accuracy for localization close to interfaces.

1. Introduction

Due to their inherent ability to use prior knowledge, Bayesian approaches have gained interest in many fields in the recent years, such as in computer vision [1, 2] and medical applications [3–6], as well as robotics (Kalman-Filter, Particle-Filter) [7, 8] and metrology [9]. In this work, we present a Bayesian framework for the localization of objects with constraints derived from interfaces (surfaces) or boundaries specific to a given localization problem.

Methods to find the locations of objects based on several projections are known from photogrammetry for a long time. These methods are also called triangulation in a multiple view geometry (e.g., see Hartley and Zisserman [1] for an overview

of established methods). Introducing probabilities in order to deal with the Gaussian measurement noise is a well-known approach for estimating a 3-dimensional (3D) point based on two 2-dimensional (2D) projections, for example, see Hartley and Sturm [10]. In this context, Bedekar and Haralick [11] provided a Maximum A Posteriori (MAP) estimator by introducing prior knowledge independent of the observation for the estimation based on a set of 2D projections, which represents a starting point for this Paper but does not provide a more general perspective on estimation which is required to bring it to the level described in this paper. We are following this probabilistic interpretation of the inverse problem of triangulation and present explicit solutions in order to characterize the triangulation problem as well as

embed it into the general framework of Bayesian estimation. In particular, we demonstrate the benefit of including soft and hard constraints about the solution space into the *prior* of the Bayesian triangulation problem. We compare three stochastic estimation methods: Maximum Likelihood (ML), Maximum A Posteriori (MAP), and Minimum Mean Square Error (MMSE), and show that, depending on the prior knowledge, MAP or MMSE is a suitable approach for this kind of problem. We show how the estimation methods can be utilized for a 3D nonlinear/non-Gaussian geometrical model in physical space without modifications or approximations of the observation geometry.

In the localization problem the input data consists of 2D projections (or images) of the 3D objects and the goal is the optimal localization of features contained in these objects, by employing prior knowledge. In general, projections of 3D objects onto 2D planes contain information about their coordinates, but due to the experimental conditions these coordinates may not be accurate or may not be easily extractable from these projections. Thus the localization problem is typically an ill-posed inverse problem. Fortunately, in many applications there exists prior knowledge about the probability densities of these uncertainties. Furthermore, if there is a general probabilistic knowledge about the position of the observed object or additional constraints about the domain of its coordinates, the total prior knowledge becomes difficult to be incorporated into the localization model. Such localizations close to a constraining interface or boundary can be modeled and solved in the Bayesian approach. For instance, in the field of medicine, during image-guided procedures (interventional radiology, radiotherapy, and surgical procedures with C-arm CBCT scanners) there is often a need to obtain 3D coordinates of anatomical features or implanted markers instantly from multiple X-ray projections, which are constrained by external or internal anatomical interfaces (e.g., patient surface) and form a hard or semi-flexible barrier. Both, the general Bayesian framework and its application to a specific localization problem presented herein can also be utilized for other applications in image or data processing, such as computer vision or metrology.

The general aim of this paper is to provide a pragmatic foundation for the Bayesian estimation in a compact way suitable for applied sciences utilizing a concise and uniform mathematical notation (general parts about the Bayesian methodology of this paper are partly taken from the doctoral thesis of the first author, which is submitted (and not yet published) at the Regensburg University, Germany, at the time of this paper submission), which is difficult to find in more theoretical formulations and whose full understanding is required for the localization problem at stake. This brief introduction to Bayesian estimation may benefit several research fields in which there is a gap between the general mathematical concepts and direct applications of Bayesian methods. Thus the major motivation of presenting the Bayesian estimation theory is the linking of these two areas by providing the Bayesian formalism for researches in applied sciences. The second aim of this work is to demonstrate that a Bayesian framework can be effectively utilized for more accurate localization of

features close to interfaces, which may have many applications.

The structure of the paper is as follows: first, we present the basic terminology of Bayesian estimation required to derive the solutions presented in the subsequent sections based on a concise summary of the books of Lehmann and Casella [12] and Robert [13]. Second, we specifically present the application of stochastic localization applied to 1D and 2D projections of 2D and 3D objects. Besides such a structured presentation of Bayesian triangulation, the hope is that this course of explanation will help the readers to develop their own Bayesian estimators for their specific problems at hand.

2. General Bayesian Theory

2.1. Introduction to the Terminology of Estimation. The goal of estimation theory is to determine a concrete set of unknown real-valued parameters $\mathbf{x} \in \Omega_{\mathbf{x}} \subseteq \mathbb{R}^m$ (e.g., the 2D/3D position coordinates of the observed object in triangulation) from real-valued observations $\mathbf{u} \in \Omega_{\mathbf{u}} \subseteq \mathbb{R}^n$ (e.g., the detected or measured 1D/2D projections of the object), which are imperfectly or ambiguously related to each other (bold fonts of symbols are utilized to emphasize multidimensional vectors). Their ambiguous relation is due to the experimental uncertainties of the given problem, such as an imperfect Radon transform, or similar transforms. The estimation of parameters \mathbf{x} from observables \mathbf{u} demands a set of definitions which will be briefly introduced in this section. The concrete parameters \mathbf{x} and observations \mathbf{u} are associated with *continuous random variables* \mathbf{X} and \mathbf{U} , respectively (denoted by corresponding capital letters), which are further associated with the underlying probability distributions describing the sources of uncertainties. With this concept a connection between the deterministic and probabilistic quantities is established. For instance, the probability that \mathbf{x} takes a value lower than \mathbf{s} is defined by

$$P(\mathbf{X} \leq \mathbf{s}) := \int_{-\infty}^{\mathbf{s}} f_{[\mathbf{X}]}(\mathbf{x}) d\mathbf{x}, \quad (1)$$

with the corresponding Probability Density Function (PDF) $f_{[\mathbf{X}]}(\mathbf{x})$. This fact is summarized by stating that $\mathbf{X} \sim f_{[\mathbf{X}]}$.

In general, the occurrence of one event can influence the probability of another. It is possible to define such a residual probability by conditional probabilities, and for the corresponding PDF describing the random variable \mathbf{U} under the condition of $\mathbf{X} = \mathbf{x}$ this is formulated by

$$f_{[\mathbf{U}]}(\mathbf{u} | \mathbf{x}) := \frac{f_{[\mathbf{U}, \mathbf{X}]}(\mathbf{u}, \mathbf{x})}{f_{[\mathbf{X}]}(\mathbf{x})}. \quad (2)$$

In this formula, the expression $f_{[\mathbf{U}, \mathbf{X}]}(\mathbf{u}, \mathbf{x})$ describes the joint PDF of \mathbf{U} and \mathbf{X} . Based on this concept, the most important quantities for estimation are defined in the next subsections.

2.1.1. Likelihood, Prior, and Posterior. In Table 1 the most important types of PDFs for estimation are presented and interpreted in the following paragraphs. The PDF $f_{[\mathbf{U}]}(\mathbf{u} | \mathbf{x})$ defines the probability of the current observations \mathbf{u} if the

TABLE 1: Probability Density Functions (PDF) which are important for estimation theory.

$f_{[U]}(\mathbf{u} \mid \mathbf{x})$	PDF of the observation \mathbf{u} conditioned on \mathbf{x}
$\mathcal{L}_{\mathbf{u}}(\mathbf{x}) := f_{[U]}(\mathbf{u} \mid \mathbf{x})$	The <i>likelihood</i> function (depending on the condition \mathbf{x})
$\pi(\mathbf{x}) := f_{[X]}(\mathbf{x})$	the <i>prior</i> , that is, PDF of the prior knowledge about \mathbf{x}
$\pi(\mathbf{x} \mid \mathbf{u}) := f_{[X]}(\mathbf{x} \mid \mathbf{u})$	the <i>posterior</i> , that is, PDF of the posterior distribution of parameter \mathbf{x} conditioned on \mathbf{u}
$\int_{\Omega_{\mathbf{x}}} f_{[U]}(\mathbf{u} \mid \mathbf{x}) \cdot \pi(\mathbf{x}) d\mathbf{x}$	PDF of the probability of the occurrence of \mathbf{u} (=: marginal distribution)
$\mathbf{X} \sim f_{[X]}, \mathbf{U} \sim f_{[U]}$	random variables \mathbf{X} and \mathbf{U} are distributed according to their PDFs $f_{[X]}$ and $f_{[U]}$

true values \mathbf{x} are known. In a typical observation scenario this corresponds to the accuracy of the observation device with respect to a specific true parameter \mathbf{x} . It is common to define the likelihood function $\mathcal{L}_{\mathbf{u}}(\mathbf{x}) := f_{[U]}(\mathbf{u} \mid \mathbf{x})$ describing the same probabilities of the observation but this time with the focus on the conditioning parameter \mathbf{x} . Therefore the likelihood is regarded as a *similarity measure* between the current observation \mathbf{u} and possibly “true” parameters \mathbf{x} .

The PDF of the *prior* $\pi(\mathbf{x}) := f_{[X]}(\mathbf{x})$ defines how probable the occurrence of \mathbf{x} is; that is, $\mathbf{X} \sim \pi$. The use of the *prior* is an intensely discussed topic, since one could derive arbitrary conclusions by utilizing a specific and strong prior knowledge which supports a certain point of view which may be just one of many possible options. Although this is generally valid, the utilization of appropriate prior knowledge (noninformative priors, priors based on actual observations of the same or similar situations, or priors derived from knowledge about the range of \mathbf{x}) can reduce the uncertainties inherent to \mathbf{x} in addition to making observations \mathbf{u} .

In *Bayesian* statistics the most important PDF is the *posterior* $\pi(\mathbf{x} \mid \mathbf{u}) := f_{[X]}(\mathbf{x} \mid \mathbf{u})$, which defines the probabilities of \mathbf{x} under the condition of the observation \mathbf{u} . Similar to the *likelihood* this quantity is also often regarded as a *similarity measure* between the observation and \mathbf{x} .

The aforementioned PDFs are related to each other by *Bayes’ Theorem*, which is expressed as

$$\pi(\mathbf{x} \mid \mathbf{u}) = \frac{f_{[U]}(\mathbf{u} \mid \mathbf{x}) \cdot \pi(\mathbf{x})}{\int_{\Omega_{\mathbf{x}}} f_{[U]}(\mathbf{u} \mid \mathbf{x}) \cdot \pi(\mathbf{x}) d\mathbf{x}}. \quad (3)$$

The benefit of this theorem is that it allows to reverse the condition by introducing prior knowledge about the parameter of interest. In practical use, with this theorem the *posterior* $\pi(\mathbf{x} \mid \mathbf{u})$ can be determined by having only knowledge about the *likelihood* $\mathcal{L}_{\mathbf{u}}(\mathbf{x})$ and the *prior* $\pi(\mathbf{x})$, since the denominator on the right side is typically determined indirectly by normalizing the posterior to the integral value of 1.

2.1.2. Loss and Risk. An *estimator* is any function $\delta : \Omega_{\mathbf{U}} \rightarrow \Omega_{\mathbf{X}}$, which utilizes the current observations \mathbf{u} in order to find an approximation to the true parameter \mathbf{x} . In order to decide on the use of an estimator compared to another, one needs a quality criterion, which depends on the true value for the parameters \mathbf{x} and an estimate $\delta(\mathbf{u})$ based on the observation \mathbf{u} . In estimation theory this is realized with the concept of the *loss function* $L(\mathbf{x}, \delta(\mathbf{u}))$. An important subset is *strictly convex loss functions*, such as the popular *quadratic loss*,

$$L(\mathbf{x}, \delta(\mathbf{u})) := \|\mathbf{x} - \delta(\mathbf{u})\|^2, \quad (4)$$

since they lead to unique Bayesian estimators. An example for a *not strictly* but still convex loss is the *constant loss*

$$L(\mathbf{x}, \delta(\mathbf{u})) := \begin{cases} 0, & \text{for } \max_{i \in \{1, \dots, p\}} |x_i - \delta_i(\mathbf{u})| \leq \varepsilon, \\ 1, & \text{else,} \end{cases} \quad (5)$$

for a constant $\varepsilon > 0$, which is of technical interest for the derivation of the MAP estimator. In this context, a very attractive thought is to find an optimal estimator $\delta(\cdot)$ by minimizing the loss function $L(\mathbf{x}, \delta(\mathbf{u}))$. Unfortunately, the naive minimization of the loss would lead to $\delta(\mathbf{u}) = \mathbf{x}$, which is a practically useless estimator since it demands knowledge about the true \mathbf{x} . Essentially, the loss itself does not contain any information about the relation between the parameter \mathbf{x} and the measurements \mathbf{u} and, therefore, represents an incomplete description of the estimation problem. Due to this observation, in classical estimation theory the concept of the *risk function* is introduced, which is defined here by the conditional expectation value.

