An Approximate Bayesian Approach to Optimal Input Signal Design for System Identification

Piotr Bania * and Anna Wójcik

Posted Date: 3 October 2025

doi: 10.20944/preprints202509.1390.v3

Keywords: Design of experiment; Bayesian experimental design; optimal input signal design; system identification; optimal input signal design; entropy; information

Article

An Approximate Bayesian Approach to Optimal Input Signal Design for System Identification

Piotr Bania * and Anna Wójcik

AGH University of Krakow, Faculty of Electrical Engineering, Automatics, Computer Science, and Biomedical Engineering, Department of Automatic Control and Robotics, al. A. Mickiewicza 30, 30-059, Krakow

* Correspondence: pba@agh.edu.pl; Tel.: + 48 12 617 46 78

† Current address: AGH University of Krakow, Faculty of Electrical Engineering, Automatics, Computer Science, and Biomedical Engineering, Department of Automatic Control and Robotics, al. A. Mickiewicza 30, 30-059, Krakow, building B-1, room 310.

Abstract

The design of informatively rich input signals is essential for accurate system identification, yet classical Fisher-information–based methods are inherently local and often inadequate in the presence of significant model uncertainty and nonlinearity. This paper develops a Bayesian approach that uses the mutual information (MI) between observations and parameters as the utility function. To address the computational intractability of the MI, we maximize a tractable MI lower bound. The method is then applied to the design of an input signals for the identification of quasi-linear stochastic dynamical systems. Evaluating the MI lower bound requires inversion of large covariance matrices whose dimensions scale with the number of data points N. To overcome this problem, an algorithm that reduces the dimension of the matrices to be inverted by a factor of N is developed, making the approach feasible for long experiments. The proposed Bayesian method is compared with the average D-optimal design method, a semi-Bayesian approach, and its advantages are demonstrated. The effectiveness of the proposed method is further illustrated through four examples, including atomic sensor models, where the input signals that generates large MI are especially important for reducing the estimation error.

Keywords: design of experiment; Bayesian experimental design; optimal input signal design; system identification; entropy; information

1. Introduction

The design of informative input signals is the cornerstone of modern system identification. Without properly chosen excitation, even advanced estimation algorithms may fail to provide accurate parameter estimates, leading to unreliable prediction and control. Classical references such as [1–3] and reviews by [4–6], emphasize that identification is not only a matter of statistical estimation but also of experimental design, where the input signal determines the achievable information content. Optimal experimental design (OED) methods therefore play a crucial role in practical applications. Traditionally, OED has relied on the Fisher Information Matrix (FIM), with criteria such as D or A optimality widely used due to their computational efficiency and asymptotic guarantees [4,7]. However, FIM-based approaches are inherently local, relying on linearization and asymptotic normality. They may thus be fragile in scenarios with large model uncertainty or strongly non-linear stochastic dynamics.

A natural alternative is the Bayesian approach, which evaluates an experiment through its *expected information gain*, typically quantified by the mutual information between model parameters and observations [5,6,8–11]. Bayesian design has several advantages: it is globally valid over the parameter space, it naturally incorporates prior information, and it is applicable to non-linear, stochastic models. Its main drawback is computational intractability, since mutual information requires high-dimensional

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

integration over parameters and observations. Since the general case is challenging and difficult to solve, in this article we focus on models that can be represented in the form

$$\boldsymbol{Y} = \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}) + \boldsymbol{Z}, \tag{1}$$

where $\boldsymbol{\theta} \in \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_r\}$ is a parameter with prior distribution $P(\boldsymbol{\theta} = \boldsymbol{\theta}_j) = p_{0,j}$. The noise $\boldsymbol{Z}$ is conditionally normal, that is, $p(\boldsymbol{Z}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{Z}, 0, \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}))$, and $\boldsymbol{U}$ is a design variable. For this class of models, the density of the observations $\boldsymbol{Y}$ is a finite Gaussian mixture of the form $p(\boldsymbol{Y}) = \sum_{j=1}^{r} p_{0,j} \mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}))$. Within these Gaussian mixtures, the mutual information between $\boldsymbol{Y}$ and $\boldsymbol{\theta}$ can be estimated from below by the effective and tractable pairwise-distance-based lower bound given by Kolchinsky and Tracey [12,13]. We maximize this bound to achieve an approximate optimal design parameter $\boldsymbol{U}$ and then generalize the method to a parameter space of continuum cardinality. In particular, we show how to treat the Gaussian prior $p_0(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{m}_\theta, \mathbf{S}_\theta)$ and prior distributions with compact support.

In this work, we focus on *quasi-linear systems*, namely stochastic dynamical systems that are linear in the state variables but non-linear in the control variables. Such systems occur ubiquitously in science and engineering. In quantum mechanics, Hamiltonians and Lindblad dissipators depend on external control fields such as laser intensities, magnetic fields, or gate voltages, leading to a non-linear dependence on the control [14,15]. In chemical processes, flow rates directly determine reaction speeds in continuous stirred tank reactors [16,17]. In thermal plants, convection coefficients scale non-linearly with flow, giving rise to quasi-linear heat transfer dynamics [16]. This broad applicability makes quasi-linear systems a natural and important class for advanced input design. However, there is a notable lack of Bayesian design methods and software tools tailored to this class of systems. Therefore, in this paper, we address this gap by developing such a method. Specifically, we show that a finite sequence of observations generated by a quasi-linear system can always be expressed in the form of the model (1), and we provide an effective algorithm for calculating the lower bound on mutual information.

The study reported here constitutes a substantial and far-reaching extension of the initial results presented in [18], as well as related research reported in [19–21]. The article's main contributions can be summarized as follows. We first introduce an Information-Theoretic Lower Bound (ITB) on the estimation error of any estimator, [22,23], and briefly discuss its relation to the Bayesian Cramér–Rao Bound (BCRB), [24–26] [23]. We conclude that maximizing mutual information is superior to maximizing Bayesian or classical Fisher information, which is consistent with the arguments presented in [5,8,10,11]. Building on this result, we introduce a novel Bayesian design method for model (1). To address the intractability and computational complexity of direct mutual information evaluation, we discretize the parameter space and maximize the Kolchinsky–Tracey lower bound [12,13], and subsequently extend this approach to a parameter space of continuum cardinality. We then focus on the application to linear and quasi-linear system identification. Since the information-theoretic bound requires inversion of large covariance matrices of the observations, whose dimensions grow linearly with the number of data points N (with $N \sim 10^3$–10^6 in applications), direct inversion is computationally problematic. To overcome this challenge, we develop an algorithm that reduces the dimension of the matrices needed for inversion by a factor of N, thus making the approach feasible for long experiments. The proposed Bayesian method is compared with the average D-optimal design [1,4,27], a semi-Bayesian method, and its advantages are demonstrated. The effectiveness of the method is further illustrated by four examples. The first two, intentionally elementary, highlights the effectiveness of our approach. The third and fourth examples are drawn from atomic sensor models, a domain where optimal input design is particularly critical. We analyze a controlled harmonic oscillator with stochastic disturbances as a paradigmatic atomic sensor model [28], and a complex magnetometer model with a non-linear dependence of the system matrices on the input [29]. In the latter case, we provide a simplified model, derive the optimal input, and demonstrate that it significantly outperforms

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

the harmonic signal, which might otherwise be presumed optimal. Moreover, we show that the estimation error of the MAP estimator achieves the theoretical lower bound.

The article is organized as follows. Section 2 formulates the problem. Section 3 develops the approximate Bayesian solution for finite and infinite parameter spaces. Section 4 applies the method to quasi-linear systems. Section 5 compares the approach with classical design methods. Section 6 presents examples. Section 7 provides discussion and conclusions.

2. Formulation of the Problem

Let us consider a family of models

$$\boldsymbol{Y} = \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}) + \boldsymbol{Z}, \tag{2}$$

where $\boldsymbol{Y}, \boldsymbol{Z} \in \mathbb{R}^{n_Y}$, $\boldsymbol{U} \in \mathbb{R}^{n_U}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{n_\theta}$. The set $\boldsymbol{\Theta}$ will be called parameter space. Parameter $\boldsymbol{\theta}$ is unknown. Prior distribution of $\boldsymbol{\theta}$ is denoted by p_0. Random variable $\boldsymbol{Z}$ is conditionally normal i.e. $p(\boldsymbol{Z}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{Z}, 0, \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}))$, where $\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}) \in \mathbf{S}^+(n_Y)$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{U} \in \mathbb{R}^{n_U}$. Functions $\boldsymbol{F}$ and $\mathbf{S}$ are smooth. The variable $\boldsymbol{U}$ is called the design parameter or, in the context of dynamical systems, the input signal. The set of admissible signals is given by

$$\mathcal{U}_{ad} = \{\boldsymbol{U} \in \mathbb{R}^{n_U}; |\boldsymbol{U} - \tilde{\boldsymbol{U}}| \leqslant \varrho\}, \tag{3}$$

where $\tilde{\boldsymbol{U}}$ is the given vector and ϱ is maximal norm of the signal. We will also consider an alternative, and useful in some applications, definition of $\mathcal{U}_{ad}$:

$$\mathcal{U}_{ad} = \{\boldsymbol{U} \in \mathbb{R}^{n_U}; \boldsymbol{U}_{min} \leqslant \boldsymbol{U} \leqslant \boldsymbol{U}_{max}\}, \tag{4}$$

where $\boldsymbol{U}_{min}, \boldsymbol{U}_{max} \in \mathbb{R}^{n_U}$ are fixed vectors. Under these assumptions and after applying the Bayes rule we get the likelihood, evidence and posterior distribution of $\boldsymbol{\theta}$:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{U}) = \mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})), \tag{5}$$

$$p(\boldsymbol{Y}|\boldsymbol{U}) = \int p_0(\boldsymbol{\theta}) \mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})) d\boldsymbol{\theta}, \tag{6}$$

$$p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{U}) = \frac{p_0(\boldsymbol{\theta}) p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{U})}{p(\boldsymbol{Y}|\boldsymbol{U})}. \tag{7}$$

The Minimum Mean Squared Error (MMSE) estimator of $\boldsymbol{\theta}$ is then given by

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Y}, \boldsymbol{U}) = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{U}) d\boldsymbol{\theta}. \tag{8}$$

To avoid the difficulties involved in calculating the integral (8), instead of the MSE, the Maximum a Posteriori (MAP) estimator is typically used. Taking the negative logarithm of both sides of (7) and omitting the terms independent on $\boldsymbol{\theta}$, we get the following:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}) = \tfrac{1}{2} |\boldsymbol{Y} - \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U})|^2_{\mathbf{S}^{-1}(\boldsymbol{\theta}, \boldsymbol{U})} + \tfrac{1}{2} \ln |S(\boldsymbol{\theta}, \boldsymbol{U})| - \ln p_0(\boldsymbol{\theta}). \tag{9}$$

Thus, the MAP estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Y}, \boldsymbol{U}) = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}). \tag{10}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

Estimators (8) or (10) may be biased, so the Cramér-Rao Bound (CRB) cannot be applied directly to them. However, a Bayesian version of the CRB exists and can be used to estimate the error of any, also biased, estimator [23–26]. Let us introduce Bayesian information (BI):

$$\mathbf{J}_B = \mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{Y}|\boldsymbol{U})}\left[\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta},\boldsymbol{Y},\boldsymbol{U})(\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta},\boldsymbol{Y},\boldsymbol{U})^{\mathsf{T}})\right] = \mathbf{J}_D + \mathbf{J}_P, \tag{11}$$

where

$$\mathbf{J}_P = \mathbb{E}_{p_0(\boldsymbol{\theta})}\left[\nabla_{\boldsymbol{\theta}} \ln p_0(\boldsymbol{\theta})(\nabla_{\boldsymbol{\theta}} \ln p_0(\boldsymbol{\theta}))^{\mathsf{T}}\right], \tag{12}$$

is the fraction of BI associated to a prior and

$$\mathbf{J}_D(\boldsymbol{U}) = \mathbb{E}_{p(\boldsymbol{\theta},\boldsymbol{Y}|\boldsymbol{U})}\left[\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{U})(\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{U}))^{\mathsf{T}}\right], \tag{13}$$

is part of BI provided by observations $\boldsymbol{Y}$. The matrix $\mathbf{J}_D$ is the Bayesian equivalent of the Fisher information matrix. The Fisher information matrix can be recovered from (13) assuming $p_0(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^0)$, where $\boldsymbol{\theta}^0$ is the true value of the parameter. Let $\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})$ be any estimator of $\boldsymbol{\theta}$. Assuming a sufficiently regular prior, it can be proven that

$$\mathbb{E}|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})|^2 \geqslant \frac{n_\theta}{|\mathbf{J}_P + \mathbf{J}_D(\mathbf{U})|^{1/n_\theta}}, \tag{14}$$

which is usually known as the Van Trees inequality or Bayesian Cramér-Rao Bound (BCRB) [23, 24,30][inequality (2.9), p. 17]. Formulas (11)-(13) are well defined only under rather restrictive assumptions. In particular, the joint density $p(\boldsymbol{Y},\boldsymbol{\theta})$ must be differentiable with respect to $\boldsymbol{\theta}$, and it must satisfy the regularity condition $\int \nabla_{\boldsymbol{\theta}} p(\boldsymbol{Y},\boldsymbol{\theta}) d\boldsymbol{Y} = 0$. Moreover, the prior and likelihood distributions must guarantee the existence of the expectations in (12) and (13). This excludes the uniform and many other useful prior distributions and considerably limits the applicability of inequality (14) (see [23] for details). Beyond inequality (14), there exists a large class of Bayesian bounds reported in [26]. Many of these bounds can serve as a utility function.

Probably one of the best design criterion is the Ziv-Zakai lower bound [31]. However, to compute this estimate, an additional, and rather complex, optimization sub-problem must be solved as shown in [31]. Therefore, due to the high computational complexity of the multivariate Ziv-Zakai bound, we will not consider it here. Given the application-oriented focus of this article and following the arguments presented in [5,6], we conclude that the entropy-based lower bound [22][p. 255], [23][Sect. 2.2, p.16-17], is a reasonable optimality criterion and provides slightly tighter estimates than the BCRB (cf. [31][Sect. V.D.]). To proceed, let us define the entropies of $\boldsymbol{Y}$ and $\boldsymbol{\theta}$ and the corresponding conditional entropies:

$$H_Y(\boldsymbol{U}) = \mathbb{E}(-\ln p(\boldsymbol{Y}|\boldsymbol{U})), \tag{15}$$

$$H_\theta = \mathbb{E}(-\ln p_0(\boldsymbol{\theta})), \tag{16}$$

$$H_{Y|\theta}(\boldsymbol{U}) = \mathbb{E}(-\ln p(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{U})) = \tfrac{1}{2}\int p_0(\boldsymbol{\theta}) \ln((2\pi e)^{n_Y}|\mathbf{S}(\boldsymbol{\theta},\boldsymbol{U})|) d\boldsymbol{\theta}, \tag{17}$$

$$H_{\theta|Y}(\boldsymbol{U}) = \mathbb{E}(-\ln p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{U})). \tag{18}$$

The mutual information (MI) between $\boldsymbol{\theta}$ and $\boldsymbol{Y}$ is defined as

$$I_{\boldsymbol{\theta};Y}(\boldsymbol{U}) = H_\theta - H_{\theta|Y}(\boldsymbol{U}) = H_Y(\boldsymbol{U}) - H_{Y|\theta}(\boldsymbol{U}). \tag{19}$$

The following theorem establishes the ultimate limit of the estimation error expressed in terms of mutual information, demonstrating that the Bayesian Cramér–Rao Bound (BCRB) does not constitute a fundamental limit.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

Theorem 1. *Let $\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})$ be any estimator of $\boldsymbol{\theta}$. Then the following inequalities hold:*

$$\mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})|^2 \geqslant n_\theta(2\pi e)^{-1}e^{2n_\theta^{-1}(H_\theta - I_{\theta;Y}(\boldsymbol{U}))} \geqslant \frac{n_\theta}{|\mathbf{J}_P+\mathbf{J}_D(\boldsymbol{U})|^{1/n_\theta}}. \tag{20}$$

The proof is given in Appendix A. The first of the inequalities (20) will be called the Information-Theoretic Lower Bound (ITB). The last part of (20) is known as the Efroimovich inequality [23,25][inequality (2.7), Ch. 2.2, p. 16].

To determine the optimal signal, one may maximize the MI, the determinant of $\mathbf{J}_B$ or the determinant of FIM. The latter corresponds to the classical design methods outlined in Sect. 5. However, the right-hand side of (14) generally underestimates the estimation error, and large values of $\mathbf{J}_B$ do not necessarily guarantee a small error. We illustrate this problem in Appendix B. In contrast, the ITB (20), which plays a central role in our subsequent analysis, shows that maximizing $I_{\theta;Y}(\boldsymbol{U})$ is essential for reducing the estimation error and provides a more fundamental criterion than maximizing either the Bayesian or classical Fisher information. In particular, in the context of optimal experimental design and input signal design for system identification, maximizing the mutual information between the parameters and the observations constitutes the most principled optimality criterion, as it directly quantifies the amount of knowledge gained about the parameters [5]. Accordingly, we define the optimal signal as the solution of the following optimization problem:

$$\boldsymbol{U}^* = \arg\max_{\boldsymbol{U}\in\mathcal{U}_{ad}} I_{\theta;Y}(\boldsymbol{U}). \tag{21}$$

As $I_{\theta;Y}$ is smooth and $\mathcal{U}_{ad}$ is compact, then (21) is well defined. After solving the task, the MMSE, MAP, or any other estimator can be used to determine $\boldsymbol{\theta}$.

Computing mutual information (MI) or its lower-bound remains a significant challenge. The analyses presented in the literature [5,6,10,11], show that this can be done in three main ways: 1) using Monte Carlo or nested Monte Carlo (MC) simulations, [5][Sect. 3.1.], [6]; 2) applying variational lower-bound (VLB) estimates of the MI, [5][Sect. 3.3.1], [6]; 3) utilizing existing, easily computable estimates of conditional entropy or the MI. Since we aim to numerically maximize MI, which also depends on the design parameter U, the procedure for calculating MI will be called by the optimization solver millions of times and must therefore be sufficiently fast. Consequently, although MC methods provide good estimates of MI, they are of limited use here. The VLB methods require simultaneous optimization of the variational distribution with respect to its parameters and the signal U, [5][Sect. 3.3.1 and 4.3.4], [6]. In addition, stochastic simulations are also used to compute the expected values in the VLB. As the goal of this article is to develop a simple design method that does not require hours of computation, we focus on the third option and use existing, easily computable lower-bounds of entropy or MI provided in [12] or [32].

3. Approximate Solutions

The optimization problem (21) becomes considerably more tractable when the parameter space $\boldsymbol{\Theta}$ is finite. Consequently, we first derive an approximate solution for a finite set of parameters and subsequently extend this result to obtain an approximate solution for the case where $\boldsymbol{\Theta}$ is an uncountable subset of $\mathbb{R}^{n_\theta}$.

3.1. Finite Parameter Space

Let $\boldsymbol{\Theta}=\{\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_r\}, \boldsymbol{\theta}_i\in\mathbb{R}^{n_\theta}, \boldsymbol{\theta}_i\neq\boldsymbol{\theta}_j$ and assume that p_0 is a discrete distribution of the form

$$p_0(\boldsymbol{\theta}) = \sum_{j=1}^{r} p_{0,j}\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_j), \tag{22}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where $p_{0,j} = P(\theta = \theta_j)$. Then, on the basis of (6) and (22), the density of $\boldsymbol{Y}$ becomes a Gaussian mixture:

$$p(\boldsymbol{Y}|\boldsymbol{U}) = \sum_{j=1}^{r} p_{0,j}\mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U})). \tag{23}$$

The application of the Bayes rule gives the posterior:

$$p(\boldsymbol{\theta}_j|\boldsymbol{Y}, \boldsymbol{U}) = \frac{p_{0,j}\mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U}))}{p(\boldsymbol{Y}|\boldsymbol{U})}. \tag{24}$$

The discrete counterpart of formulas (15)-(19) takes the form:

$$H_Y(\boldsymbol{U}) = -\int p(\boldsymbol{Y}|\boldsymbol{U}) \ln p(\boldsymbol{Y}|\boldsymbol{U}) d\boldsymbol{Y}, \tag{25}$$

$$H_\theta = -\sum_{j=1}^{r} p_{0,j} \ln p_{0,j}, \tag{26}$$

$$H_{Y|\theta}(\boldsymbol{U}) = \tfrac{1}{2}\sum_{j=1}^{r} p_{0,j} \ln\big((2\pi e)^{n_Y}|\mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U})|\big), \tag{27}$$

$$H_{\theta|Y}(\boldsymbol{U}) = \int p(\boldsymbol{Y}|\boldsymbol{U}) \left(-\sum_{j=1}^{r} p(\boldsymbol{\theta}_j|\boldsymbol{Y}, \boldsymbol{U}) \ln p(\boldsymbol{\theta}_j|\boldsymbol{Y}, \boldsymbol{U}) \right) d\boldsymbol{Y}, \tag{28}$$

$$I_{\boldsymbol{\theta};\boldsymbol{Y}}(\boldsymbol{U}) = H_\theta - H_{\theta|Y}(\boldsymbol{U}) = H_Y(\boldsymbol{U}) - H_{Y|\theta}(\boldsymbol{U}). \tag{29}$$

The direct computation of mutual information (29) remains difficult and, in many cases, intractable. **Hence, our central idea is to overcome this difficulty by replacing mutual information (29) with a computationally tractable and non-trivial lower bound.** In particular, we observe that (23) is a finite Gaussian mixture. For such mixtures, one of the most effective lower bounds on $I_{\boldsymbol{\theta}|\boldsymbol{Y}}$ is the inequality introduced in [12].

Lemma 1. *(Information bounds, [12]). For the Gaussian mixture (23) with $p_{0,j} = P(\theta = \theta_j)$, the following inequalities holds:*

$$I_l(\boldsymbol{U}) \leqslant I_{\theta;Y}(\boldsymbol{U}) \leqslant H_\theta, \tag{30}$$

where

$$I_l(\boldsymbol{U}) = -\sum_{i=1}^{r} p_{0,i} \ln\left(\sum_{j=1}^{r} p_{0,j} e^{-d_{i,j}(\boldsymbol{U})} \right), \tag{31}$$

$$d_{i,j}(\boldsymbol{U}) = \tfrac{1}{8}\boldsymbol{\Delta}_{i,j}^{\mathsf{T}} \Big(\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j) \Big)^{-1} \boldsymbol{\Delta}_{i,j} + \tfrac{1}{2}\ln|\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)| - \tfrac{1}{4}\ln\big(|\mathbf{S}_i||\mathbf{S}_j|\big), \tag{32}$$

$$\boldsymbol{\Delta}_{i,j} = \boldsymbol{F}(\boldsymbol{\theta}_i, \boldsymbol{U}) - \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}_i = \mathbf{S}(\boldsymbol{\theta}_i, \boldsymbol{U}), \mathbf{S}_j = \mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U}). \tag{33}$$

Detailed proof is given in [13][Sect. IIIb, inequality (11), and Sect. IV, formula (15) with $\alpha = 0.5$] and also in [12]. Now, the approximate solution of (21) is given by

$$\boldsymbol{U}^* = \arg\max_{\boldsymbol{U}\in\mathcal{U}_{ad}} I_l(\boldsymbol{U}). \tag{34}$$

Since $\mathcal{U}_{ad}$ is compact and I_l is smooth and bounded, the solution of (34) exists. We also note that in the case of two alternatives, that is, when $r = 2$ in (31), we get

$$e^{-I_l(\boldsymbol{U})} = \left(p_{0,1} + p_{0,2} e^{-d_{1,2}(\boldsymbol{U})} \right)^{p_{0,1}} \left(p_{0,1} e^{-d_{1,2}(\boldsymbol{U})} + p_{0,2} \right)^{p_{0,2}}. \tag{35}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

Accordingly, the optimal signal in this case arises as the solution of a somewhat simplified optimization problem:

$$\max_{\boldsymbol{U}\in\mathcal{U}_{ad}} d_{1,2}(\boldsymbol{U}). \tag{36}$$

If the function $\boldsymbol{F}$ in (2) is affine with respect to $\boldsymbol{U}$ and the covariance $\mathbf{S}$ does not depend on $\boldsymbol{U}$, then it follows from Lemma 1, that $d_{1,2}$ is a positive (semi-) definite quadratic form with respect to $\boldsymbol{U}$. For constraints (4), we thus obtain a convex quadratic programming problem. In the case of constraints (3), one needs to find the minimum of $d_{1,2}$ on a closed ball in $\mathbb{R}^N$. This is also a convex problem, and it can be reduced to finding zeros of a scalar function [33][Thm. 4.1, p. 70 and Sect. 4.3]. Furthermore, if $\boldsymbol{F}(\boldsymbol{\theta}_i, \boldsymbol{U}) = \boldsymbol{F}_i\boldsymbol{U}, i = 1, 2$, and the constraints are defined by (3), then the solution of (36) is the eigenvector of the matrix $\mathbf{Q} = (\boldsymbol{F}_1 - \boldsymbol{F}_2)^T(\mathbf{S}_1 + \mathbf{S}_2)^{-1}(\boldsymbol{F}_1 - \boldsymbol{F}_2)$, corresponding to its largest eigenvalue (see [18] [Sect. 2.1], for details).