Definition 1. The risk function (or alternatively average loss) is defined as (It is inherent to the classical frequentist’s approach that the parameters are not interpreted as random variables \mathbf{X} . Since in this work a general Bayesian perspective is focused on, in the definition of the frequentist risk, the parameters are written as random variables but simultaneously conditioned on $\mathbf{X} = \mathbf{x}$, which essentially leads to the same risk function.)

$$\begin{aligned} R(\mathbf{x}, \delta) &:= \mathbb{E}[L(\mathbf{X}, \delta(\mathbf{U})) \mid \mathbf{X} = \mathbf{x}] \\ &= \int_{\Omega_{\mathbf{U}}} L(\mathbf{x}, \delta(\mathbf{u})) \cdot f_{[U]}(\mathbf{u} \mid \mathbf{x}) d\mathbf{u}. \end{aligned} \quad (6)$$

The risk function $R(\mathbf{x}, \delta)$ incorporates all possible outcomes \mathbf{u} of the observation and weights them with their likelihood $f_{[U]}(\mathbf{u} \mid \mathbf{x})$. For this reason, the long-term average loss of a given estimator $\delta(\cdot)$ is obtained by the risk function of (6). In consequence, based on the risk function a reasonable comparison of estimators can be performed using the concept of *admissibility* (see Lehmann and Casella [12, pp. 309–376] or Robert [13, pp. 74–77, 391–415] for details). On the other side, the risk function still depends on the unknown parameter \mathbf{x} , and, therefore, direct minimization of the risk function can only be performed for a fixed \mathbf{x} . In order to incorporate all possible values of \mathbf{x} into a single quality criterion an extension of the concept of the risk function is introduced by incorporating (possibly noninformative) prior knowledge about \mathbf{x} by $\pi(\mathbf{x})$.

Definition 2. The integrated risk function is defined as the expectation value taken over \mathbf{X} and \mathbf{U}

$$r(\pi, \delta) := \mathbb{E}[L(\mathbf{X}, \delta(\mathbf{U}))] \\ = \int_{\Omega_{\mathbf{X}}} \int_{\Omega_{\mathbf{U}}} L(\mathbf{x}, \delta(\mathbf{u})) \cdot f_{[\mathbf{U}]}(\mathbf{u} | \mathbf{x}) \cdot \pi(\mathbf{x}) d\mathbf{u} d\mathbf{x}. \quad (7)$$

With this criterion one gets a single real value for a given estimator $\delta(\cdot)$ which incorporates all parameters \mathbf{x} and possible observations \mathbf{u} and allows a straight comparison between different estimators (and especially a meaningful optimization). Unfortunately, to find an optimal estimator directly from the minimization of the integrated risk of (7) is a nontrivial problem but can be solved by Bayesian estimation as presented in Section 2.2. Further, a delicate point is the introduction of a prior distribution $\pi(\mathbf{x})$ before any measurement is performed since this could bias the estimation result in a disadvantageous way. On the other side, if $\Omega_{\mathbf{X}}$ is bounded and $\pi(\mathbf{x}) = 1/\text{vol}(\Omega_{\mathbf{X}}) = \text{const.}$ for all $\mathbf{x} \in \Omega_{\mathbf{X}}$, all possible values \mathbf{x} are treated equally likely in order to erase the subjective notion when introducing the prior. In consequence, the application of the integrated risk function in order to derive estimators could also be of interest in cases in which one wants to avoid such a subjective notion and still get an estimator working best for all possible values \mathbf{x} and \mathbf{u} in average.

2.2. Maximum Likelihood, Maximum A Posteriori, and Bayesian Estimation. In this section, three central estimation methods are introduced in a general way (for a brief characterization of these approaches, see Section 2.3).

2.2.1. Maximum Likelihood (ML). Maximum Likelihood estimation is typically applied in situations in which no prior knowledge about $\mathbf{x} \in \Omega_{\mathbf{X}}$ is available.

Definition 3. A Maximum Likelihood (ML) estimator for a given observation $\mathbf{u} \in \Omega_{\mathbf{U}}$ is

$$\delta(\mathbf{u}) := \arg \max_{\mathbf{x}} \mathcal{L}_{\mathbf{u}}(\mathbf{x}) = \arg \max_{\mathbf{x}} f_{[\mathbf{U}]}(\mathbf{u} | \mathbf{x}), \quad (8)$$

where “arg max” means that we want to find the position \mathbf{x} which maximizes the likelihood $\mathcal{L}_{\mathbf{u}}(\mathbf{x})$. This is also valid for the logarithm of the likelihood since the logarithm is a strictly increasing function, and the likelihood is positive. Since the ML estimation depends on the shape of the likelihood function which is problem-specific, under fairly general assumptions, the existence but not uniqueness of the estimate can be guaranteed. The theorem by Lehmann and Casella [12, p. 463] assures the existence of an asymptotically unbiased, consistent, and asymptotically efficient estimator which is the solution of $\nabla_{\mathbf{x}} \mathcal{L}_{\mathbf{u}}(\mathbf{x}) = \mathbf{0}$, the so-called likelihood equations. If there is a unique solution to these equations, then in consequence, the corresponding ML fulfills these properties. From a practical perspective the determination of the estimate results in an optimization problem, and numerically this may lead to problems with respect to running into local maxima instead of global maxima.

2.2.2. Bayesian Estimator (BE). Bayesian estimation utilizes prior knowledge about $\mathbf{x} \in \Omega_{\mathbf{X}}$, and focuses on the given observation. For this reason, a modified risk function needs to be introduced at first.

Definition 4. The Bayesian risk function is defined for a given loss L , observation $\mathbf{u} \in \Omega_{\mathbf{U}}$ and the posterior $\pi(\mathbf{x} | \mathbf{u})$ as

$$\rho_{\mathbf{u}}(\pi, \delta) := \mathbb{E}[L(\mathbf{X}, \delta(\mathbf{U})) | \mathbf{U} = \mathbf{u}] \\ = \int_{\Omega_{\mathbf{X}}} L(\mathbf{x}, \delta(\mathbf{u})) \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}. \quad (9)$$

The main difference between the Bayesian risk (9) and the integrated risk (7) is that the integrated risk does average over all possible observations $\mathbf{u} \in \Omega_{\mathbf{U}}$, and the Bayesian risk utilizes only the concretely acquired observation \mathbf{u} .

Definition 5. A Bayesian estimator (BE) for a given observation $\mathbf{u} \in \Omega_{\mathbf{U}}$ is the estimator $\delta(\mathbf{u})$ which minimizes the Bayesian risk function

$$\delta(\mathbf{u}) := \arg \min_{\xi} \rho_{\mathbf{u}}(\pi, \xi). \quad (10)$$

At the first glance, the Bayesian estimator therefore looks unrelated to the optimization of the previously introduced risk functions in Section 2.1. A powerful statement is achieved by the following theorem.

Theorem 6. The estimation function which minimizes the integrated risk $r(\pi, \delta)$ can be obtained by selecting for every specific observation $\mathbf{u} \in \Omega_{\mathbf{U}}$ the estimate $\delta(\mathbf{u})$ which minimizes the Bayesian risk $\rho_{\mathbf{u}}(\pi, \delta)$.

For the proof see Robert [13, p. 63]. A major property of Bayesian estimators is that if the loss is strictly convex, then the corresponding BE is unique. This will be shown explicitly in the next paragraph for the quadratic loss. For this loss the properties are similar as for the ML estimator, leading to consistent and asymptotically efficient estimators under some weak assumptions on the prior; see Lehmann and Casella [12, p. 490]. From a practical point of view, the numerical computation of the integral might be limiting the applications of BE if analytical solutions cannot be derived. Compared to the optimization problem in ML, the general calculation of these multiple integrals is typically computationally more challenging.

2.2.3. The Minimum Mean Square Error (MMSE). The Bayesian estimator utilizing the quadratic loss is called the Minimum Mean Square Error (MMSE). The derivation leads to

$$\delta(\mathbf{u}) := \arg \min_{\xi} \mathbb{E}[L(\mathbf{X}, \xi) | \mathbf{U} = \mathbf{u}] \\ = \arg \min_{\xi} \int_{\Omega_{\mathbf{X}}} \|\mathbf{x} - \xi\|^2 \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}. \quad (11)$$

Since the loss is strictly convex and (11) represents a parameter integral with respect to ξ , the unique solution can be derived from

$$\nabla_{\xi} \mathbb{E} [L(\mathbf{X}, \xi) | \mathbf{U} = \mathbf{u}] \stackrel{!}{=} \mathbf{0}. \quad (12)$$

This means for every coordinate $i \in \{1, \dots, m\}$

$$\begin{aligned} \int_{\Omega_{\mathbf{x}}} 2(x_i - \xi_i) \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x} &\stackrel{!}{=} 0, \\ \xi_i \cdot \underbrace{\int_{\Omega_{\mathbf{x}}} \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}}_{=1} &= \int_{\Omega_{\mathbf{x}}} x_i \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}. \end{aligned} \quad (13)$$

This leads to the corresponding MMSE for each coordinate $i \in \{1, \dots, m\}$ with

$$\delta_i(\mathbf{u}) = \mathbb{E} [X_i | \mathbf{U} = \mathbf{u}] = c \cdot \int_{\Omega_{\mathbf{x}}} x_i \cdot \mathcal{L}_{\mathbf{u}}(\mathbf{x}) \cdot \pi(\mathbf{x}) d\mathbf{x}, \quad (14)$$

which is the *expectation value of the posterior* for every coordinate $i \in \{1, \dots, m\}$, with the normalization constant c . This example already points out the importance of the shape of the posterior $\pi(\mathbf{x} | \mathbf{u})$ for the Bayesian approach.

2.2.4. Maximum A Posteriori (MAP). An alternative approach, conceptually in between ML and BE, is the Maximum A Posteriori (MAP) estimation which already utilizes prior knowledge about $\mathbf{x} \in \Omega_{\mathbf{x}}$, but has many characteristics of the ML approach.

Definition 7. The Maximum A Posteriori estimator (MAP) for a given observation $\mathbf{u} \in \Omega_{\mathbf{U}}$ is

$$\boldsymbol{\delta}(\mathbf{u}) := \arg \max_{\mathbf{x}} \pi(\mathbf{x} | \mathbf{u}) = \arg \max_{\mathbf{x}} \mathcal{L}_{\mathbf{u}}(\mathbf{x}) \cdot \pi(\mathbf{x}), \quad (15)$$

since the marginal distribution is positive and independent of \mathbf{x} . In other words, MAP is the mode of the posterior.

When applying the logarithm to $\mathcal{L}_{\mathbf{u}}(\mathbf{x}) \cdot \pi(\mathbf{x})$, this approach can be identified with a specific Tikhonov regularization of the ML approach (also known as penalized Maximum Likelihood), as discussed in Eldar [14]. Since the MAP estimation depends on the shape of the likelihood and prior function which is problem specifically defined, under fairly general assumptions, the existence but not uniqueness of the estimate can be guaranteed. The asymptotic properties are the same as for the ML case, with some further regularity conditions on $\mathcal{L}_{\mathbf{u}}(\mathbf{x})$ and $\pi(\mathbf{x})$; for example, see Robert [13, p. 166]. Although MAP utilizes prior knowledge about $\mathbf{x} \in \Omega_{\mathbf{x}}$, the numerical problem is the same as for ML, that is, resulting in a numerical optimization problem with the problem of identifying local and global maxima.