3.2. Infinite Parameter Space

Let us assume that $\Theta = \mathbb{R}^{n_\theta}$ and consider the Gaussian prior

$$p_0(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{m}_\theta, \mathbf{S}_\theta), \mathbf{S}_\theta > 0. \tag{37}$$

Then the integral in (6) can be approximated with a finite Gaussian mixture

$$p(\boldsymbol{Y}|\boldsymbol{U}) = \int p_0(\boldsymbol{\theta})\mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}))d\boldsymbol{\theta} \approx \sum_{j=1}^{N_a} p_{0,j}\mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U})), \tag{38}$$

where $p_{0,j} \geqslant 0$ and $\sum_{j=1}^{N_a} p_{0,j} = 1$. The weights $p_{0,j}$ and the nodes $\boldsymbol{\theta}_j$ in (38) can be calculated by using the multidimensional Gauss-Hermite quadrature rule or any other suitable method.

The Gauss-Hermite quadrature of order p is exact for polynomials of degree at most $2p - 1$. The approximation error of the integral (38) using the Gauss-Hermite quadrature of order p depends on the $2p$-th derivatives of the functions F and S. In the single-parameter case, with Gaussian prior $\mathcal{N}(\theta, m_\theta, \sigma_\theta)$, the error estimate is given by the formula:

$$e \leqslant \sigma_\theta \frac{4^p C}{p!} \sup_{\theta,\boldsymbol{U},\boldsymbol{Y}} \frac{d^{2p}}{d\theta^{2p}}\mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\theta, \boldsymbol{U}), \mathbf{S}(\theta, \boldsymbol{U})), \tag{39}$$

where C is constant. The error tends to zero as $p \to \infty$ or $\sigma_\theta \to 0$. Therefore, the Gauss-Hermite approximation of the integral (38) is especially useful when the prior is narrow (small σ_θ) or when the integrand, in the neighbourhood of the point m_θ, can be well approximated by low-degree polynomials. To illustrate the method, we will show only a very simple second-order Gaussian quadrature rule with $2n_\theta$ points.

Lemma 2. *Approximate value of the integral $J(f) = \int \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{m}_\theta, \mathbf{S}_\theta)f(\boldsymbol{\theta})d\boldsymbol{\theta}$ is given by*

$$J(f) \approx \frac{1}{2n_\theta} \sum_{j=1}^{2n_\theta} f(\boldsymbol{\theta}_j), \tag{40}$$

where

$$\boldsymbol{\theta}_{2i-1} = \boldsymbol{m}_\theta - \mathbf{S}_\theta^{0.5}\sqrt{n_\theta}\boldsymbol{e}_i, \boldsymbol{\theta}_{2i} = \boldsymbol{m}_\theta + \mathbf{S}_\theta^{0.5}\sqrt{n_\theta}\boldsymbol{e}_i, i = 1, ..., n_\theta \tag{41}$$

and $\boldsymbol{e}_i$ is i^{th} basis vector in $\mathbb{R}^{n_\theta}$. If $f(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^\mathsf{T}\mathbf{A}\boldsymbol{\theta} + \boldsymbol{b}^\mathsf{T}\boldsymbol{\theta} + c$, then the equality holds in (40).

Proof. Direct calculation. □

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

An analogous method can be used for prior distributions defined on compact subsets of $\mathbb{R}^{n_\theta}$ (e.g. an n-dimensional hypercube), but the formulas for nodes and weights in (38) will then change. For example, if θ is a scalar parameter and the prior distribution is uniform, that is, $p_0(\theta) = U[a,b]$, then, using a second-order Gauss-Legendre quadrature, the approximate value of the integral $\int_a^b p_0(\theta)f(\theta)d\theta$ is computed using the formula:

$$\int_a^b p_0(\theta)f(\theta)d\theta \approx p_{0,1}f(\theta_1) + p_{0,2}f(\theta_2), \tag{42}$$

where $p_{0,1} = p_{0,2} = 0.5$ and

$$\theta_1 = \tfrac{1}{2}\left(a+b-\frac{b-a}{\sqrt{3}}\right), \theta_2 = \tfrac{1}{2}\left(a+b+\frac{b-a}{\sqrt{3}}\right). \tag{43}$$

The formula (42) is exact for polynomials of degree 3. The error estimate is analogous to (39) and tends to zero when $b-a \to 0$. More general multidimensional formulas, integration methods and error estimates are given in [34–36].

The application of lemma 2 to (38) gives $N_a = 2n_\theta$, $p_{0,j} = (2n_\theta)^{-1}$. Now, since (38) is approximated by a Gaussian mixture, the results of section 3.1 can be used. Based on Eqs. (31-33), (38) and Lemma 1, the information lower bound takes the form:

$$I_l(\boldsymbol{U}) = -\frac{1}{2n_\theta}\sum_{i=1}^{2n_\theta}\ln\left(\frac{1}{2n_\theta}\sum_{j=1}^{2n_\theta}e^{-d_{i,j}(\boldsymbol{U})}\right), \tag{44}$$

where $d_{i,j}$ and $\boldsymbol{\theta}_j$ are given by Eqs. (31-33) and (41) or (43). The approximate solution of (21) can be found by maximizing (44) with constraints (3) or (4).

4. Bayesian Input Signal Design in Quasi-Linear Control Systems

Consider the family of quasi-linear systems

$$\boldsymbol{x}_{k+1} = \mathbf{A}(\boldsymbol{\theta},\boldsymbol{u}_k)\boldsymbol{x}_k + \mathbf{B}(\boldsymbol{\theta},\boldsymbol{u}_k) + \mathbf{G}(\boldsymbol{\theta},\boldsymbol{u}_k)\boldsymbol{w}_k, \tag{45}$$

$$\boldsymbol{y}_k = \mathbf{C}\boldsymbol{x}_k + \boldsymbol{v}_k, \tag{46}$$

where $k = 0,1,...,N$, $N \geqslant 1$, $\boldsymbol{x}_k \in \mathbb{R}^n, \boldsymbol{y}_k \in \mathbb{R}^{n_y}, \boldsymbol{w}_k \in \mathbb{R}^{n_w}, \boldsymbol{v}_k \in \mathbb{R}^{n_y}$, $\boldsymbol{w}_k \sim \mathcal{N}(0,\mathbb{I})$, $\boldsymbol{v}_k \sim \mathcal{N}(0,\mathbf{S}_v), \mathbf{S}_v > 0$. Variables $\boldsymbol{x}_0, \boldsymbol{w}_0, ..., \boldsymbol{w}_{N-1}, \boldsymbol{v}_0, ..., \boldsymbol{v}_N$ are mutually independent. The initial state $\boldsymbol{x}_0$ is conditionally normal i.e. $p(\boldsymbol{x}_0|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{m}_0^-(\boldsymbol{\theta}), \mathbf{S}_0^-(\boldsymbol{\theta}))$, where $\boldsymbol{m}_0^-, \mathbf{S}_0^-$ are smooth and $\mathbf{S}_0^-(\boldsymbol{\theta}) > 0$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The joint prior distribution of the initial state $\boldsymbol{x}_0$ and the parameter $\boldsymbol{\theta}$ is given by $p_0(\boldsymbol{x}_0,\boldsymbol{\theta}) = p_0(\boldsymbol{\theta})\mathcal{N}(\boldsymbol{x}_0, \boldsymbol{m}_0^-(\boldsymbol{\theta}), \mathbf{S}_0^-(\boldsymbol{\theta}))$. Let's define $\mathbf{A}_k = \mathbf{A}(\boldsymbol{\theta},\boldsymbol{u}_k)$, $\mathbf{B}_k = \mathbf{B}(\boldsymbol{\theta},\boldsymbol{u}_k)$, $\mathbf{G}_k = \mathbf{G}(\boldsymbol{\theta},\mathbf{u}_k)$. Then the solution of (45) has the form:

$$\boldsymbol{x}_0 = \mathbb{I}\boldsymbol{x}_0, \tag{47}$$

$$\boldsymbol{x}_1 = \mathbf{A}_0\boldsymbol{x}_0 + \mathbf{B}_0 + \mathbf{G}_0\boldsymbol{w}_0, \tag{48}$$

$$\boldsymbol{x}_2 = \mathbf{A}_1\boldsymbol{x}_1 + \mathbf{B}_1 + \mathbf{G}_1\boldsymbol{w}_1 = \mathbf{A}_1\mathbf{A}_0\boldsymbol{x}_0 + \mathbf{A}_1\mathbf{B}_0 + \mathbf{B}_1 + \mathbf{A}_1\mathbf{G}_0\boldsymbol{w}_0 + \mathbf{G}_1\boldsymbol{w}_1, \tag{49}$$

$$\vdots \tag{50}$$

$$\boldsymbol{x}_N = \boldsymbol{\Phi}(N,0)\boldsymbol{x}_0 + \sum_{j=0}^{N-1}\boldsymbol{\Phi}(N,j+1)\boldsymbol{B}_j + \sum_{j=0}^{N-1}\boldsymbol{\Phi}(N,j+1)\mathbf{G}_j\boldsymbol{w}_j, \tag{51}$$

Peer-reviewed version available at *Entropy* **2025**, *27*, 1041; doi:https://doi.org/10.3390/e27101041

where $\mathbf{\Phi}(n,n) = \mathbb{I}$ and

$$\mathbf{\Phi}(n,j) = \prod_{i=1}^{n-j} \mathbf{A}_{n-i}, j < n. \tag{52}$$

Now, if we denote $\boldsymbol{X} = \mathrm{col}(\boldsymbol{x}_0, ..., \boldsymbol{x}_N)$, $\boldsymbol{Y} = \mathrm{col}(\boldsymbol{y}_0, ..., \boldsymbol{y}_N)$, $\boldsymbol{U} = \mathrm{col}(\boldsymbol{u}_0, ..., \boldsymbol{u}_{N-1})$, $\boldsymbol{W} = \mathrm{col}(\boldsymbol{w}_0, ..., \boldsymbol{w}_{N-1})$, $\boldsymbol{V} = \mathrm{col}(\boldsymbol{v}_0, ..., \boldsymbol{v}_N)$, we can rewrite Eqs. (47-51) and (46), in matrix-vector form

$$\boldsymbol{X} = \mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})\boldsymbol{x}_0 + \mathcal{B}(\boldsymbol{\theta}, \boldsymbol{U}) + \mathcal{G}(\boldsymbol{\theta}, \boldsymbol{U})\boldsymbol{W}, \tag{53}$$

$$\boldsymbol{Y} = \mathcal{C}\boldsymbol{X} + \boldsymbol{V}, \tag{54}$$

where the matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{G}$, $\mathcal{C} = \mathbb{I}_{N+1} \otimes \mathbf{C}$ follow directly from Eqs. (47-51), (46), and $\boldsymbol{W} \sim \mathcal{N}(0, \mathbb{I}_{Nn_w})$, $\boldsymbol{V} \sim \mathcal{N}(0, \mathbb{I}_{N+1} \otimes \mathbf{S}_v)$. Substituting (53) into (54) and taking into account that $p(\boldsymbol{x}_0|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{m}_0^-(\boldsymbol{\theta}), \mathbf{S}_0^-(\boldsymbol{\theta}))$, we get

$$\boldsymbol{Y} = \mathcal{C}\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})\boldsymbol{m}_0^-(\boldsymbol{\theta}) + \mathcal{C}\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{U}) + \boldsymbol{Z}, \tag{55}$$

$$\boldsymbol{Z} = \mathcal{C}\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})(\boldsymbol{x}_0 - \boldsymbol{m}_0^-(\boldsymbol{\theta})) + \mathcal{C}\mathcal{G}(\boldsymbol{\theta}, \boldsymbol{U})\boldsymbol{W} + \boldsymbol{V}. \tag{56}$$

The conditional density of variable $\boldsymbol{Z}$ has the form $p(\boldsymbol{Z}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{Z}, 0, \mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}))$, where the covariance matrix $\mathbf{S}$ is given by

$$\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U}) = \mathcal{C}(\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})\mathbf{S}_0^-(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})^{\mathsf{T}} + \mathcal{G}(\boldsymbol{\theta}, \boldsymbol{U})\mathcal{G}(\boldsymbol{\theta}, \boldsymbol{U})^{\mathsf{T}})\mathcal{C}^{\mathsf{T}} + \mathbb{I}_{N+1} \otimes \mathbf{S}_v. \tag{57}$$

Finally, if we define

$$\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}) = \mathcal{C}\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{U})\boldsymbol{m}_0(\boldsymbol{\theta}) + \mathcal{C}\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{U}), \tag{58}$$

we can rewrite (55) in the form $\boldsymbol{Y} = \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U}) + \boldsymbol{Z}$, which is exactly the model (2). To find the optimal input signal, we maximize one of the criteria (31) or (44) with constraints (3) or (4).

With a large number of data (large N), calculating the inverse and determinant of a very large matrix $\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})$ in (9) and calculating the quantities $d_{i,j}(\boldsymbol{U})$ in Eqs. (31-33) is numerically ill-conditioned and requires special treatment. The algorithms below reduce the size of the matrices necessary to invert by a factor of $N + 1$.

Lemma 3. *The following identities hold:*

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{U}) = \prod_{k=0}^{N} \mathcal{N}(\boldsymbol{y}_k, \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta}), \mathbf{\Sigma}_k(\boldsymbol{\theta})), \tag{59}$$

$$|\boldsymbol{Y} - \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U})|^2_{\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})^{-1}} = \sum_{k=0}^{N} |\boldsymbol{y}_k - \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta})|^2_{\mathbf{\Sigma}_k^{-1}(\boldsymbol{\theta})}, \tag{60}$$

$$|\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})| = \prod_{k=0}^{N} |\mathbf{\Sigma}_k(\boldsymbol{\theta})|, \tag{61}$$

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}) = \tfrac{1}{2} \sum_{k=0}^{N} \left(|\boldsymbol{y}_k - \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta})|^2_{\mathbf{\Sigma}_k^{-1}(\boldsymbol{\theta})} + \ln |\mathbf{\Sigma}_k(\boldsymbol{\theta})| \right) - \ln p_0(\boldsymbol{\theta}), \tag{62}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where $\mathcal{L}$ is given by (9) and $\boldsymbol{m}_k^-$, $\boldsymbol{\Sigma}_k$, are calculated recursively by the Kalman filter

$$\boldsymbol{\Sigma}_k(\boldsymbol{\theta}) = \mathbf{S}_v + \mathbf{C}\mathbf{S}_k^-(\boldsymbol{\theta})\mathbf{C}^\mathsf{T}, \tag{63}$$
$$\mathbf{L}_k(\boldsymbol{\theta}) = \mathbf{S}_k^-(\boldsymbol{\theta})\mathbf{C}^\mathsf{T}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\theta}), \tag{64}$$
$$\boldsymbol{m}_k(\boldsymbol{\theta}) = \boldsymbol{m}_k^-(\boldsymbol{\theta}) + \mathbf{L}_k(\boldsymbol{\theta})(\boldsymbol{y}_k - \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta})), \tag{65}$$
$$\mathbf{S}_k(\boldsymbol{\theta}) = \mathbf{S}_k^-(\boldsymbol{\theta}) - \mathbf{L}_k(\boldsymbol{\theta})\boldsymbol{\Sigma}_k(\boldsymbol{\theta})\mathbf{L}_k(\boldsymbol{\theta})^\mathsf{T}, \tag{66}$$
$$\boldsymbol{m}_{k+1}^-(\boldsymbol{\theta}) = \mathbf{A}_k\boldsymbol{m}_k(\boldsymbol{\theta}) + \mathbf{B}_k, \tag{67}$$
$$\mathbf{S}_{k+1}^-(\boldsymbol{\theta}) = \mathbf{A}_k\mathbf{S}_k(\boldsymbol{\theta})\mathbf{A}_k^\mathsf{T} + \mathbf{G}_k\mathbf{G}_k^\mathsf{T}, k = 0, 1..., N, \tag{68}$$

with initial conditions $\boldsymbol{m}_0^-(\boldsymbol{\theta})$, $\mathbf{S}_0^-(\boldsymbol{\theta})$.

The proof is given in Appendix A. The Eqs. (63-68), are, in fact, a family of discrete-time Kalman filters indexed by $\boldsymbol{\theta}$. The first four formulas describe the correction step. The prediction step is given by the last two equations. The matrix $\mathbf{L}_k$ is the Kalman gain and $\boldsymbol{\Sigma}_k$ is the covariance matrix of the output prediction error $\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - \mathbf{C}\boldsymbol{m}_k^-$.

Lemma 4. *Let us define*

$$\tilde{\mathbf{A}}_k = \begin{bmatrix} \mathbf{A}(\boldsymbol{\theta}_i, \boldsymbol{u}_k) & 0 \\ 0 & \mathbf{A}(\boldsymbol{\theta}_j, \boldsymbol{u}_k) \end{bmatrix}, \tilde{\mathbf{B}}_k = \begin{bmatrix} \mathbf{B}(\boldsymbol{\theta}_i, \boldsymbol{u}_k) \\ \mathbf{B}(\boldsymbol{\theta}_j, \boldsymbol{u}_k) \end{bmatrix}, \tag{69}$$

$$\tilde{\mathbf{G}}_k = \begin{bmatrix} \mathbf{G}(\boldsymbol{\theta}_i, \boldsymbol{u}_k) & 0 \\ 0 & \mathbf{G}(\boldsymbol{\theta}_j, \boldsymbol{u}_k) \end{bmatrix}, \tilde{\mathbf{C}} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{C} & -\mathbf{C} \end{bmatrix} \tag{70}$$

and let

$$\tilde{\boldsymbol{\Sigma}}_k = \mathbf{S}_v + \tilde{\mathbf{C}}\tilde{\mathbf{S}}_k^-\tilde{\mathbf{C}}^\mathsf{T}, \tag{71}$$
$$\tilde{\mathbf{L}}_k = \tilde{\mathbf{S}}_k^-\tilde{\mathbf{C}}^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_k^{-1}, \tag{72}$$
$$\tilde{\mathbf{S}}_k = \tilde{\mathbf{S}}_k^- - \tilde{\mathbf{L}}_k\tilde{\boldsymbol{\Sigma}}_k\tilde{\mathbf{L}}_k^\mathsf{T}, \tag{73}$$
$$\tilde{\boldsymbol{m}}_{k+1}^- = \tilde{\mathbf{A}}_k(\mathbb{I} - \tilde{\mathbf{L}}_k\tilde{\mathbf{C}})\tilde{\boldsymbol{m}}_k^- + \tilde{\mathbf{B}}_k, \tag{74}$$
$$\tilde{\mathbf{S}}_{k+1}^- = \tilde{\mathbf{A}}_k\tilde{\mathbf{S}}_k\tilde{\mathbf{A}}_k^\mathsf{T} + \tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^\mathsf{T}, k = 0, 1, ..., N, \tag{75}$$

with initial conditions

$$\tilde{\boldsymbol{m}}_0^- = \begin{bmatrix} \boldsymbol{m}_0^-(\boldsymbol{\theta}_i) \\ \boldsymbol{m}_0^-(\boldsymbol{\theta}_j) \end{bmatrix}, \tilde{\mathbf{S}}_0^- = \begin{bmatrix} \mathbf{S}_0^-(\boldsymbol{\theta}_i) & 0 \\ 0 & \mathbf{S}_0^-(\boldsymbol{\theta}_j) \end{bmatrix}. \tag{76}$$

Then the quantity $d_{i,j}(\boldsymbol{U})$ in formula (31) is given by:

$$d_{i,j}(\boldsymbol{U}) = \frac{1}{4}\sum_{k=0}^{N} |\tilde{\boldsymbol{C}}\tilde{\boldsymbol{m}}_k^-|^2_{\tilde{\boldsymbol{\Sigma}}_k^{-1}} + \frac{1}{2}\sum_{k=0}^{N} \ln|\tilde{\boldsymbol{\Sigma}}_k| - \frac{1}{4}\ln(|\mathbf{S}_i||\mathbf{S}_j|), \tag{77}$$

where $|\mathbf{S}_i| = |\mathbf{S}(\boldsymbol{\theta_i}, \boldsymbol{U})|$, $|\mathbf{S}_j| = |\mathbf{S}(\boldsymbol{\theta_j}, \boldsymbol{U})|$, are calculated accordingly to Lemma 3, Eq. (61).

The proof is given in Appendix A. Let us observe that, instead of calculating the inverse and determinant of the large matrices $\mathbf{S}_i$, $\mathbf{S}_j$, $\frac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)$, of dimension $(N+1)n_y$, we only need to calculate the determinants and inverses of the much smaller matrices $\boldsymbol{\Sigma}_k$, $\tilde{\boldsymbol{\Sigma}}_k$, whose dimension is n_y, which is usually a small number.

5. Comparison with Classical Methods of Input Signal Design

Classical methods for input signal design in system identification are primarily concerned with LTI state space or transfer function models (such as ARMAX) and are usually based on maximizing

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

some functions of error covariance or Fisher information matrix. For the prediction error method (PEM) estimator, the asymptotic form of the error covariance matrix (or Fisher information) is well known, both in the time and frequency domains. In the time domain, the solution corresponds to a specific input signal, whereas in the frequency domain, the solution yields the optimal power spectral density of the input signal. Below, we provide a brief overview of these methods, following the methodology presented in [1,2][Ch.9. Sect. 9.3, 9.4] and [4][Sect. 6.1].

Consider the LTI, SISO system

$$\boldsymbol{x}_{k+1} = \mathbf{A}(\boldsymbol{\theta})\boldsymbol{x}_k + \mathbf{B}(\boldsymbol{\theta})u_k + \mathbf{G}(\boldsymbol{\theta})\boldsymbol{w}_k, \tag{78}$$

$$y_k = \boldsymbol{C}\boldsymbol{x}_k + v_k, \tag{79}$$

under the assumptions stated in Section 4. System (78), (79) is equivalent to the transfer function model

$$y_k = G(\boldsymbol{\theta}, z)u_k + H(\boldsymbol{\theta}, z)e_k, \tag{80}$$

where $e_k \sim \mathcal{N}(0, \sigma_e^2)$ is a sequence of mutually independent Gaussian variables. The filters G and H are determined by the formulas

$$G(\boldsymbol{\theta}, z) = \mathbf{C}(z\mathbb{I} - \mathbf{A}(\boldsymbol{\theta}))^{-1}\mathbf{B}(\boldsymbol{\theta}), H(\boldsymbol{\theta}, z) = 1 + \mathbf{C}(z\mathbb{I} - \mathbf{A}(\boldsymbol{\theta}))^{-1}\mathbf{K}(\boldsymbol{\theta}), \tag{81}$$

where the Kalman gain $\mathbf{K}(\boldsymbol{\theta})$ is given by

$$\mathbf{K}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\mathbf{S}(\boldsymbol{\theta})\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{S}(\boldsymbol{\theta})\mathbf{C}^{\mathsf{T}} + \sigma_v^2)^{-1}, \tag{82}$$

with a non-negative matrix $\mathbf{S}$ being a solution of the Riccati equation (cf. [2])

$$\mathbf{S} = \mathbf{A}\mathbf{S}\mathbf{A}^{\mathsf{T}} + \mathbf{G}\mathbf{G}^{\mathsf{T}} - \mathbf{A}\mathbf{S}\mathbf{C}^{\mathsf{T}}(\mathbf{C}\mathbf{S}\mathbf{C}^{\mathsf{T}} + \sigma_v^2)^{-1}\mathbf{C}\mathbf{S}\mathbf{A}^{\mathsf{T}}. \tag{83}$$

The prediction errors are given by the recurrence

$$\epsilon_k(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}) = H^{-1}(\boldsymbol{\theta}, z)(y_k - G(\boldsymbol{\theta}, z)u_k). \tag{84}$$