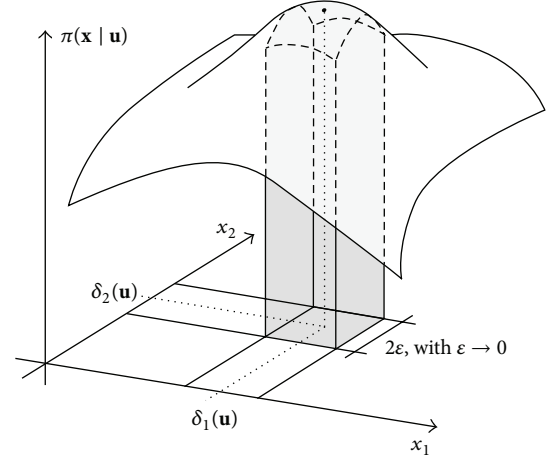


FIGURE 1: The MAP estimator as a limit of Bayesian estimators with $\mathbf{x} \in \mathbb{R}^2$. For every fixed ε the Bayesian estimator is the midpoint $\boldsymbol{\delta}(\mathbf{u})$ which maximizes the gray shaded volume.

In the context of Bayesian estimation, the MAP estimation can be derived as a limit of Bayesian estimators with respect to the constant loss function (5). It is shown that

$$\begin{aligned} \boldsymbol{\delta}(\mathbf{u}) &:= \arg \min_{\xi} \mathbb{E} [L(\mathbf{X}, \xi) | \mathbf{U} = \mathbf{u}] \\ &= \arg \min_{\xi} \int_{\Omega_{\mathbf{x}}} L(\mathbf{x}, \xi) \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x} \\ &= \arg \max_{\xi} \int_{C_{\xi}} \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}, \end{aligned} \quad (16)$$

with C_{ξ} as the m -dimensional cube with side length 2ε and midpoint ξ . In other words, the aim is to find the midpoint of the m -dimensional cube C_{ξ} , so that the volume under the graph of the posterior is maximized. By assuming an at least continuous posterior and letting $\varepsilon \rightarrow 0$, the best midpoint $\boldsymbol{\delta}(\mathbf{u})$ is the mode of the posterior, that is, the MAP estimator. This is illustrated in Figure 1. Since the constant loss function is not strictly convex, the corresponding BEs are not necessarily unique, and this implicates the nonuniqueness of the MAP as a limit of these BEs.

2.3. Characterization of the Estimators. The presented estimators are closely related which is graphically illustrated in Figure 2. A major distinction is drawn in between the use of prior knowledge about $\mathbf{x} \in \Omega_{\mathbf{x}}$ or not, which is shown as a thick boundary in Figure 2. In detail, the main connections are summarized in the following list.

- (i) BE \rightarrow MMSE: the general BE reduces to the MMSE when applying the quadratic loss of (4).
- (ii) MMSE \rightarrow MAP: if the posterior distribution is symmetric, that is, $\pi(\boldsymbol{\mu} + \mathbf{x} | \mathbf{u}) = \pi(\boldsymbol{\mu} - \mathbf{x} | \mathbf{u})$ around the global maximum $\boldsymbol{\mu}$, then the MMSE estimator is identical with the MAP estimator. In this case, due to

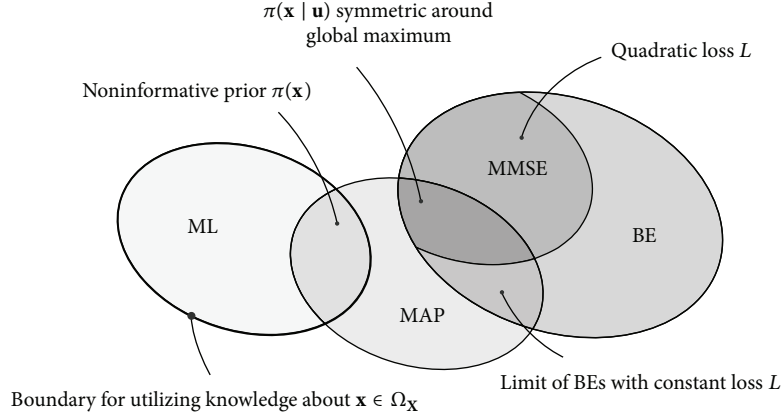


FIGURE 2: Illustration of the relation between estimation based on ML, MAP, MMSE, and Bayesian estimation. The intersections state under which conditions the estimators are equivalent. Every condition appears only once; that is, since the conditions are not excluding each other, multiple conditions lead to further equalities.

symmetry the expectation and the maximum of the posterior coincide.

- (iii) BE \rightarrow MAP: the BE converges to the MAP estimator when applying the constant loss of (5) with $\varepsilon \rightarrow 0$.
- (iv) MAP \rightarrow ML: if the prior is noninformative, that is, constant for a finite Ω_X or asymptotically noninformative for an infinite Ω_X as the deviation parameters of the prior go to infinity, then MAP coincides with the ML estimate.

Of course, the combination of the stated properties is possible, for example, for a symmetric posterior distribution around a global maximum and a noninformative prior the ML and MMSE coincide. There are several other concepts utilized in the literature for deriving non-Bayesian estimators, that is, without (possibly noninformative) prior knowledge about $\mathbf{x} \in \Omega_X$, such as the following.

- (i) *Unbiasedness*, that is, the search of an unbiased estimator which minimizes the frequentist risk locally or uniformly. This property has been regarded as very important from the beginning of estimation theory (e.g., see Cramér [15]), although it could be shown that an unbiased estimator does not always exist (see, e.g., Lehmann and Casella [12, p. 83]), and lower risk function values (6) can be achieved when introducing a suitable bias (Eldar [16]). In this context, the Minimum Variance Unbiased Estimator (MVUE) is of special interest, since it represents the efficient unbiased estimator.
- (ii) *Minimaxity*, that is, the minimization of the maximum value of the risk function of (6), which typically leads to rather conservative estimates, since they can be approximated by Bayesian estimators with a prior distributions concentrated at the worst case (Robert [13, p. 67]). Unfortunately, these estimates do not always exist (Robert [13, p. 69]).
- (iii) *Equivariance*, that is, the invariance of the estimation problem under certain transformations of the

estimator. This property is important for cases in which a given estimator is applied to modified estimation situations, but the subsequent investigation of the optimality of the modified estimator should be avoided. This is important for practical cases, in which the estimator is modified routinely or a certain symmetry of the estimator is demanded by the nature of the parameters one wants to estimate. For details, see Lehmann and Casella [12, p. 147–225].

3. Application to Stochastic Localization in Euclidean Space

The general Bayesian framework described in the previous section will be applied to a localization problem in which 2D (3D) objects are projected onto 1D (2D) planes, and the projection signals (images) are detected for multiple views around these objects. This localization problem will be referred to as multiple view triangulation close to interfaces with a 2D or 3D observation model. In principle, multiple projections can be measured for views adjacent to each other and forming a contiguous arc or could be separated by larger angles (e.g., 90 deg). The formalism works for all sets of projections and all input data containing information about the projected coordinates of the objects. Two cases are presented in the following: linear parallel projections of 2D objects and nonlinear cone beam projection of 3D objects in a convex solution space Ω_X . All transformations are purposefully presented in Euclidean space (instead of the projective space with homogeneous coordinates [1]) in order to allow intuitive interpretations of the model and the results, as well as demonstrate direct solutions for the nonlinear 3D case, and utilize straightforwardly interpretable priors. Further, we exactly know the geometry of image acquisition (i.e., use calibrated cameras) in order to specifically focus on the noisy measurements and the incorporation of prior knowledge, instead of the challenging problem of camera calibration [1, pp. 434–457].

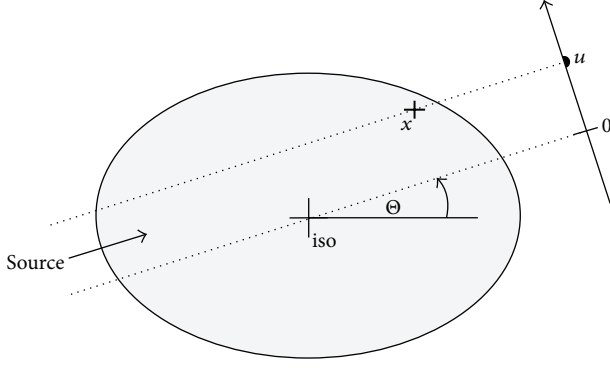


FIGURE 3: Geometrical projection $u \in \mathbb{R}$ of the coordinate $\mathbf{x} \in \mathbb{R}^2$ in the detector plane.

The 2D geometry is presented in Figure 3 and shows a region which contains a feature within a larger object with interfaces, whose coordinates $\mathbf{x} \in \Omega_{\mathbf{x}}$ we want to estimate. It is known that \mathbf{x} might be at a different position, but it cannot be outside the region (e.g., is constrained by the surface of the object or regional interfaces). Furthermore, the PDFs are assumed to be Gaussians or truncated Gaussians (or uniform distributions in the Results section), since these are valid assumptions in many applications [1], and the resulting estimators can partly be presented in terms of explicit relationships. In general, the estimation we present works for arbitrary PDFs and can be directly applied to nonlinear/non-Gaussian problems.

Constraints due to interfaces or surfaces, for example, might occur in radiotherapy (or interventional radiology, image-guided surgical procedure) where the location of an implanted fiducial marker or another feature has to be determined in real-time based on X-ray projections (radiographs or tomosynthesis projections) [3, 4]. Let us focus on a specific application of breast cancer patient setup in radiotherapy. In this case, surgical clips implanted during tumor resection are used for localization of the tumor bed. Due to deformations of the breast in between the original CT images and the treatment, the markers might be at arbitrary locations, but they will never be outside of the patient's body, and they will not be inside the chest defined by the ribs. Also it is likely that they will not be displaced by more than a given range known from clinical practice (e.g., 2 cm from the original location in the planning CT scan). In radiotherapy, the goal is to localize the tumor (or the tumor surrogate such as the marker) with 1 mm accuracy. This is achievable by performing a CT scan. But such a CT scan is not possible in many clinical situations either due to undesired dose from the imaging or due to time constraints or other practical concerns. Additionally, in the case of a given marker within the flexible breast tissue, the surface can be measured with optical or infrared cameras in real-time or based on a tomosynthesis reconstruction, which is an intensely discussed topic [17–19]. We can thus assume that the knowledge about the breast surface is available for the given posture of the patient. In a similar context, in interventional radiology or image based surgeries there is a need to work with a few projections or even single images

in order to find the location of certain objects in 3D. Many other applications are thinkable, such as robotic arm position estimation in a closed room or computer vision approaches with spatial restrictions [1].

3.1. Parallel Beam in 2D. In this case $\mathbf{x} \in \Omega_{\mathbf{x}} \subset \mathbb{R}^2$. The observation $u \in \mathbb{R}$ is performed by projecting the feature in a detector plane at different angles Θ . The aim of this example is to estimate the coordinate \mathbf{x} based on the observed coordinates u at N different angles between 0° and 90° , which are merged to a single vector $\mathbf{u} = (u_1, \dots, u_N)^T \in \Omega_{\mathbf{u}} \subset \mathbb{R}^N$.

Please note that although the presented situation might have several applications, the numerical example in the Results section uses arbitrary dimensionless parameters in order to demonstrate the principal differences of the estimation methods. Furthermore, the linear relation between observation and the coordinate in the parallel beam model allows to derive explicit estimators, which is not possible for the nonlinear cone beam geometry in Section 3.2.

3.1.1. Observation without Uncertainties and Prior Knowledge.

If it is assumed that the observation is without any errors, then no prior knowledge is necessary because there is no ambiguity in the detection, and the linear relation between \mathbf{x} and a single observation u can be derived from

$$u_i := -x_1 \sin \Theta_i + x_2 \cos \Theta_i. \quad (17)$$

This is a linear equation in the two coordinates of \mathbf{x} and; therefore, two equations need to be derived for a unique solution, that is, two observations u_1 and u_2 at angles Θ_1 and Θ_2 with $\Theta_1 \neq \Theta_2$ modulo 360° . The explicit solution for this straightforward estimation is given by the estimator $\delta(\mathbf{u})$

$$\delta(\mathbf{u}) := \begin{pmatrix} -\sin \Theta_1 & \cos \Theta_1 \\ -\sin \Theta_2 & \cos \Theta_2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. \quad (18)$$

In the numerical examples we will assume the two angles to be $\Theta_1 = 0^\circ$ and $\Theta_2 = 90^\circ$ which represents the best separation between two projections. In the further text the estimator of (18) will be denoted by the 2×2 approach.

3.1.2. Observation with Uncertainties and without Prior Knowledge.

If we know that there are observational uncertainties, that is, errors in determining \mathbf{u} , then one can assume an error distribution, which is represented in this example by the normally distributed random variable at angle Θ_i

$$\Delta U_i \sim \mathcal{N}(0, \sigma_{\Delta U}^2). \quad (19)$$

In the geometrical model, this means we are adding a random error to the observed projected coordinate at every angle

$$U_i := -x_1 \sin \Theta_i + x_2 \cos \Theta_i + \Delta U_i, \quad (20)$$

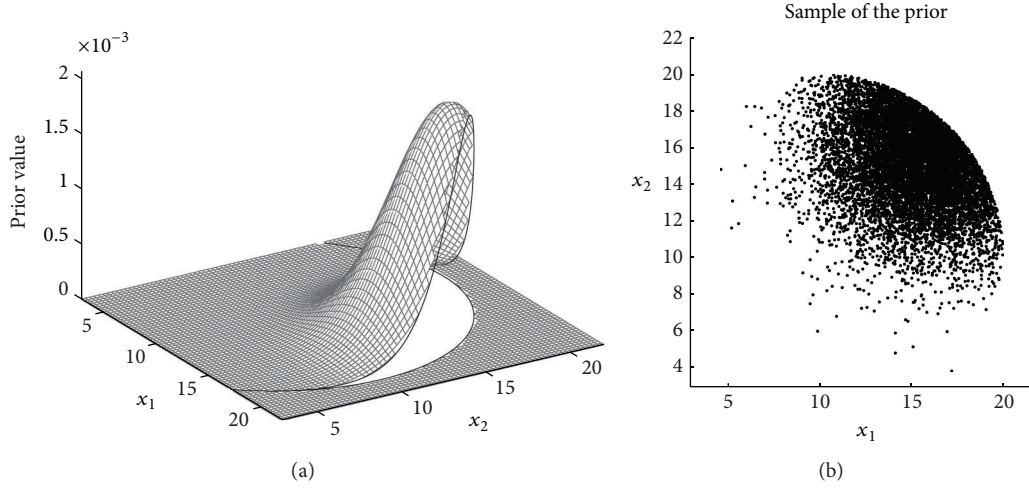


FIGURE 4: (a) Example for the prior density function with the parameters specified in the Results section. (b) 10,000 samples drawn from the prior distribution which are utilized in the Results section.

which makes now the observation itself to a random variable with the PDF

$$f_{[U_i]}(u_i | \mathbf{x}) = \frac{1}{\sigma_{\Delta U} \sqrt{2\pi}} \times \exp \left[-\frac{(u_i - (-x_1 \sin \Theta_i + x_2 \cos \Theta_i))^2}{2\sigma_{\Delta U}^2} \right]. \quad (21)$$

By acquiring $N \geq 2$ stochastically independent observations \mathbf{u} at angles $\Theta_1, \dots, \Theta_N$, one can summarize these observations in a single PDF, which is the 2D *likelihood function* in \mathbf{x}

$$\begin{aligned} f_{[U]}(\mathbf{u} | \mathbf{x}) &= \prod_{i=1}^N f_{[U_i]}(u_i | \mathbf{x}) \\ &= \frac{1}{(\sigma_{\Delta U} \sqrt{2\pi})^N} \\ &\quad \times \exp \left[-\sum_{i=1}^N \frac{(u_i - (-x_1 \sin \Theta_i + x_2 \cos \Theta_i))^2}{2\sigma_{\Delta U}^2} \right]. \end{aligned} \quad (22)$$

In consequence, by applying the logarithm the *Maximum Likelihood* estimator is determined by

$$\delta(\mathbf{u}) := \arg \min_{\mathbf{x}} \sum_{i=1}^N (u_i - (-x_1 \sin \Theta_i + x_2 \cos \Theta_i))^2, \quad (23)$$

which is exactly the *least squares* solution. This is a well-known result in estimation theory and indicates the success of *least squares*, since the sum of many independent errors can often be approximated well by a normal distribution [1, pp. 102, 121, 134, 434, and 568]. An important observation is that in this approach the solution is totally independent of

the knowledge about the measurement uncertainty $\sigma_{\Delta U}$. The explicit solution is derived from solving (23) analytically, and it follows (cp. (18))

$$\delta(\mathbf{u}) = \begin{pmatrix} \sum_{i=1}^N \sin^2 \Theta_i & -\sum_{i=1}^N \sin \Theta_i \cos \Theta_i \\ -\sum_{i=1}^N \sin \Theta_i \cos \Theta_i & \sum_{i=1}^N \cos^2 \Theta_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} -\sum_{i=1}^N u_i \sin \Theta_i \\ \sum_{i=1}^N u_i \cos \Theta_i \end{pmatrix}. \quad (24)$$

3.1.3. Observation with Uncertainties and Prior Knowledge. If it is further known that \mathbf{x} is stochastically distributed, that is, a random variable \mathbf{X} , one can additionally introduce prior knowledge about this distribution into the estimation model. In the 2D example the prior is defined as follows

$$\mathbf{X} \sim \pi(\mathbf{x}) := a \cdot \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}})(\mathbf{x}) \cdot \chi_{C(\mathbf{m}, r)}(\mathbf{x}), \quad (25)$$

with a normalization constant a , the characteristic function χ which has the value 1 in the circle $C(\mathbf{m}, r)$ with midpoint \mathbf{m} and radius r , and 0 elsewhere (representing an example of a convex region), and $\boldsymbol{\mu}_{\mathbf{X}} \in \mathbb{R}^2$ the mean of the normal distribution, and

$$\Sigma_{\mathbf{X}} := \begin{pmatrix} \sigma_{X_1}^2 & 0 \\ 0 & \sigma_{X_2}^2 \end{pmatrix}. \quad (26)$$

This is essentially a Gaussian density function which is truncated by the circle $C(\mathbf{m}, r)$, which is illustrated in Figure 4. In the geometrical model now also the parameters become random variables; that is,

$$U_i := -X_1 \sin \Theta_i + X_2 \cos \Theta_i + \Delta U_i. \quad (27)$$

Utilizing Bayes' theorem, the *posterior* PDF can be calculated by (c is the normalization constant of the posterior)

$$\pi(\mathbf{x} | \mathbf{u}) = c \cdot f_{[U]}(\mathbf{u} | \mathbf{x}) \cdot \pi(\mathbf{x}) = c \cdot \prod_{i=1}^N f_{[U_i]}(u_i | \mathbf{x}) \cdot \pi(\mathbf{x}) \quad (28)$$

$$= \tilde{c} \cdot \exp \left[- \sum_{i=1}^N \frac{(u_i - (-x_1 \sin \Theta_i + x_2 \cos \Theta_i))^2}{2\sigma_{\Delta U}^2} - \sum_{j=1}^2 \frac{(x_j - \mu_{X_j})^2}{2\sigma_{X_j}^2} \right] \cdot \chi_{C(\mathbf{m}, r)}(\mathbf{x}), \quad (29)$$

$$\tilde{c} := \frac{ca}{(\sigma_{\Delta U})^N \sigma_{X_1} \sigma_{X_2} (2\pi)^{N/2+1}} > 0. \quad (30)$$

Based on the posterior the *Maximum A Posteriori* estimator is therefore defined as

$$\delta(\mathbf{u}) := \arg \min_{\mathbf{x} \in C(\mathbf{m}, r)} \sum_{i=1}^N \frac{(u_i - (-x_1 \sin \Theta_i + x_2 \cos \Theta_i))^2}{\sigma_{\Delta U}^2} + \sum_{j=1}^2 \frac{(x_j - \mu_{X_j})^2}{\sigma_{X_j}^2}. \quad (31)$$

This formula essentially shows explicitly how the compromise between measurement and prior knowledge, as two different sources of information, is gained: the measurement is weighted with $1/\sigma_{\Delta U}^2$ against the prior knowledge with $1/\sigma_{X_j}^2$ ($j = 1, 2$), which is the case of the Tikhonov regularization for the constrained least squares problem. Since this is a convex function on a convex domain $C(\mathbf{m}, r)$, there is a guaranteed global minimum. If this maximum is in the interior of $C(\mathbf{m}, r)$, the explicit solution is determined by solving (31), and it follows (cp. (18) and (24))

$$\delta(\mathbf{u}) = \begin{pmatrix} \sum_{i=1}^N \sin^2 \Theta_i + \left(\frac{\sigma_{\Delta U}}{\sigma_{X_1}}\right)^2 & -\sum_{i=1}^N \sin \Theta_i \cos \Theta_i \\ -\sum_{i=1}^N \sin \Theta_i \cos \Theta_i & \sum_{i=1}^N \cos^2 \Theta_i + \left(\frac{\sigma_{\Delta U}}{\sigma_{X_2}}\right)^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} -\sum_{i=1}^N u_i \sin \Theta_i + \mu_{X_1} \left(\frac{\sigma_{\Delta U}}{\sigma_{X_1}}\right)^2 \\ \sum_{i=1}^N u_i \cos \Theta_i + \mu_{X_2} \left(\frac{\sigma_{\Delta U}}{\sigma_{X_2}}\right)^2 \end{pmatrix}, \quad (32)$$

From a practical perspective, if the $\delta(\mathbf{u})$ calculated by (32) is outside of $C(\mathbf{m}, r)$, then the maximum of (31) must be on the border (circle in Figure 4), which can then be solved by *Lagrangian multipliers* applied to (31) or directly by solving (31) with a general optimization algorithm. The explicit

solution for the interior of $C(\mathbf{m}, r)$ depends now on the ratios of $\sigma_{\Delta U}/\sigma_{X_i}$ ($i \in \{1, 2\}$), which again shows a weighting of the measurement uncertainties and the prior knowledge. Particularly, if $\sigma_{X_{1/2}} \rightarrow \infty$, the prior gets noninformative, and the interior estimator converges to the ML estimator.

Based on the posterior also the Minimum *Mean Square Error* estimator can be determined by

$$\delta_i(\mathbf{u}) := \mathbb{E}[X_i | \mathbf{u}] = \iint_{C(\mathbf{m}, r)} x_i \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x} \quad (33)$$

for $i \in \{1, 2\}$ with the posterior in (29), which cannot be calculated elementarily anymore. For the specific case of $r \rightarrow \infty$, the MMSE estimator is coinciding with the MAP estimator of (32) due to the symmetry of the posterior around its global maximum when using Gaussians. The difference between MAP and MMSE will therefore be enhanced by the estimation close to the boundary. Please note that both MAP and MMSE guarantee an estimate which is not outside of the (convex) circle $C(\mathbf{m}, r)$, effectively bounding the feasibility region with the prior. In contrast, for the 2×2 and ML approach no bounded feasibility region is possible.

3.2. Cone Beam in 3D. In this case $\mathbf{x} \in \Omega_{\mathbf{x}} \subset \mathbb{R}^3$. The observation $\mathbf{u} \in \mathbb{R}^2$ is again performed by projecting the feature in a detector plane at different angles Θ . There are N observed coordinates \mathbf{u} at different angles between 0° and 90° , which are merged to a single matrix $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_N)^T \in \Omega_U \subset \mathbb{R}^{2 \times N}$ (which can be identified with a vector in \mathbb{R}^{2N}).