The cost function used in the Prediction Error Method (PEM) is expressed as:

$$V(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}) = \frac{1}{2N\sigma_e^2}\sum_{k=1}^{N}\epsilon_k^2(\boldsymbol{\theta}, \boldsymbol{U}). \tag{85}$$

Minimization of (85) with reference to $\boldsymbol{\theta}$ yields the PEM estimator

$$\hat{\boldsymbol{\theta}}(\boldsymbol{Y}, \boldsymbol{U}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} V(\boldsymbol{\theta}, \boldsymbol{Y}, \boldsymbol{U}). \tag{86}$$

The above estimator, under rather weak identifiability conditions, is consistent, asymptotically normal, and efficient, i.e. it achieves the Cramér-Rao lower bound. Following the reasoning presented in [1,2][Ch. 9, Sect. 9.3 and 9.4] or [4][Sect. 6.1], we divide the parameter vector into two groups related to the parameters appearing in G and H, that is, $\boldsymbol{\theta} = \mathrm{col}(\boldsymbol{\theta_H}, \boldsymbol{\theta_G})$. The sensitivity of ϵ_k to changes in $\boldsymbol{\theta}_G$ is calculated recursively according to the following equations:

$$\boldsymbol{\psi}_k(\boldsymbol{\theta}, \boldsymbol{U}) = H^{-1}(\boldsymbol{\theta}, z)\left(\nabla_{\boldsymbol{\theta}_G} G(\boldsymbol{\theta}, z)\right)u_k = F_z(\boldsymbol{\theta}, z)u_k, \tag{87}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where $\nabla_{\boldsymbol{\theta}_G}$ means differentiation only with respect to the parameters that occur in G. The information matrix, which is also the inverse of the error covariance $\mathbf{P}_{\boldsymbol{\theta}_G}$, is given by:

$$\mathbf{M}(\boldsymbol{\theta}, \boldsymbol{U}) = \mathbf{P}_{\boldsymbol{\theta}_G}^{-1}(\boldsymbol{\theta}, \boldsymbol{U}) = \mathbf{R}_e(\boldsymbol{\theta}) + \frac{1}{N\sigma_e^2} \sum_{k=1}^{N} \boldsymbol{\psi}_k(\boldsymbol{\theta}, \boldsymbol{U}) \boldsymbol{\psi}_k(\boldsymbol{\theta}, \boldsymbol{U})^{\mathsf{T}}, \tag{88}$$

where $\mathbf{R}_e$ does not depend on $\boldsymbol{U}$. Using the D-optimal criterion, the optimal signal is given by maximization of $\det\big(\mathbf{M}(\boldsymbol{\theta}^0, \boldsymbol{U})\big)$, where $\boldsymbol{\theta}^0$ is the true value of the parameter. Since $\boldsymbol{\theta}^0$ is unknown, one can use the prior distribution and maximize the average D-optimal criterion:

$$Q(\boldsymbol{U}) = \mathbb{E}_{p_0(\boldsymbol{\theta})} \det(\mathbf{M}(\boldsymbol{\theta}, \boldsymbol{U})), \tag{89}$$

with constraints (3) or (4). The asymptotic error covariance can also be expressed in terms of the power spectral density of the input signal u_k. Let Φ_u denote the spectral density of u_k. As it was shown in [2][p. 291] or [4] [Sect. 6.2], we have:

$$\mathbf{M}(\Phi_u, \boldsymbol{\theta}^0) = \mathbf{P}_{\boldsymbol{\theta}_G}^{-1}(\Phi_u, \boldsymbol{\theta}^0) = \frac{N}{2\pi\sigma_e^2} \int_{-\pi}^{\pi} F_z(e^{i\omega}, \boldsymbol{\theta}^0) F_z(e^{-i\omega}, \boldsymbol{\theta}^0)^{\mathsf{T}} \Phi_u(\omega) d\omega + \mathbf{R}_e(\boldsymbol{\theta}^0), \tag{90}$$

where F_z is defined by (87) and the term $\mathbf{R}_e$ in (90) does not depend on Φ_u. Similarly as before, the parameter-averaged determinant of the matrix $\mathbf{M}$ is maximized with respect to Φ_u, subject to the signal power and frequency constraints. Typically, the spectrum Φ_u is parametrized by a finite number of coefficients c_k, so that the resulting optimization problem is convex; see [37] for details. After performing spectral factorization of Φ_u, a filter is obtained, whose input is white noise and whose output yields the optimal signal u_k. This has been implemented in the MOOSE-2 solver [38]. Unfortunately, MOOSE-2 does not allow for averaging over prior and involves *true* unknown value of the parameter.

Numerous variants of the aforementioned methods can be found in the literature. For example, instead of the D-optimality criterion, one may also consider maximizing $\mathrm{tr}(\mathbf{M})$ or $\lambda_{\min}(\mathbf{M})$. However, the vast majority of methods are based on the principles stated above (see, e.g., [4]), that is, maximization of some functions of the Fisher information matrix. Finally, we note that the above methods employ the classical optimality criterion, averaged only over the prior distribution. Consequently, they are not fully Bayesian and, following the terminology of [11], should rather be referred to as *pseudo-Bayesian* methods.

6. Examples of Input Signal Design

In the following, we present four examples of optimal input signal design using both Bayesian and classical methods. Examples 1-3 are classical in nature and concern time-invariant linear stochastic systems. Examples 1 and 2 are elementary, while Example 3, taken from [28], addresses the design of a control signal for a paradigmatic model of the atomic sensor. The sensor is modelled as a harmonic oscillator with the natural frequency being the parameter of interest. In examples 1-3, the Bayesian approach is compared with classical methods.

Example 4, adapted from [29], is more advanced and considers the design of the pump laser control signal in an optically pumped magnetometer. The magnetometer is modelled as a quasi-linear stochastic system, where the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{G}$ depend nonlinearly on the control signal u. For this system, classical methods cannot be applied. Therefore, estimation errors are compared with the Information-Theoretic Lower Bound (ITB) provided in Theorem 1 and with the errors obtained by using an appropriately selected harmonic input signal.

In all examples, the errors were computed using the Monte Carlo method. The parameter $\boldsymbol{\theta}$ and the initial conditions $\boldsymbol{x}_0$ were sampled from the prior distribution $p_0(\boldsymbol{x}_0, \boldsymbol{\theta})$. Observations $y_0, \ldots, y_N$ corresponding to the sampled parameters and initial conditions were then generated, and the error of

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

the MAP estimator was calculated. This error was subsequently averaged over many repetitions of the procedure.

6.1. Elementary Example

We begin with a very simple first-order system

$$x_{k+1} = \theta_1 x_k + \theta_2 u_k + g w_k, \tag{91}$$

$$y_k = x_k + \sigma_v v_k, k = 0, 1, ..., N, \tag{92}$$

with $\sigma_v = 0.1$, $g = 0.01$. The parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2]^{\mathsf{T}}$, has a prior distribution $p_0(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{m}_\theta, \mathbf{S}_\theta)$, where $\boldsymbol{m}_\theta = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix}^{\mathsf{T}}$, $\mathbf{S}_\theta = 10^{-2}\mathbb{I}$. As assumed in section 4, the initial condition x_0 is conditionally Gaussian, that is, $p(x_0|\boldsymbol{\theta}) = \mathcal{N}(x_0, m_0^-(\boldsymbol{\theta}), s_0^-(\boldsymbol{\theta}))$, with $m_0^-(\boldsymbol{\theta}) = 0$, $s_0^-(\boldsymbol{\theta}) = 0.01$. The length of the signal $N = 100$ and the set of admissible signals is given by (3) with $\tilde{\boldsymbol{U}} = 0$, that is, the norm of the signal cannot be greater than ϱ. To minimize the averaged D-optimal criterion (89), we need to calculate the sensitivity of the prediction error. The sensitivity equations (87) now take the form

$$(1 + (K(\theta_1) - \theta_1)z^{-1})(1 - \theta_1 z^{-1})\psi_{1,k} = \theta_2 z^{-2} u_k, \tag{93}$$

$$(1 + (K(\theta_1) - \theta_1)z^{-1})\psi_{2,k} = z^{-1} u_k, \tag{94}$$

where the Kalman gain $K(\theta_1)$ is given by (82), (83) with $\mathbf{A} = \theta_1$, $\mathbf{G} = g$, $\mathbf{C} = 1$.

The optimal input signals were designed by maximizing the Bayesian criterion (44), the averaged D-optimal criterion (89), and the spectral criterion (90), subject to the constraint (3) with $\tilde{\boldsymbol{U}} = 0$.

Maximization of the spectral criterion (90) was performed using the MOOSE-2 solver [38], evaluated at $\boldsymbol{\theta} = \boldsymbol{m}_\theta$ with default parameters, that is, the input spectrum was FIR-type with 20 lags and the spectrum power constraint was set to 1 (prob.spectrum.signal.power.ub=1). There were no additional constraints on the shape of the spectrum.

The optimal signals and the corresponding estimation errors of θ_1 and θ_2 are shown in Figures 1 and 2. In Fig. 2 we also calculate the estimation errors for the constant (step) signal, which is certainly not optimal. The constant signal and the MOOSE signal were always assigned a norm equal to ϱ.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

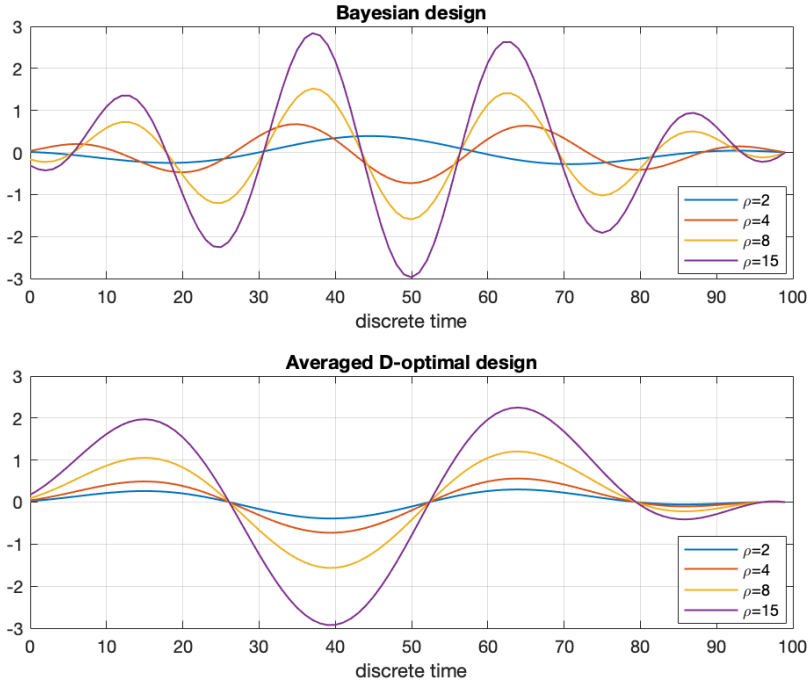

Figure 1. Optimal input signals resulting from the maximization of the Bayesian criterion (44) (top) and of the averaged D-optimal criterion (89) (bottom), shown for several values of ϱ.

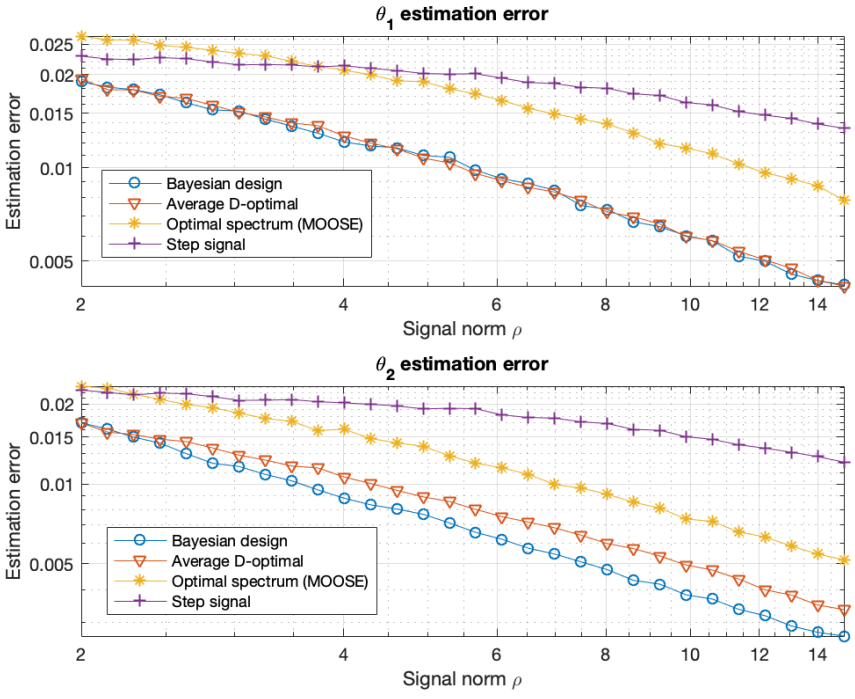

Figure 2. Mean estimation errors of the parameters θ_1 and θ_2, obtained using the MAP estimator (10), as functions of the maximum admissible signal norm ϱ. The results are based on a Monte Carlo simulation with 3000 repetitions. The constant (step) signal and the MOOSE signal were always assigned a norm equal to ϱ.