If it is assumed that the observation is without error and no prior knowledge is available to the estimation, the nonlinear relation between \mathbf{x} and a single observation \mathbf{u}_i at angle Θ_i is

$$\mathbf{u}_i = \boldsymbol{\psi}_i(\mathbf{x}) := \frac{g + h}{x_1 \cos \Theta_i + x_2 \sin \Theta_i + g} \times \begin{pmatrix} -x_1 \sin \Theta_i + x_2 \cos \Theta_i \\ x_3 \end{pmatrix}, \quad (34)$$

with the imaging source to isocenter of rotation distance g and isocenter to detection plane distance h ; see Figure 5 for an illustration of this geometry. This projection excludes points \mathbf{x} which are on the plane $x_1 \cos \Theta_i + x_2 \sin \Theta_i + g = 0$ orthogonal to the beam through the imaging source. It is directly recognizable that $\boldsymbol{\psi}_i$ is nonlinear (in Euclidean space), since $\boldsymbol{\psi}_i(\mathbf{x} + \mathbf{y}) \neq \boldsymbol{\psi}_i(\mathbf{x}) + \boldsymbol{\psi}_i(\mathbf{y})$. We recognize that this projection can be represented as a linear transformation in homogeneous coordinates (e.g., see Hartley and Zisserman [1]), but we purposefully ignore this in order to show how one can apply the proposed estimation approaches directly to a nonlinear situation as well as define intuitive priors in Euclidean space.

A direct solution, such as the 2×2 case in 2D, is not uniquely possible since there are equation numbers as a multiple of 2, but we have 3 unknowns. We avoid this ambiguity by not defining a corresponding 3×3 case.

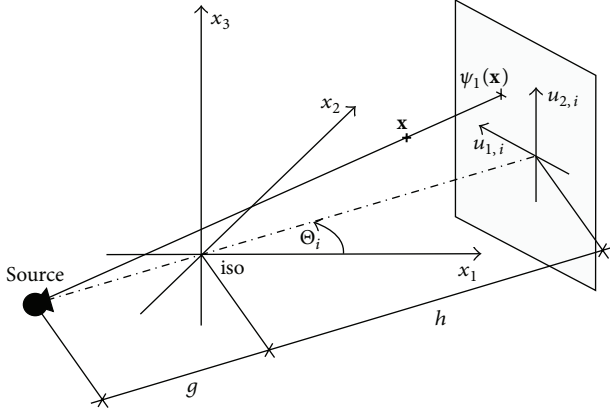


FIGURE 5: Geometrical projection $\mathbf{u} \in \mathbb{R}^2$ of the coordinate $\mathbf{x} \in \mathbb{R}^3$ in the detector plane.

3.2.1. Observation with Uncertainties and without Prior Knowledge. The observation uncertainty at angle Θ_i is in this case described by

$$\Delta \mathbf{U}_i \sim \mathcal{N}(0, \sigma_{\Delta U} \cdot \mathbf{I}_{2 \times 2}), \quad (35)$$

with the 2×2 identity matrix $\mathbf{I}_{2 \times 2}$. The geometrical model is therefore modified to

$$\mathbf{U}_i := \boldsymbol{\psi}_i(\mathbf{x}) + \Delta \mathbf{U}_i. \quad (36)$$

In consequence, the observation PDF is

$$f_{[\mathbf{U}_i]}(\mathbf{u}_i | \mathbf{x}) = \frac{1}{\sigma_{\Delta U}^2 2\pi} \exp \left[-\sum_{k=1}^2 \frac{(u_{k,i} - \psi_{k,i}(\mathbf{x}))^2}{2\sigma_{\Delta U}^2} \right]. \quad (37)$$

By acquiring $N \geq 2$ stochastically independent observations \mathbf{u}_i at angles $\Theta_1, \dots, \Theta_N$, one can summarize these observation in a single PDF, which is the 3D *likelihood function* in \mathbf{x}

$$f_{[\mathbf{U}]}(\mathbf{u} | \mathbf{x}) = \frac{1}{(\sigma_{\Delta U}^2 2\pi)^N} \times \exp \left[-\sum_{i=1}^N \sum_{k=1}^2 \frac{(u_{k,i} - \psi_{k,i}(\mathbf{x}))^2}{2\sigma_{\Delta U}^2} \right]. \quad (38)$$

This leads then to the *Maximum Likelihood estimator*

$$\begin{aligned} \boldsymbol{\delta}(\mathbf{u}) &:= \arg \min_{\mathbf{x}} \\ &\times \sum_{i=1}^N \left(u_{1,i} - \frac{(-x_1 \sin \Theta_i + x_2 \cos \Theta_i)(g+h)}{x_1 \cos \Theta_i + x_2 \sin \Theta_i + g} \right)^2 \\ &+ \left(u_{2,i} - \frac{x_3(g+h)}{x_1 \cos \Theta_i + x_2 \sin \Theta_i + g} \right)^2, \end{aligned} \quad (39)$$

which is the nonlinear *least squares* solution. Due to the nonlinearity, no elementary solution can be derived, and this expression is directly minimized.

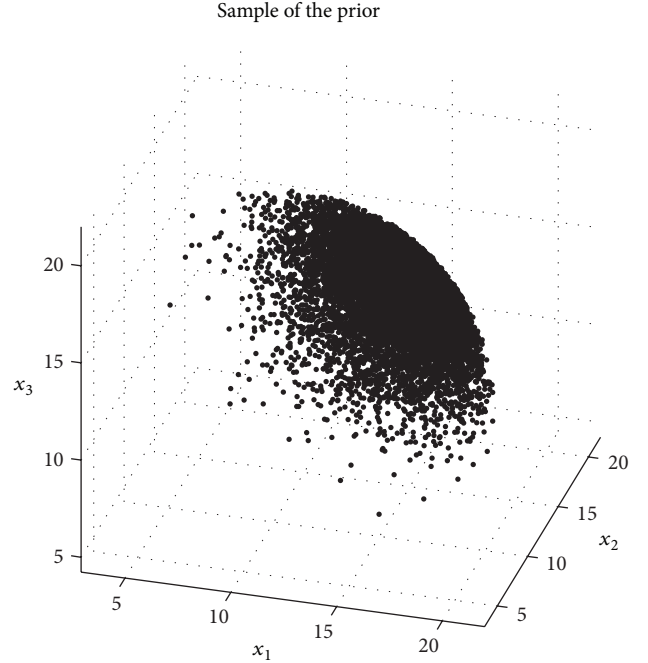


FIGURE 6: 10,000 samples drawn from the prior distribution which are utilized in the Results section.

3.2.2. Observation with Uncertainties and Prior Knowledge. The prior knowledge is in the 3D case modeled with

$$\mathbf{X} \sim \pi(\mathbf{x}) := a \cdot \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_X)(\mathbf{x}) \cdot \chi_{S(\mathbf{m}, r)}(\mathbf{x}), \quad (40)$$

with a normalization constant a , the characteristic function χ which has the value 1 in the sphere $S(\mathbf{m}, r)$ with midpoint \mathbf{m} and radius r , and 0 elsewhere (again as an example of a convex region), and $\boldsymbol{\mu}_X \in \mathbb{R}^3$ the mean of the normal distribution, and

$$\Sigma_X := \begin{pmatrix} \sigma_{X_1}^2 & 0 & 0 \\ 0 & \sigma_{X_2}^2 & 0 \\ 0 & 0 & \sigma_{X_3}^2 \end{pmatrix}. \quad (41)$$

This is essentially a Gaussian density function which is truncated by the sphere $S(\mathbf{m}, r)$. A sample from this prior is illustrated in Figure 6. The geometrical model is then

$$\mathbf{U}_i := \boldsymbol{\psi}_i(\mathbf{X}) + \Delta \mathbf{U}_i. \quad (42)$$

The posterior distribution then gets (c is the normalization constant of the posterior)

$$\begin{aligned} \pi(\mathbf{x} | \mathbf{u}) &= \tilde{c} \cdot \exp \left[-\sum_{i=1}^N \sum_{k=1}^2 \frac{(u_{k,i} - \psi_{k,i}(\mathbf{x}))^2}{2\sigma_{\Delta U}^2} - \sum_{j=1}^3 \frac{(x_j - \mu_{X_j})^2}{2\sigma_{X_j}^2} \right] \\ &\cdot \chi_{S(\mathbf{m}, r)}(\mathbf{x}), \end{aligned} \quad (43)$$

$$\tilde{c} := \frac{c a}{(\sigma_{\Delta U})^{2N} \sigma_{X_1} \sigma_{X_2} \sigma_{X_3} (2\pi)^{N+3/2}} > 0. \quad (44)$$

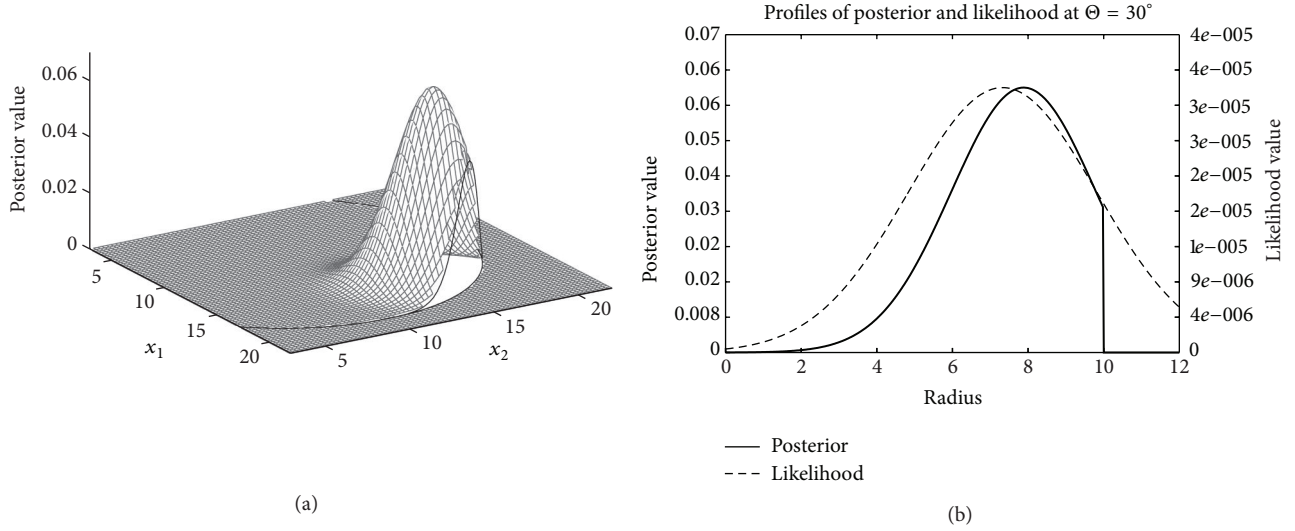


FIGURE 7: (a) Posterior density function for a specific estimation of case (A) and full prior. (b) Profile of the likelihood and posterior function at $\mathbf{x} = r \cdot (\cos \pi/6, \sin \pi/6)^T$ with $r \in [0, 12]$.

In consequence, the *Maximum A Posteriori* estimator is

$$\begin{aligned} \delta(\mathbf{u}) &:= \arg \min_{\mathbf{x} \in S(\mathbf{m}, r)} \\ &\times \sum_{i=1}^N \left(\left(u_{1,i} - \frac{(-x_1 \sin \Theta_i + x_2 \cos \Theta_i)(g+h)}{x_1 \cos \Theta_i + x_2 \sin \Theta_i + g} \right)^2 \right. \\ &\quad \left. + \left(u_{2,i} - \frac{x_3(g+h)}{x_1 \cos \Theta_i + x_2 \sin \Theta_i + g} \right)^2 \right) \\ &\times (\sigma_{\Delta U}^2)^{-1} + \sum_{j=1}^3 \frac{(x_j - \mu_{X_j})^2}{\sigma_{X_j}^2}, \end{aligned} \quad (45)$$

representing a constrained and regularized nonlinear least squares optimization problem. The corresponding *Minimum Mean Square Error* estimator is analogously

$$\delta_i(\mathbf{u}) := \mathbb{E}[X_i | \mathbf{u}] = \iiint_{S(\mathbf{m}, r)} x_i \cdot \pi(\mathbf{x} | \mathbf{u}) d\mathbf{x}, \quad (46)$$

for $i \in \{1, 2, 3\}$ with the posterior of (43), which again cannot be calculated elementarily anymore.

4. Results

In order to show the practical differences, we present 5 different cases for the comparison of the estimators with 10,000 sample points \mathbf{x} each (recall that \mathbf{X} is a random variable), in 2D parallel beam and 3D cone beam geometry. The standard parameter set (A) is as follows: with $N = 5$ projections equidistantly distributed in between 0° and 90° and $\sigma_{\Delta U} = 3$, $\sigma_{X_1} = \sigma_{X_2} = \sigma_{X_3} = 3$, $\mu_{\mathbf{X}} = (16.5, 16.5)^T$, or $\mu_{\mathbf{X}} = (16.5, 16.5, 16.5)^T$, respectively, $\mathbf{m} = (10, 10)^T$ or $\mathbf{m} = (10, 10, 10)^T$ respectively, $r = 10$, $g = 100$ and $h = 50$.

Besides the standard parameter set presented above, four other parameter sets (cases) are investigated (by changing only one parameter at a time while setting all other parameters to the standard): (B) applying only $N = 2$ projections, (C) applying $N = 10$ projections, (D) narrowing the prior by $\sigma_{X_i} = 1.5$, and (E) reducing the measurement uncertainties $\sigma_{\Delta U} = 1.5$. In addition, for the 2D parallel beam case, we present the results of the MAP and MMSE estimations if there is a noninformative prior (np) utilized, that is, uniform distribution inside the circle $C(\mathbf{m}, r)$ and zero outside,

$$\sigma_{X_i} \rightarrow \infty, \quad i \in \{1, 2\} \implies \pi(\mathbf{x}) := \frac{1}{r^2 \pi} \cdot \chi_{C(\mathbf{m}, r)}(\mathbf{x}), \quad (47)$$

which are denoted with $\text{MAP}^{\text{np}} := \lim_{\sigma_{X_i} \rightarrow \infty} \text{MAP}$ and MMSE^{np} accordingly. This means, only the hard constraints about the feasibility region are utilized in these estimations, without prior knowledge about the concentration of \mathbf{x} at the border (circle in Figure 4). In a broader context, such a noninformative MAP^{np} estimator can be identified with a ML estimator that has ad hoc constraints about the solution space in optimization.

4.1. Parallel Beam in 2D. For the standard parameters of case (A), in Figure 7 an example of the posterior PDF in 2D for a single sample point \mathbf{x} is presented. In this plot on the right side a profile of the likelihood and the posterior function is presented, which demonstrates the general behavior: the width of the posterior is reduced compared to the likelihood function by the cost of introducing a shift (bias) of the maximum position by prior knowledge. Further, in the posterior profile the circular boundary is presented by the truncation of the profile. By incorporating the truncation into the prior distribution, the MMSE (which for simplicity one

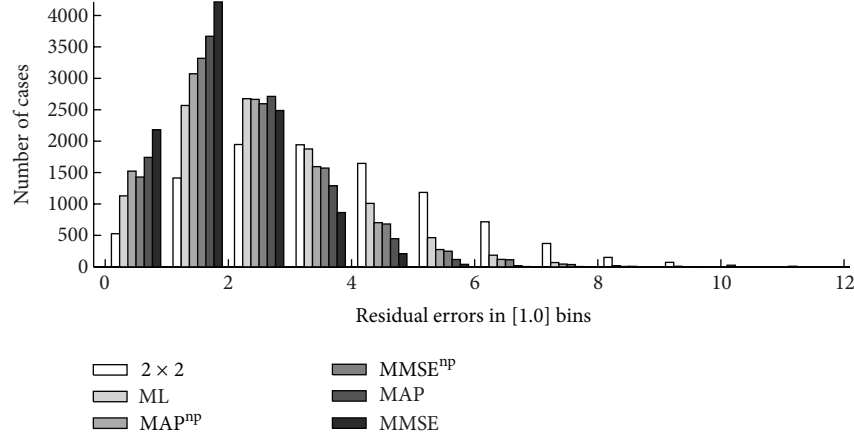


FIGURE 8: Histogram of the residual radial errors of 2D parallel beam geometry for case (A) after estimation with the 2×2 approach (radial RMSE: 4.24), ML (3.05), MAP^{np} (2.71), MMSE^{np} (2.67), MAP (2.33), and MMSE (2.03).

can think of as the expectation value of the posterior profile) is compensating for the bias of the maximum position in MAP.

In Figure 8 the estimation performance, for case (A) of the 2×2 approach of (18), the ML of (24), the MAP of (32) and (31), and the MMSE of (33) (with noninformative prior and truncated Gaussian prior) are presented. The complete estimation results for all five cases (A)–(E) are presented in Table 2. In addition to Table 2 the RMSEs of a broader variation of the essential parameters are illustrated in Figure 9 by starting with case (A) and varying only the specified parameter of each subfigure. Examples for the posterior distributions cases (A)–(E) (for the same specific sample \mathbf{x}) are presented in Figure 10, in order to see the qualitative differences due to the according changes in the parameter set.

Based on Table 2 and Figure 9 several characteristic observations can be made: at first, in all cases the estimation accuracy, which can be summarized with the RMSE, is improving mainly along the chain $2 \times 2 \rightarrow \text{ML} \rightarrow \text{MAP}^{\text{np}} \rightarrow \text{MMSE}^{\text{np}} \rightarrow \text{MAP} \rightarrow \text{MMSE}$.

Second, in order to compare the empirical bias of each estimator one has to focus on the mean values of the coordinate directions x_1 and x_2 which are presented in brackets in the third column of Table 2. In theory, the 2×2 and the ML approaches are (asymptotically) unbiased, which was also observed in this empirical evaluation (i.e., a bias of 0.0 in each coordinate direction). MAP and MMSE estimators are typically biased estimators, which is certainly true in this application for the MAP estimation and partly true for the MMSE estimator. In detail, for the estimators with uniform priors, MAP^{np} and MMSE^{np} , a bias towards the center of the constraining circle can be recognized (e.g., negative bias values in each coordinate direction). This is plausible, since estimates outside the circle are prohibited with this prior and, on the other side, no information about the concentration of \mathbf{x} at the border of the circle is introduced. When focusing on the full prior (constraining circle and the concentration at the border) with MAP, a bias in the opposite direction, towards the border, is introduced (e.g.,

see Figure 7 for an illustration of this effect). Please note that the introduction of a significant bias in these three estimations, MAP^{np} , MMSE^{np} , and MAP, still leads to better results compared to the unbiased 2×2 and ML estimations, which is a characteristic behavior in Bayesian estimation. Remarkably, the use of the Bayesian MMSE with the full prior leads to an empirically unbiased estimator in this application. This can be understood by realizing that the bias is in balance between the concentration of \mathbf{x} at the border (leading to a bias towards the border) and the constraints of the border (leading to a bias towards the center of the circle) when averaging over the whole posterior distribution (e.g., see Figure 7 again for an illustration). This allows to conclude that for the spatial estimation at constraining interfaces a highly accurate and practically unbiased estimator can be constructed with MMSE estimation.

Third, compared to the standard parameter set, utilizing $N = 2$ projections reduces the observation information and increases the RMSE. Specifically, ML is converging exactly to the 2×2 approach, increasing its RMSE from 3.05 to 4.28. Compared to the uniform prior in MAP^{np} and MMSE^{np} , the full prior in MAP and MMSE leads to robust estimates with much less increasing RMSEs.

Fourth, by utilizing $N = 10$ projections, the performances of all estimators are improved. Essentially, the relative use of the prior knowledge is decreased since it is weighted less by the MAP and MMSE estimators, which, for example, can be seen in (32).

Fifth, when the sample distribution is getting narrower and simultaneously the corresponding prior PDF is utilized, the RMSEs of MAP and MMSE decrease significantly. The relative benefit from MAP to MMSE is decreased in this case, since, for the narrower prior, the influence of the truncation of the posterior by the boundary is reduced. On the other side, MAP^{np} and MMSE^{np} lead to RMSEs much closer to ML, since they introduce a strong bias (not knowing the concentration of \mathbf{x} at the border). In this context, the increase in the bias is much stronger for MMSE^{np} compared to MAP^{np} , leading overall to even worse RMSE values.

TABLE 2: Statistical values about the estimation results of 2D parallel beam geometry for cases (A)–(E) with $M = 10,000$ sample points \mathbf{x} each are presented for the six 2D estimators. The radial estimation error is presented at first and the coordinate estimation errors of (x_1, x_2) are shown in brackets. In the following the formulas for the statistical values are presented for the radial error: the *mean value* $= \|\delta(\mathbf{u})\| - \|\mathbf{x}\| = 1/M \sum_{j=1}^M \|\delta(\mathbf{u}_j)\| - \|\mathbf{x}_j\|$, the *standard deviation* $= (1/(M-1) \sum_{j=1}^M (\|\delta(\mathbf{u}_j)\| - \|\delta(\mathbf{u})\| - \|\mathbf{x}_j\| + \|\mathbf{x}\|)^2)^{1/2}$, the *absolute maximum* $= \max_{j=1 \dots M} \|\delta(\mathbf{u}_j)\| - \|\mathbf{x}_j\|$, the *Root Mean Square Error (RMSE)* $= (1/M \sum_{j=1}^M (\|\delta(\mathbf{u}_j)\| - \|\mathbf{x}_j\|)^2)^{1/2}$ and the *gain of method X relative to ML* $= (\text{RMSE}_{\text{ML}} - \text{RMSE}_X) / \text{RMSE}_{\text{ML}}$.