6.2. Example with a Non-Gaussian Prior Distribution

Consider the following system:

$$d\boldsymbol{x} = (\mathbf{A}_c(\theta) + \mathbf{B}_c(\theta)u)dt + \mathbf{G}_c(\theta)d\boldsymbol{w}, \tag{95}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

$$y_k = \mathbf{C}\boldsymbol{x}_k + s_v v_k, \tag{96}$$

where

$$\mathbf{A}_c(\theta) = \begin{bmatrix} 0 & 1 \\ 0 & -\theta \end{bmatrix}, \mathbf{B}_c = \begin{bmatrix} 0 \\ \theta \end{bmatrix}, \mathbf{G}_c(\theta) = \begin{bmatrix} 0 \\ \sqrt{d_c\theta} \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{97}$$

$\boldsymbol{x}_k = \boldsymbol{x}(t_k)$, $t_k = k\Delta$, $\Delta = 0.05 \cdot 10^{-3}$, $d_c = 0.01$, $s_v = 0.1$. This system can be considered as controlled Brownian motion or DC motor with stochastic disturbances. The parameter θ is the unknown damping rate of the system. Assuming that $u(t) = u_k, t \in [t_k, t_{k+1}]$, the discrete-time system corresponding to (95) has the form

$$\boldsymbol{x}_{k+1} = \mathbf{A}(\theta)\boldsymbol{x}_k + \mathbf{B}(\theta)u_k + \mathbf{G}(\theta)w_k, \tag{98}$$

where, according to the procedure given in Appendix C:

$$\begin{aligned} \mathbf{A}(\theta) &= \begin{bmatrix} 1 & \frac{1-e^{-\theta\Delta}}{\theta} \\ 0 & e^{-\theta\Delta} \end{bmatrix}, \mathbf{B}(\theta) = \begin{bmatrix} \Delta - \frac{1-e^{-\theta\Delta}}{\theta} \\ 1 - e^{-\theta\Delta} \end{bmatrix}, \\ \mathbf{G}(\theta) &= \sqrt{d_c\theta} \begin{bmatrix} \sqrt{D_{1,1}(\theta)} & 0 \\ \frac{D_{1,2}(\theta)}{\sqrt{D_{1,1}(\theta)}} & \sqrt{\frac{D_{1,1}(\theta)D_{2,2}(\theta) - D_{1,2}(\theta)^2}{D_{1,1}(\theta)}} \end{bmatrix}, \end{aligned} \tag{99}$$

where

$$D_{1,1}(\theta) = \frac{4e^{-\theta\Delta} - e^{-2\theta\Delta} + 2\theta\Delta - 3}{2\theta^3}, \tag{100}$$

$$D_{1,2}(\theta) = \frac{1 - 2e^{-\theta\Delta} + e^{-2\theta\Delta}}{2\theta^2}, D_{2,2}(\theta) = \frac{1 - e^{-2\theta\Delta}}{2\theta}. \tag{101}$$

The initial condition is Gaussian with $\boldsymbol{m}_0^- = 0$, $\mathbf{S}_0^- = \mathrm{diag}[0.001, 0.005]$. Unlike in the previous example, here **we assume that the prior distribution of θ is uniform**, that is, $p_0(\theta) = U[a, b]$ with $a = 0.05$, $b = 2$. Following the Gauss-Legendre formula (43), we get $\theta_1 = 0.5(a + b - (b - a)/\sqrt{3}) \approx 0.462$, $\theta_2 = 0.5(a + b + (b - a)/\sqrt{3}) \approx 1.588$, $p_{0,1} = p_{0,2} = 0.5$. Thus, $r = 2$ in (31) and, accordingly to (35), the Bayesian optimal signal is a solution of the simplified and convex optimization problem (36) with $d_{1,2}$ defined by Lemmas 3 and 4. Moreover, since the matrices in (99), do not depend on u_k, the last two terms in (77) can be omitted. The set of admissible signals is given by (3) with $\tilde{\boldsymbol{U}} = 0$, that is, the signal norm cannot be greater than ϱ.

In order to employ the classical methods described in Sect. 5, it is necessary to first evaluate the sensitivity of the prediction error. The transfer functions G and H in (80) have the form

$$G(\theta, z) = \frac{B(\theta, z)}{A(\theta, z)} z^{-1}, H(\theta, z) = \frac{C(\theta, z)}{A(\theta, z)}, \tag{102}$$

where

$$A(\theta, z) = 1 - (1 + e^{\theta\Delta}))z^{-1} + e^{-\theta\Delta}z^{-2}, \tag{103}$$

$$B(\theta, z) = \Delta - \frac{1 - e^{-\theta\Delta}}{\theta} + \left(\frac{1}{\theta} - \left(\frac{1}{\theta} + \Delta\right)e^{-\theta\Delta}\right)z^{-1}, \tag{104}$$

$$C(\theta, z) = 1 + \left(\mathbf{K}_1(\theta) - \left(1 - e^{-\theta\Delta}\right)\right)z^{-1} + \left(\frac{1 - e^{-\theta\Delta}}{\theta}\mathbf{K}_2(\theta) + (1 - \mathbf{K}_1(\theta))e^{-\theta\Delta}\right)z^{-2}, \tag{105}$$

and the Kalman gain $\mathbf{K}$ is given by (82). Since we only have one parameter, the sensitivity ψ_k is a number, and the sensitivity equation (87) now takes the form

$$A(\theta, z)C(\theta, z)\psi_k(\theta, \boldsymbol{U}) = \left(\frac{\partial B(\theta, z)}{\partial\theta}A(\theta, z) - B(\theta, z)\frac{\partial A(\theta, z)}{\partial\theta}\right)z^{-1}u_k. \tag{106}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

The D-optimal signal is then obtained by maximization of the averaged D-optimal criterion

$$Q(\boldsymbol{U}) = \mathbb{E}_{p_0(\theta)}\left[\frac{1}{N}\sum_{k=1}^{N}\psi_k^2(\theta,\boldsymbol{U})\right], \tag{107}$$

with constraints (3).

The optimal input signals were designed by maximizing the Bayesian criterion (44) and the averaged D-optimal criterion (107), subject to the constraint (3) with $\tilde{\boldsymbol{U}} = 0$. The results are presented in Figures 3 and 4. Figure 4 also shows the estimation error for the step signal (constant) and the PRBS signal. The constant and PRBS signals were always assigned a norm equal to ϱ.

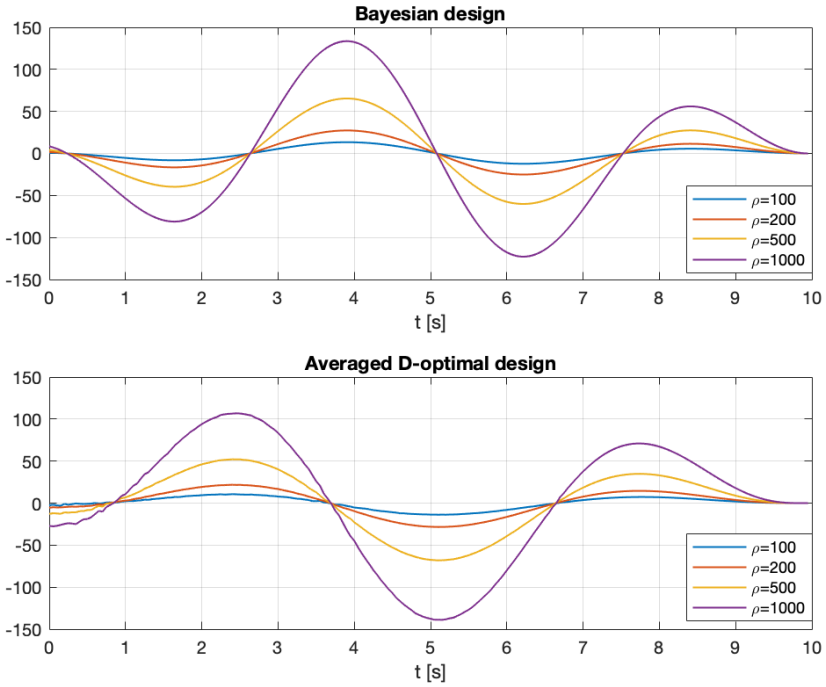

Figure 3. Optimal input signals (left) and corresponding system outputs (right) obtained by maximizing the Bayesian criterion (36) (top) and the averaged D-optimal criterion (107) (bottom) subject to the constraint (3) with $\tilde{\boldsymbol{U}} = 0$.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

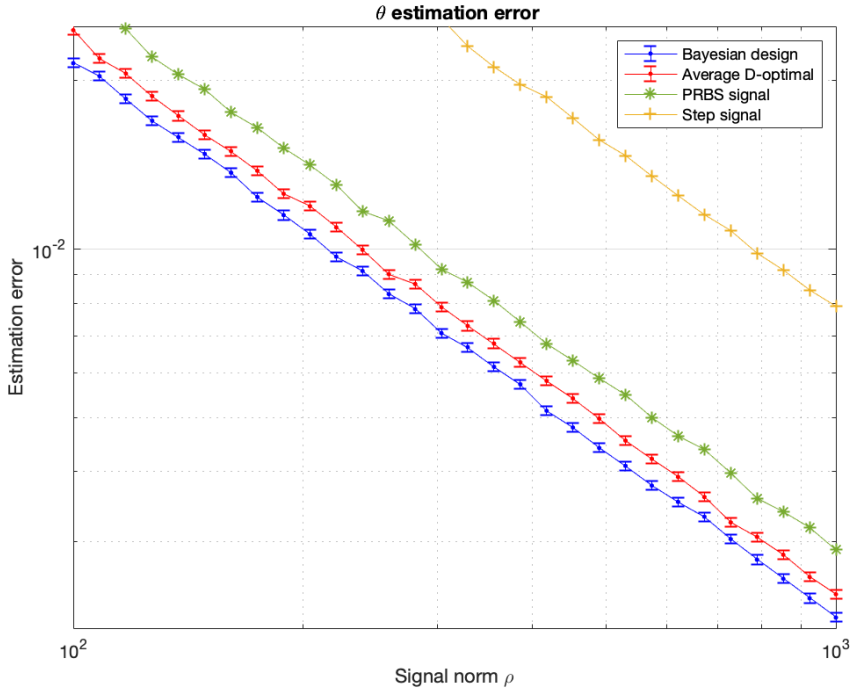

Figure 4. Mean estimation errors of the parameter θ, obtained using the MAP estimator (10), as functions of the maximum admissible signal norm ϱ for different signals. The results are based on a Monte Carlo simulation with 6000 repetitions.