Case	Method	Mean	st. deviation	abs. maximum	Radial RMSE	[%]
A	2×2	3.77 (−0.0,0.0)	1.94 (3.0,3.0)	11.90 (11.1,10.9)	4.24 (3.0,3.0)	−39
	ML	2.67 (−0.0,0.0)	1.48 (2.2,2.2)	9.75 (9.2,8.3)	3.05 (2.2,2.2)	—
	MAP ^{np}	2.35 (−0.3,−0.3)	1.35 (1.9,1.9)	9.75 (7.5,8.3)	2.71 (1.9,1.9)	+11
	MMSE ^{np}	2.31 (−0.9,−0.9)	1.32 (1.6,1.7)	9.65 (7.5,8.2)	2.67 (1.9,1.9)	+13
	MAP	2.05 (0.6,0.6)	1.11 (1.5,1.5)	8.38 (6.5,6.3)	2.33 (1.7,1.7)	+24
	MMSE	1.79 (0.0,0.0)	0.96 (1.4,1.4)	7.93 (6.1,5.8)	2.03 (1.4,1.4)	+33
B	ML	3.79 (−0.0,−0.0)	1.99 (3.0,3.0)	13.43 (13.2,11.7)	4.28 (3.0,3.0)	—
	MAP ^{np}	3.29 (−0.4,−0.4)	1.83 (2.7,2.6)	12.77 (12.6,11.5)	3.76 (2.7,2.7)	+12
	MMSE ^{np}	3.10 (−1.2,−1.2)	1.68 (2.2,2.2)	10.99 (10.3,10.9)	3.52 (2.5,2.5)	+18
	MAP	2.56 (0.7,0.7)	1.34 (1.9,1.9)	8.95 (8.8,8.3)	2.89 (2.0,2.1)	+33
	MMSE	2.28 (0.0,0.0)	1.20 (1.8,1.8)	8.84 (8.6,8.1)	2.58 (1.8,1.8)	+40
C	ML	1.99 (−0.0,−0.0)	1.13 (1.6,1.6)	8.40 (7.0,7.2)	2.29 (1.6,1.6)	—
	MAP ^{np}	1.81 (−0.2,−0.2)	1.05 (1.5,1.5)	7.00 (5.8,6.2)	2.09 (1.5,1.5)	+9
	MMSE ^{np}	1.76 (−0.6,−0.6)	1.02 (1.3,1.3)	7.00 (5.8,6.2)	2.03 (1.4,1.5)	+11
	MAP	1.67 (0.4,0.4)	0.94 (1.3,1.3)	7.18 (5.5,5.7)	1.92 (1.4,1.4)	+16
	MMSE	1.48 (0.0,0.0)	0.82 (1.2,1.2)	6.95 (5.0,5.6)	1.70 (1.2,1.2)	+26
D	ML	2.68 (0.0,0.0)	1.48 (2.2,2.2)	9.47 (7.8,9.0)	3.06 (2.2,2.2)	—
	MAP ^{np}	2.23 (−0.4,−0.4)	1.32 (1.8,1.8)	9.47 (7.4,7.9)	2.59 (1.8,1.8)	+15
	MMSE ^{np}	2.35 (−1.2,−1.2)	1.32 (1.5,1.5)	9.45 (7.5,7.9)	2.69 (1.9,1.9)	+12
	MAP	1.38 (0.4,0.4)	0.76 (1.1,1.1)	5.78 (5.0,4.4)	1.57 (1.1,1.1)	+49
	MMSE	1.29 (0.0,0.0)	0.67 (1.0,1.0)	5.69 (4.6,4.2)	1.45 (1.0,1.0)	+53
E	2×2	1.88 (0.0,0.0)	0.99 (1.5,1.5)	6.48 (6.4,6.2)	2.13 (1.5,1.5)	−40
	ML	1.33 (0.0,−0.0)	0.74 (1.1,1.1)	4.93 (3.8,4.2)	1.52 (1.1,1.1)	—
	MAP ^{np}	1.24 (−0.1,−0.1)	0.70 (1.0,1.0)	4.93 (3.8,4.2)	1.42 (1.0,1.0)	+7
	MMSE ^{np}	1.20 (−0.3,−0.3)	0.68 (0.9,0.9)	4.93 (3.8,4.2)	1.38 (1.0,1.0)	+9
	MAP	1.22 (0.2,0.2)	0.67 (1.0,1.0)	4.74 (3.5,4.6)	1.39 (1.0,1.0)	+9
	MMSE	1.11 (0.0,0.0)	0.61 (0.9,0.9)	4.72 (3.4,4.6)	1.27 (0.9,0.9)	+16

Sixth, if the measurement accuracy is increased, all estimators benefit from it significantly. Particularly, the benefit of using (partial) prior knowledge (in MAP^{np}, MMSE^{np}, MAP, and MMSE) compared to not using it (in 2×2 and ML) is decreased but still not negligible.

4.2. Cone Beam in 3D. In Figure 11 the estimation performances for this standard parameter set of the ML of (39), the MAP of (45), and the MMSE of (46) are presented. Further, in Table 3, the results of the three estimators are presented for all five cases (A)–(E), similar to the results of 2D parallel beam geometry.

Several observations can be made: at first, in general the same behavior as for the 2D parallel beam geometry can be made regarding the presented five cases, and, therefore, the mentioned statements remain valid; especially, the estimation improves in the order $\text{ML} \rightarrow \text{MAP} \rightarrow \text{MMSE}$.

Second, when focusing on the RMSE column, for all three methods the estimation of the third coordinate x_3 is

in general more accurate than the estimation of x_1 and x_2 . This can be understood by the fact that x_3 is almost directly observed in the cone beam geometry of (34) and, therefore, is estimated with higher confidence.

Third, the estimation of x_1 and x_2 is slightly better in the 3D cone beam geometry than in the 2D parallel beam geometry. This is due to the fact that also in the third coordinate there is a slight information about the first two coordinates due to the scaling in cone beam geometry. Therefore, the information content about the first two coordinates is higher in 3D cone beam geometry, having two measurements per projection \mathbf{u} for three unknowns \mathbf{x} .

Fourth, the calculation times on the test computer (Intel(R) Core2 Quad CPU, 4 Cores, 3.00 GHz, 4.00 GB RAM) for a single data point were the same for ML and MAP with in average $1.9 \cdot 10^{-3}$ seconds utilizing the MATLAB implemented Nelder-Mead simplex method and for the MMSE implementation in average 5.5 seconds utilizing 1/3 Simpson's rule for integration. The optimization in ML

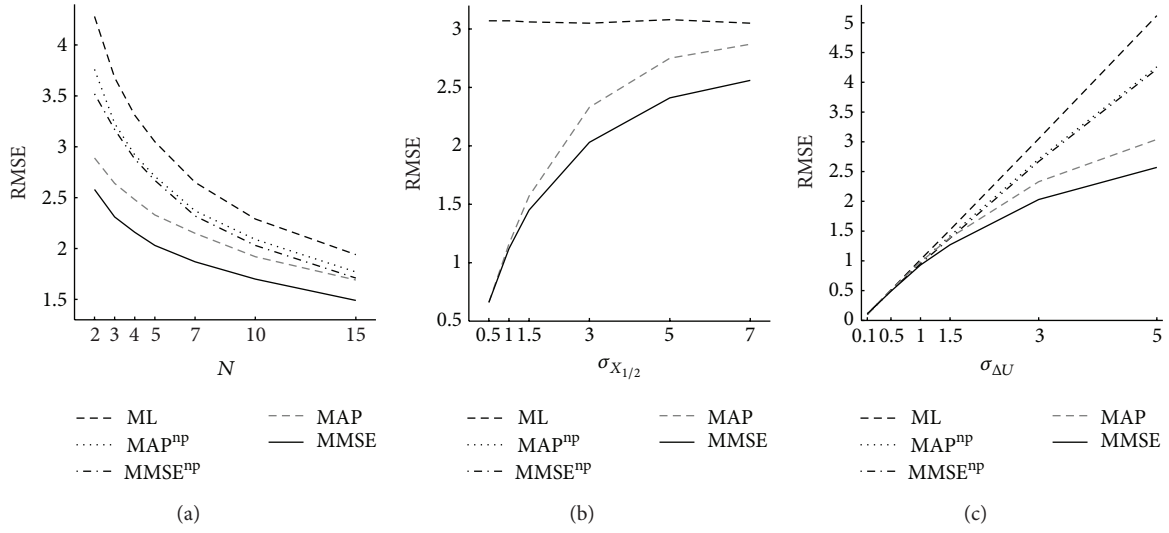


FIGURE 9: The RMSEs of ML, MAP^{np}, MMSE^{np}, MAP, and MMSE for broadly varying parameters, such as the number of projections N (a), the standard deviation of the prior knowledge $\sigma_{x_1} = \sigma_{x_2}$ (b), and the standard deviation of the observation uncertainty $\sigma_{\Delta U}$ (c). The constant parameters of each subfigure are set to the standard values of case (A). In the central subfigure MAP^{np} and MMSE^{np} are not presented since they are independent of $\sigma_{x_{1/2}}$.

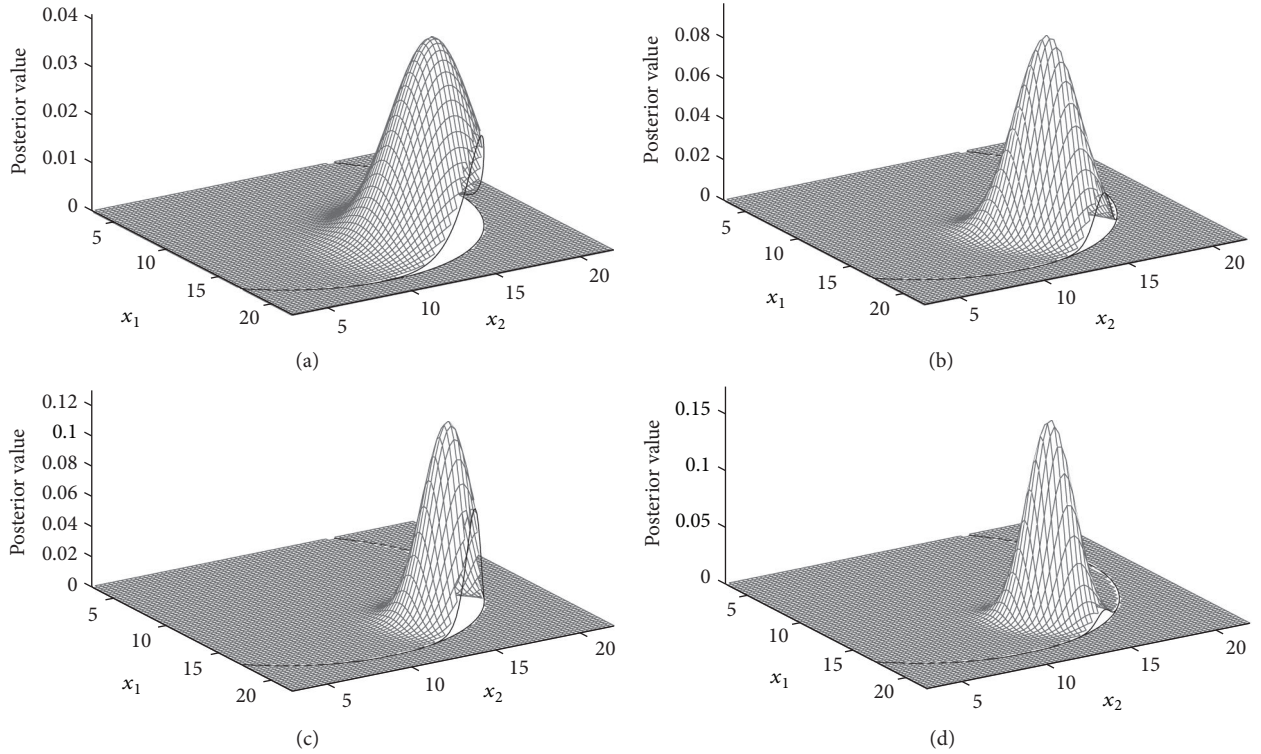


FIGURE 10: Posterior distributions for the same \mathbf{x} and full prior. (a) Case (B), using 2 projections. (b) Case (C), using 10 projections. (c) Case (D), using strong prior knowledge. (d) Case (E), using higher measurement accuracy.

and MAP was terminated if the difference of the estimated position between iterations was less than 10^{-6} , and for the MMSE integration a $100 \times 100 \times 100$ grid with a constant spatial step size of 0.2 was utilized. It could be

observed that a further refinement of these criteria did not lead to practically better results. Several ways for a further speedup of the MMSE estimator can be implemented, for example, taking symmetries into account when

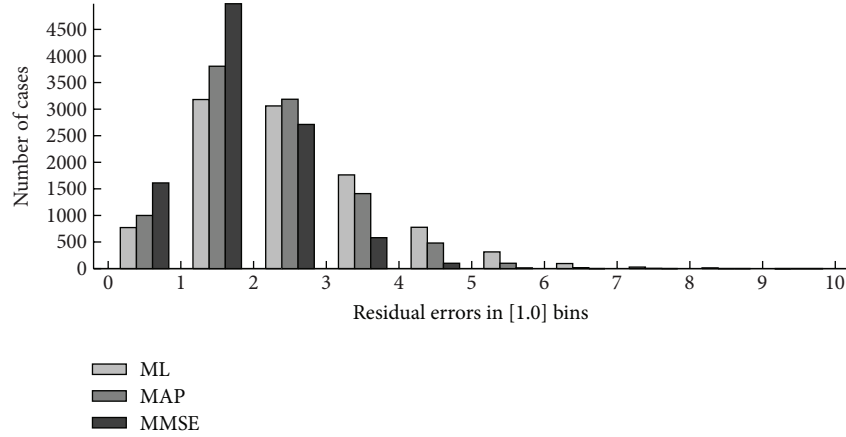


FIGURE 11: Histogram of the residual radial errors of 3D cone beam geometry for case (A) after estimation with ML (radial RMSE: 2.81), MAP (2.43), and MMSE (1.93).