6.3. Optimal Input Design for Atomic Sensor Model

In [28], a simplified paradigmatic model of an atomic sensor (an atomic magnetometer, [39,40]) is introduced, in which the dynamics is governed by oscillations of the collective spin of an atomic ensemble subjected to an external magnetic field. The system is driven by circularly polarized light from a pump laser, whose frequency acts as the input signal. A linearly polarized probe laser illuminates the atoms, and upon transmission through the medium, its polarization undergoes a Faraday rotation. The J_z component of the collective spin is inferred from the measurement of the probe laser's polarization angle. The model presented in [28] describes the dynamics of the spin components $\boldsymbol{J} = [J_y, J_z]^{\mathsf{T}}$ and has the form

$$d\boldsymbol{J} = \begin{bmatrix} -\frac{1}{T_2} & \omega_L \\ -\omega_L & -\frac{1}{T_2} \end{bmatrix} J dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathcal{E}(t) dt + d\boldsymbol{w}^{(J)}, \tag{108}$$

where ω_L is the Larmor frequency, $T_2 = 0.87$ ms, is the relaxation time, $\mathcal{E}$ is the pumping laser frequency, and $\boldsymbol{w}^{(J)}$ is the Wiener process with known covariance $q\mathbb{I}$. The observation has the form $I_k = g_D J_z(k\Delta) + \xi_k$, $k = 0, 1, \ldots$, where I_k is the photocurrent, $\Delta = 5$ μs, is the sampling time, $\xi_k \sim \mathcal{N}(0, \sigma_\xi^2)$ and g_D, σ_ξ are known parameters. The Larmor frequency and the external magnetic field B are related to each other by the formula $\omega_L = \gamma_e B$, where γ_e is the gyromagnetic ratio. Hence, by measuring ω_L, one can determine the field B. Taking T_2 as the time unit and after rescaling the time, state variables, observations, and the input signal $\mathcal{E}$, we get the following, equivalent to (108), stochastic system:

$$d\boldsymbol{x} = (\mathbf{A}_c(\theta) + \mathbf{B}_c u) dt + \mathbf{G}_c d\boldsymbol{w}, \tag{109}$$

$$y_k = \mathbf{C}\boldsymbol{x}_k + s_v v_k, \tag{110}$$

where

$$\mathbf{A}_c(\theta) = \begin{bmatrix} -1 & \theta \\ -\theta & -1 \end{bmatrix}, \mathbf{B}_c = \begin{bmatrix} 0 \\ b_c \end{bmatrix}, \mathbf{G}_c = \sqrt{2}\mathbb{I}, \mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \tag{111}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

$\boldsymbol{x}_k = \boldsymbol{x}(t_k)$, $t_k = k\Delta$, $\Delta = 5.7471 \cdot 10^{-3}$, $s_v = 11.85$, $b_c = 10^5$. The input signal u in (109) corresponds to $\mathcal{E}$ in (108). The parameter θ in (109) is related to the Larmor frequency ω_L in (108), by formula $\theta = \omega_L T_2$. Since the estimation error of the parameter θ depends on the input signal u, a natural question arises as to what form this signal should take. To solve this problem, we will go to discrete time and apply the methodology described in Sections 4 and 5. Assuming that $u(t) = u_k, t \in [t_k, t_{k+1}]$, the discrete-time system corresponding to (109) has the form

$$\boldsymbol{x}_{k+1} = \mathbf{A}(\theta)\boldsymbol{x}_k + \mathbf{B}(\theta)u_k + \mathbf{G}w_k, \tag{112}$$

where, according to the procedure given in Appendix C:

$$\mathbf{A}(\theta) = e^{-\Delta}\begin{bmatrix} \cos\theta\Delta & \sin\theta\Delta \\ -\sin\theta\Delta & \cos\theta\Delta \end{bmatrix}, \mathbf{B}(\theta) = \frac{b_c}{1+\theta^2}\begin{bmatrix} \theta - e^{-\Delta}(\theta\cos\theta\Delta + \sin\theta\Delta) \\ 1 - e^{-\Delta}(\cos\theta\Delta - \theta\sin\theta\Delta) \end{bmatrix}, \tag{113}$$

$$\mathbf{G} = \sqrt{1 - e^{-2\Delta}}\mathbb{I}. \tag{114}$$

We assume that the prior distribution of θ is Gaussian, that is, $p_0(\theta) = \mathcal{N}(\theta, m_\theta, s_\theta)$ with $m_\theta = 54.6637$, $s_\theta = 10.76$, which corresponds to the Larmor frequency of 10 kHz and its initial uncertainty of the order of 600 Hz (3σ). At the beginning of the process, the system is in thermal equilibrium, that is, $p(\boldsymbol{x}_0|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_0, 0, \mathbb{I})$. Following Lemma 2, we get $\theta_1 = m_\theta - \sqrt{s_\theta} \approx 51$, $\theta_2 = m_\theta + \sqrt{s_\theta} \approx 58$, $p_{0,1} = p_{0,2} = 0.5$. Thus, $r = 2$ in (31) and, accordingly to (35), the Bayesian optimal signal is a solution of the simplified and convex optimization problem (36) with $d_{1,2}$ defined by Lemmas 3 and 4. Moreover, since the matrices in (113), (114) do not depend on u_k, the last two terms in (77) can be omitted. The set of admissible signals is given by (3) with $\tilde{\boldsymbol{U}} = 0$, that is, the signal norm cannot be greater than ϱ.

In order to employ the classical methods described in Sect. 5, it is necessary to first evaluate the sensitivity of the prediction error. Similarly as in the previous example, the polynomials A, B, C have the form

$$A(\theta, z) = 1 - 2e^{-\Delta}\cos(\theta\Delta)z^{-1} + e^{-2\Delta}z^{-2}, \tag{115}$$

$$B(\theta, z) = \mathbf{B}_2(\theta) - e^{-\Delta}(\mathbf{B}_1(\theta)\sin(\theta\Delta) + \mathbf{B}_2(\theta)\cos(\theta\Delta))z^{-1}, \tag{116}$$

$$\begin{aligned} C(\theta, z) &= 1 + \Big(\mathbf{K}_2(\theta) - 2e^{-\Delta}\cos(\theta\Delta)\Big)z^{-1} + \\ &+ e^{-\Delta}\Big(e^{-\Delta} - \mathbf{K}_1(\theta)\sin(\theta\Delta) - \mathbf{K}_2(\theta)\cos(\theta\Delta)\Big)z^{-2}, \end{aligned} \tag{117}$$

and the Kalman gain $\mathbf{K}$ and the vector $\mathbf{B}$ are given by (82) and (113), respectively. The sensitivity equation (87) is given by (106). The D-optimal signal is then obtained by maximization of the averaged D-optimal criterion (107) with constraints (3).

The optimal input signals were designed by maximizing the Bayesian criterion (44), the averaged D-optimal criterion (107), and the spectral criterion (90), subject to the constraint (3) with $\tilde{\boldsymbol{U}} = 0$. The results are presented in Figures 5 and 6. Figure 6 also shows the estimation error for the step (constant) signal and harmonic signal $u(t) = a\cos(m_\theta t)$. Frequency of the harmonic signal was equal to the expected value of the a priori distribution of the parameter θ. The constant, MOOSE and harmonic signals were always assigned a norm equal to ϱ.

Maximization of the spectral criterion (90) was performed using the MOOSE-2 solver [38], evaluated at $\boldsymbol{\theta} = \boldsymbol{m}_\theta$ with default parameters, that is, the input spectrum was FIR-type with 20 lags and the spectrum power constraint was set to 1 (prob.spectrum.signal.power.ub=1). There were no additional constraints on the shape of the spectrum.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

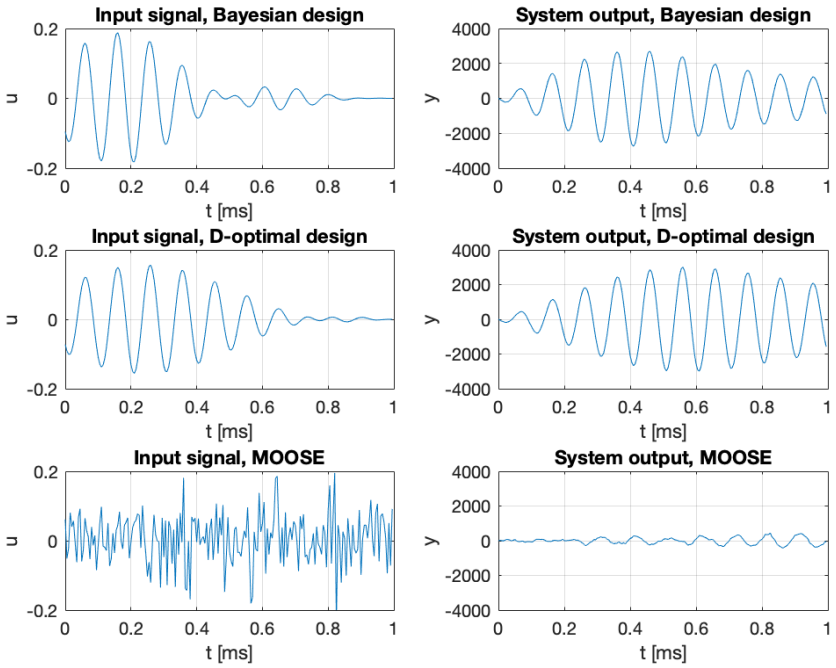

Figure 5. Optimal input signals (left) and corresponding system outputs (right) obtained by maximizing the Bayesian criterion (36) (top), the averaged D-optimal criterion (107) (middle), and the spectral criterion (90) (bottom), subject to the constraint (3) with $\tilde{\boldsymbol{u}} = 0$. Maximization of the spectral criterion (90) was performed using the MOOSE-2 solver evaluated at $\theta = m_\theta$. The norm of all signals is equal to 1, and the scale is consistent across all plots.

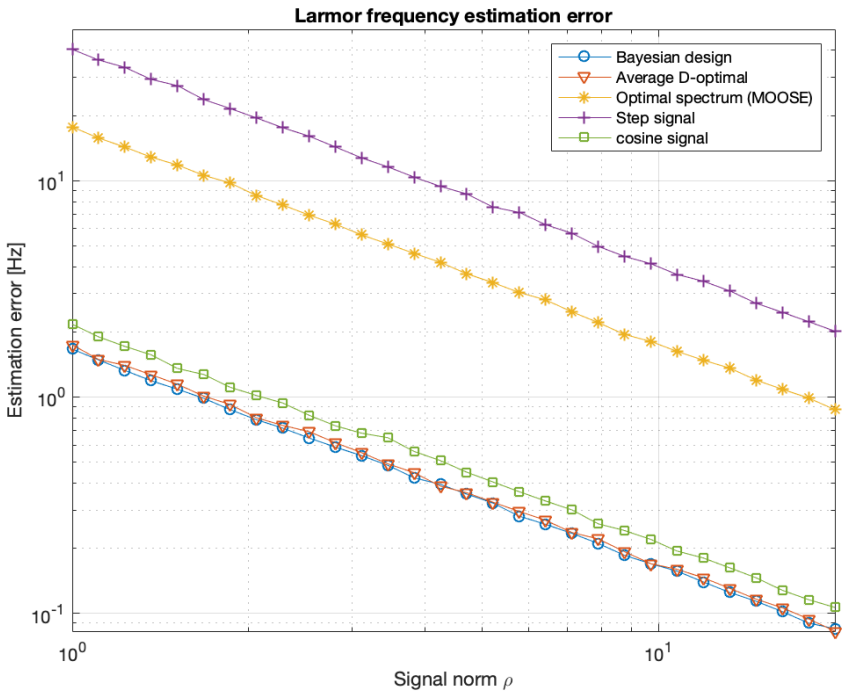

Figure 6. Mean estimation errors of the Larmor frequency $f_L = \frac{\omega_L}{2\pi}$, obtained using the MAP estimator (10), as functions of the maximum admissible signal norm ϱ for different signals. The results are based on a Monte Carlo simulation with 3000 repetitions.