TABLE 3: Statistical values about the estimation results of 3D cone beam geometry for cases (A)–(E) with $M = 10,000$ sample points \mathbf{x} each are presented for the three 3D estimators. The radial estimation error is presented at first and the coordinate estimation errors of (x_1, x_2, x_3) are shown in brackets. The same formulas described as in the caption of Table 2 results are utilized.

Case	Method	Mean	st. deviation	abs. maximum	Radial RMSE	[%]
A	ML	2.52 (−0.0,−0.0,−0.0)	1.25 (1.8,1.8,1.1)	8.89 (7.7,7.1,4.7)	2.81 (1.8,1.8,1.1)	—
	MAP	2.20 (0.8,0.8,0.3)	1.03 (1.4,1.4,1.0)	7.86 (6.1,6.9,4.1)	2.43 (1.6,1.6,1.0)	+14
	MMSE	1.77 (0.0,0.0,0.0)	0.78 (1.2,1.2,0.9)	5.99 (4.8,5.2,4.1)	1.93 (1.2,1.2,0.9)	+31
B	ML	3.36 (0.1,0.1,0.0)	1.50 (2.3,2.3,1.7)	10.55 (9.5,9.1,7.2)	3.68 (2.3,2.3,1.7)	—
	MAP	2.79 (0.9,0.9,0.6)	1.21 (1.6,1.6,1.4)	8.72 (7.9,8.3,5.5)	3.04 (1.9,1.9,1.5)	+17
	MMSE	2.27 (0.0,0.0,0.1)	1.00 (1.5,1.5,1.3)	7.80 (7.1,7.4,5.2)	2.48 (1.5,1.5,1.3)	+33
C	ML	1.89 (0.0,0.0,0.0)	0.98 (1.4,1.4,0.8)	7.57 (5.5,6.9,2.9)	2.13 (1.4,1.4,0.8)	—
	MAP	1.74 (0.6,0.6,0.2)	0.88 (1.1,1.1,0.7)	7.04 (4.8,5.5,3.0)	1.95 (1.3,1.3,0.8)	+8
	MMSE	1.42 (0.0,0.0,0.0)	0.68 (1.0,1.0,0.7)	5.18 (4.3,3.9,3.0)	1.57 (1.0,1.0,0.7)	+36
D	ML	2.56 (0.0,0.0,0.0)	1.28 (1.9,1.9,1.1)	10.99 (7.4,8.7,4.7)	2.86 (1.9,1.9,1.1)	—
	MAP	1.86 (0.9,0.9,0.5)	0.77 (0.9,0.9,0.8)	5.73 (5.7,5.3,3.6)	2.02 (1.3,1.3,0.9)	+29
	MMSE	1.31 (0.1,0.1,0.0)	0.59 (0.9,0.9,0.8)	5.00 (4.7,4.0,3.2)	1.43 (0.9,0.9,0.8)	+50
E	ML	1.27 (−0.0,−0.0,−0.0)	0.63 (0.9,0.9,0.5)	4.78 (3.9,3.8,2.2)	1.41 (0.9,0.9,0.5)	—
	MAP	1.21 (0.3,0.3,0.1)	0.60 (0.8,0.8,0.5)	4.55 (3.8,3.5,2.3)	1.35 (0.9,0.9,0.5)	+4
	MMSE	1.07 (0.0,0.0,0.0)	0.51 (0.8,0.8,0.5)	3.93 (3.0,2.8,2.0)	1.19 (0.8,0.8,0.5)	+16

utilizing Gaussians, more adaptive grid generation in the numerical integration, more effectively utilizing parallel programming and a full implementation in C compared to MATLAB.

5. Discussion

We presented briefly the general Bayesian framework for point estimation and its application to the stochastic localization/triangulation problem. This compact presentation of the estimation methods is suitable for a broad group of applications. In the general introduction, we presented the foundations of Maximum Likelihood, Maximum A Posteriori, and Bayesian estimation applied to the Minimum Mean Square Error and how they are related to each other and other popular estimation methods.

Several points can be made considering the specific application to stochastic localization/triangulation, which hold true in general for Bayesian approaches as well: as seen in the proposed application, there is a tradeoff between benefit and effort. The more the information is contained in observations (= quality of observation, e.g., more projections or higher quality measurement/detection), the smaller the benefit is due to prior knowledge. The other way round is true as well: the less the information in observations, the more important is the prior knowledge. With the Bayesian estimation (MAP^{np}, MMSE^{np}, MAP, or MMSE), these two probabilistic sources of information can be naturally included in the model.

The introduction of a *prior* can be utilized to benefit the estimation in two ways: it can place more weight on possible values of locations \mathbf{x} and effectively restrict $\Omega_{\mathbf{x}}$ by setting the prior probabilities to zeros outside of the feasibility region

(boundaries). If the constrained region of interest for \mathbf{x} is a convex region (as in the example), then the MMSE and the MAP will derive estimates in the constrained region, and, therefore, this imposes an effective constraint. However, if the region is nonconvex, only the MAP will still lead to an estimate within the restricted region, while MMSE will lead to an estimate in the convex hull of the constrained region (possibly outside the constrained region). In a broader context, the prior can also be interpreted as a regularizing influence to the presented inverse problem, such as the Tikhonov regularization as in MAP or in a more interweaved form in MMSE.

It is pointed out that in this simulation study the standard deviations of the probabilistic knowledge about the prior knowledge (σ_{x_i} for $i = \{1, 2, 3\}$) and the observation uncertainties ($\sigma_{\Delta U}$) are known before estimation. In some practical applications these parameters are not known and have to be estimated or assumed, which would introduce additional uncertainties to the presented estimation methods. The investigation of these additional uncertainties is not part of this study, and it is recommended to extend the presented methods for this purpose utilizing *hierarchical Bayesian models* [13].

Focusing on the presented 2D study, the following statements can be formulated: if no prior knowledge is available, utilizing ML with 5 or 10 projections compared to a single 2×2 with 2 projections decreases effectively the measurement uncertainties and increases the estimation accuracy by about 40%. If only information about the constraints of \mathbf{x} is available but no information about its possible distribution (e.g., truncated Gaussian), MAP^{np} should be preferred to MMSE^{np} since it achieves a similar accuracy (case (A): 11% increase for MAP^{np} relative to ML compared to 13% for MMSE^{np}) with less bias and much faster calculation times. In contrast to this, if both information are available, information about the constraints and concentration of \mathbf{x} , then utilizing MMSE should be preferred to MAP since it allows finding a better compromise close to the surfaces between constraints and concentration. This is due to the fact that MAP neglects the constraints totally as long as the estimates are inside the circle, leading to worse results in average (case (A): 33% increase for MMSE relative to ML compared to 24% for MAP). In general, it is shown in this application that with the utilization of the MMSE for spatial coordinate estimation at constraining interfaces it is possible to construct an empirically unbiased estimator with high accuracy.

Further testing of these methods in medical imaging or other practical applications of stochastic localization at boundaries is highly encouraged. This paper summarizes the basic theory for this purpose and provides clearly interpretable results in the simulation studies. However, in each practical application the true benefit of the estimation methods has to be evaluated separately or individually. For instance, in radiotherapy due to differences in patient anatomy, the problem of breast tumor localization is dramatically different from prostate localization. For breast the boundary is the breast surface, and, for prostate, hard constraints can arise from the bony anatomy or an endorectal balloon which fixes the anatomy differently.

The presented estimation with constraints incorporated into the prior is very general and not limited to linear acquisition geometries, as demonstrated for the nonlinear 3D case. Further, the results of the 3D study show that the scaling of the 3D cone beam geometry increases the estimation accuracy for the first two coordinates of \mathbf{x} compared to the parallel beam geometry in 2D. Eventually, ML and MAP estimations lead to fast optimization problems and the MMSE estimator to a more challenging high-dimensional integration. In practical situations it depends on the specific application, such as real-time imaging, to find a suitable compromise between computational effort and accuracy.

6. Conclusion

With the 2D parallel beam and 3D cone beam geometry examples it is illustrated that the general framework of Bayesian estimation can be applied to a variety of estimation problems even when no explicit formulas can be derived, such as for nonlinear/non-Gaussian distributions of the uncertainties or prior knowledge. In this context, it is demonstrated that constraints about the solution space can be efficiently incorporated into the prior for stochastic localization by utilizing Maximum A Posteriori and Minimum Mean Square Error estimation.

Acknowledgment

This work was supported by the German Research Foundation (DFG) within the funding programme Open Access Publishing.

References

- [1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.
- [2] D. E. Clark and S. Iveković, "The Cramer-Rao Lower Bound for 3-D state estimation from rectified stereo cameras," in *Proceedings of the 13th Conference on Information Fusion (Fusion '10)*, pp. 1–8, July 2010.
- [3] W. Hoegele, R. Loeschel, B. Dobler, J. Hesser, O. Koelbl, and P. Zygmanski, "Stochastic formulation of patient positioning using linac-mounted cone beam imaging with prior knowledge," *Medical Physics*, vol. 38, no. 2, pp. 668–681, 2011.
- [4] W. Hoegele, P. Zygmanski, B. Dobler, M. Kroiss, O. Koelbl, and R. Loeschel, "Localization of deformable tumors from short-arc projections using Bayesian estimation," *Medical Physics*, vol. 39, no. 12, pp. 7205–7214, 2012.
- [5] W. Hoegele, R. Loeschel, B. Dobler, O. Koelbl, C. Beard, and P. Zygmanski, "Stochastic triangulation for prostate positioning during radiotherapy using short arc arcs," *Radiotherapy & Oncology*, 2013.
- [6] C. B. Guure, N. A. Ibrahim, and A. O. M. Ahmed, "Bayesian estimation of two-parameter weibull distribution using extension of Jeffreys' prior information with three loss functions," *Mathematical Problems in Engineering*, vol. 2012, Article ID 589640, 13 pages, 2012.
- [7] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," *IEEE Proceedings F*, vol. 140, no. 2, pp. 107–113, 1993.

- [8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [9] A. Weckenmann, X. Jiang, K.-D. Sommer et al., "Multisensor data fusion in dimensional metrology," *CIRP Annals - Manufacturing Technology*, vol. 58, no. 2, pp. 701–721, 2009.
- [10] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.
- [11] A. S. Bedekar and R. M. Haralick, "A Bayesian method for triangulation and its application to finding corresponding points," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 2, pp. 362–365, October 1995.
- [12] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, NY, USA, 1998.
- [13] C. P. Robert, *The Bayesian Choice*, Springer, New York, NY, USA, 2007.
- [14] Y. C. Eldar, "Minimum variance in biased estimation: bounds and asymptotically optimal estimators," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1915–1930, 2004.
- [15] H. Cramér, *Mathematical Methods of Statistics*, Asia Publishing House, 1946.
- [16] Y. C. Eldar, "Rethinking biased estimation: improving maximum likelihood and the cramér-rao bound," *Foundations and Trends in Signal Processing*, vol. 1, pp. 305–449, 2008.
- [17] S. Rohl, S. Bodenstedt, S. Suwelack et al., "Dense gpu-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration," *Medical Physics*, vol. 39, no. 3, pp. 1632–1645, 2012.
- [18] G. Zheng, "Statistical shape model-based reconstruction of a scaled, patient-specific surface model of the pelvis from a single standard AP x-ray radiograph," *Medical Physics*, vol. 37, no. 4, pp. 1424–1439, 2010.
- [19] R. Schulze, U. Heil, O. Weinheimer et al., "Accurate registration of random radiographic projections based on three spherical references for the purpose of few-view 3D reconstruction," *Medical Physics*, vol. 35, no. 2, pp. 546–555, 2008.