Peer-reviewed version available at *Entropy* **2025**, *27*, 1041; doi:https://doi.org/10.3390/e27101041

6.4. Bayesian Input Signal Design for Pump Laser in Optically Pumped Magnetometer

Optically pumped magnetometers operate by aligning atomic spins with a circularly polarized pump laser, after which the spins precess around the external magnetic field at the Larmor frequency. The probe laser measures this precession via polarization rotation (Faraday effect), linking the detected signal to the magnetic field [39,40]. The pump laser's frequency strongly affect spin polarization and coherence time making precise laser control central to minimizing estimation error. The advanced control strategies can then suppress the noise and enhance sensitivity. Consequently, accurate control of the pumping laser is a key factor in achieving high-resolution and low-error magnetometer. We consider here the magnetometer model given by equation (S9) in the article [29]:

$$\frac{d\boldsymbol{F}}{d\tau} = (-\gamma_e \boldsymbol{B} + GS_3\hat{z}) \times \boldsymbol{F} - \gamma \boldsymbol{F} + P(\tau)(\hat{z}F_{max} - \boldsymbol{F}) + \mathbf{G}_0(P(\tau))\boldsymbol{w}, \tag{118}$$

where $\boldsymbol{F} = (F_x, F_y, F_z)^\mathsf{T}$ is the collective atomic spin, γ_e is electron gyromagnetic ratio, $\boldsymbol{B} = (B_x, B_y, B_z)^\mathsf{T}$ is constant magnetic field vector, G is known positive constant, $GS_3\hat{z}$ is the effective field produced by ac-Stark shifts due to the probe laser where S_3 is white Gaussian noise with variance σ_3^2. The optical pumping rate $P(\tau) \geqslant 0$ is an input signal. The atomic spin noise $\mathbf{G}_0(P(\tau))\boldsymbol{w}$ is modelled as white Gaussian where $\boldsymbol{w} = (w_x, w_y, w_z)^\mathsf{T}$ is a vector of standard and mutually independent Wiener increments. The $\mathbf{G}_0$ matrix is diagonal and is given by

$$\mathbf{G}_0(P(\tau)) = \sqrt{\tfrac{2}{3}F(F+1)N_A(\gamma + P(\tau))}\mathbb{I}, \tag{119}$$

where N_A is the number of atoms and F is known atomic spin number. Parameter $F_{max} = N_A F$ is the maximum possible polarization. The transverse relaxation rate γ depends on the number of atoms and is given by $\gamma(N_A) = T_2(N_A)^{-1} = \gamma_0 + 10^{-12}\alpha N_A$, where γ_0 and α are known positive constants and T_2 is the effective coherence time. The observation equation has the form

$$S_2 = F_z + N_{S_2}, \tag{120}$$

where N_{S_2} denote measurement noise with variance σ_2^2. In the experiment, the S_2 component of the Stokes vector is measured at discrete time moments $t_k = k\Delta$, where Δ is sampling period. The realistic parameters of the model are given in Table 1.

Table 1. Typical parameters.

Parameter	Abbreviation	Typical value
Number of atoms	N_A	10^{12}
Spin number	F	1
Larmor frequencies	$\gamma_e \boldsymbol{B}$	$2\pi[-50, 50]$ kHz
Parameter	γ_0	600 Hz
Parameter	α	550 Hz
Typical relaxation time	T_2	0.87 ms
Typical relaxation rate	γ	1149 Hz
Pumping rate	P	0-200 kHz
Measurement noise level	σ_2	$9.6755 \cdot 10^6$
Sampling time	Δ	5μs

In what follows, equation (118) will be interpreted in the Itö sense. **Moreover, we assume that the noise GS_3 in (118) is small and can be omitted.**

By introducing state variables $\boldsymbol{\xi} = \boldsymbol{F}\sqrt{\frac{3}{F(F+1)N_A}}$, control variable $u = P/\gamma$ and non-dimensional time $t = \gamma\tau$, and after multiplying both sides of (120) by $\sqrt{\frac{3}{F(F+1)N_A}}$, we get the following model:

$$d\boldsymbol{\xi} = (\mathbf{A}_c(\boldsymbol{\theta}, u)\boldsymbol{\xi} + \mathbf{B}_c u)dt + \mathbf{G}_c(u)d\boldsymbol{\eta}, y_k = \boldsymbol{\xi}_3(t_k) + \sigma_v v_k, \tag{121}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where $\boldsymbol{\eta}$ is the three-dimensional standard Wiener process with unit covariance and

$$\mathbf{A}_c(\boldsymbol{\theta}, u) = \begin{bmatrix} -(1+u) & \theta_3 & -\theta_2 \\ -\theta_3 & -(1+u) & \theta_1 \\ \theta_2 & -\theta_1 & -(1+u) \end{bmatrix}, \mathbf{B}_c = \begin{bmatrix} 0 \\ 0 \\ b_c \end{bmatrix}, \mathbf{G}_c(u) = \sqrt{2(1+u)}\mathbb{I}, \tag{122}$$

with $b_c = \sqrt{\frac{3N_A F}{F+1}}$, $\sigma_v = \sigma_2\sqrt{\frac{3}{F(F+1)N_A}}$. Taking the parameters from Table 1 we have $b_c = 1.22 \cdot 10^6$, $\sigma_v = 11.85$. The parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^\mathsf{T}$, represents the external magnetic field due to the relation $\boldsymbol{\theta} = \gamma_e T_2 \boldsymbol{B}$. If u is a constant signal, then system (121) approaches thermodynamic equilibrium, with $\mathbb{E}\boldsymbol{x}(t) = -\mathbf{A}_c(\boldsymbol{\theta}, u)^{-1}\mathbf{B}_c u$ and $\mathrm{cov}(\boldsymbol{x}(t)) = \mathbb{I}$.

A closer examination of equation (121) shows that the component $\boldsymbol{\xi}_3(t, \boldsymbol{\theta})$ of its solution remains invariant under rotations of the vector $\boldsymbol{\theta}$ about the z-axis. As a result, $\boldsymbol{\theta}$, and hence the field $\boldsymbol{B}$, cannot be uniquely identified from the observations $y_0, \ldots, y_N$. The only quantities that can be uniquely identified in this model are the magnitude of the vector $\boldsymbol{B}$ and the angle η between $\boldsymbol{B}$ and one of the coordinate axes, say the $\hat{z}$ axis. However, to simplify the problem as much as possible, we introduce here the additional assumption that the field $\boldsymbol{B}$ always lies in the x–y plane, that is, $\boldsymbol{B} = (B_x, B_y, 0)^\mathsf{T}$. With this assumption, the change of variables

$$x_1 = \xi_1 \sin\varphi - \xi_2 \cos\varphi, x_2 = \xi_3, \tag{123}$$

$$\cos\varphi = \frac{\theta_1}{\sqrt{\theta_1^2 + \theta_2^2}}, \sin\varphi = \frac{\theta_2}{\sqrt{\theta_1^2 + \theta_2^2}}, \tag{124}$$

reduces model (121) to a two-dimensional system:

$$d\boldsymbol{x} = (\mathbf{A}_c(\theta, u)x + \mathbf{B}_c u)dt + \mathbf{G}_c(u)d\boldsymbol{w}, y_k = \mathbf{C}\boldsymbol{x}_k + \sigma_v v_k, \tag{125}$$

where $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$, $\theta = \gamma_e T_2\sqrt{B_x^2 + B_y^2}$, $\boldsymbol{w}$ is a two-dimensional standard Wiener process with unit covariance and

$$\mathbf{A}_c(\theta, u) = \begin{bmatrix} -(1+u) & -\theta \\ \theta & -(1+u) \end{bmatrix}, \mathbf{B}_c = \begin{bmatrix} 0 \\ b_c \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \mathbf{G}_c(u) = \sqrt{2(1+u)}\mathbb{I}. \tag{126}$$

Hence, under the assumption $B_z = 0$, the observations $y_0, \ldots, y_N$, variable $\boldsymbol{\xi}_3$, and the $\boldsymbol{F}_z$ component of the collective spin, are fully characterized by the reduced model (125). Furthermore, within this reduced model it can be readily verified that θ is uniquely identifiable. Naturally, the accuracy of estimating θ depends on the choice of input u. To determine an input u that maximizes the information about θ, we now turn to the discrete-time formulation of (125) and apply the methods described in Sect. 3 and 4. Assuming the control signal is piecewise constant, that is, $u(t) = u_k, t \in [t_k, t_{k+1}]$, the process $\boldsymbol{x}_k = \boldsymbol{x}(t_k)$ satisfies the difference equation

$$\boldsymbol{x}_{k+1} = \mathbf{A}(\theta, u_k)\boldsymbol{x}_k + \mathbf{B}(\theta, u_k) + \mathbf{G}(u_k)\boldsymbol{w}_k, \tag{127}$$

where $\boldsymbol{w}_k \sim \mathcal{N}(0, \mathbb{I})$ and the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$ can be calculated following the procedure given in Appendix C. Upon completion of straightforward calculations, we get:

$$\mathbf{A}(\theta, u_k) = e^{-(1+u_k)\Delta}\begin{bmatrix} \cos\theta\Delta & -\sin\theta\Delta \\ \sin\theta\Delta & \cos\theta\Delta \end{bmatrix}, \mathbf{G}(u_k) = \sqrt{1 - e^{-2(1+u_k)\Delta}}\mathbb{I}, \tag{128}$$

$$\mathbf{B}(\theta, u_k) = \frac{b_c u_k}{(1+u_k) + \theta^2}\begin{bmatrix} e^{-(1+u_k)\Delta}(\theta\cos(\theta\Delta) + (1+u_k)\sin(\theta\Delta)) - \theta \\ e^{-(1+u_k)\Delta}(\theta\sin(\theta) - (1+u_k)\cos(\theta\Delta)) + (1+u_k) \end{bmatrix}. \tag{129}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

At the beginning of the process, the system is in a thermal equilibrium corresponding to $u \equiv 0$. Hence, $p(\boldsymbol{x}_0|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_0, 0, \mathbb{I})$. We also assume that the prior distribution of θ is Gaussian, that is, $p_0(\theta) = \mathcal{N}(\theta, m_\theta, s_\theta)$ with $m_\theta = 54.6637$, $s_\theta = 3 \cdot 10^{-3}$, which corresponds to the Larmor frequency of 10 kHz and its initial uncertainty of the order of 30 Hz (3σ). Similarly as in Sect. 6.3, we get from Lemma 2: $\theta_1 = m_\theta - \sqrt{s_\theta} \approx 54.61$, $\theta_2 = m_\theta + \sqrt{s_\theta} \approx 54.72$, $p_{0,1} = p_{0,2} = 0.5$. Since $r = 2$ in (31), then accordingly to (35), the Bayesian optimal signal is a solution of the simplified optimization problem (36) with $d_{1,2}$ calculated by Lemmas 3 and 4. Unlike in the previous examples, in this problem we maximize criterion (36) with constraints on the signal amplitude, that is, $0 \leqslant u_k \leqslant u_{max}$, which is preferable in realistic scenarios.

The results are presented in Figures 7, 8, and 9. The optimal input signal consistently lies on the boundary of the admissible set. For small $u_{\max}$, the optimal signal is rectangular, with a frequency close to the *a priori* Larmor frequency. Since large values of $u(t)$ strongly damp spin oscillations and increases the noise, the optimal signals for large $u_{\max}$ consist of short pulses at the maximum admissible amplitude. Once the oscillations decay, the system should be re-excited by a new sequence of short pulses, repeated periodically, as illustrated in Figure 9. The harmonic signal $u(t) = 0.5u_{\max}(1 + \cos(m_\theta t))$ is nearly optimal for small $u_{\max}$ but becomes ineffective for large $u_{\max}$, as it strongly damps the oscillations (see the lower right panel of Figure 8). As a result, the measurements carry less information about the Larmor frequency, and the estimation error increases despite the higher signal amplitude. An analogous behaviour is observed for rectangular signals. More generally, let $s(t) \in [0, 1]$, be any signal and define $u(t) = as(t)$ with $a \geqslant 0$. Then, as illustrated in Figure 7, estimation error reaches a minimum for some non-zero value of the parameter a.

Extending the experimental duration from 2 to 5 ms reduces the estimation error by a factor of 2 compared to the case shown in Figure 7. For $u_{\max} = 200$ and an experiment duration of 5 ms, the harmonic input signal yields an estimation error of 7 mHz, while the optimal signal, shown in the lower left panel of Figure 9, reduces the error to 0.48 mHz, that is, approximately 14 times smaller. Finally, the estimation error attains the Information-Theoretic Lower Bound (20), demonstrating that in this case the MAP estimator (10) achieves optimal performance.

It should be noted that the above models assume the Markovian environment and this condition should be checked in an experiment. To do this, one can use the criterion given in [41]. Non-Markovian models are much more complicated (see e.g. [42]) and one would need to employ a noise model with long memory. To model long-memory noise, fractional-order stochastic equations can be used instead of (118). Such models capture long-memory effects, and their noise correlation function decays slowly, for example as $t^{-1/2}$.

To implement the proposed method in real time, one can proceed as follows. First, observe that the pump signal has a simple structure, consisting of short pulses at the maximum admissible amplitude, each lasting approximately 5 μs. These pulses should be repeated with a period of roughly $2T_2$, and each pulse should be triggered when the vector $[F_y, F_z]$ forms an angle of about 30° with the z-axis (i.e., 30° before the maximum of F_z). To estimate the unknown vector $\boldsymbol{F}$ and the Larmor frequency, the MAP estimator is too slow for real-time applications. Instead, an Extended Kalman Filter (EKF) can be employed in a manner roughly similar to that described in [43] and [44]. This approach is considered feasible for implementation in an experimental setup.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

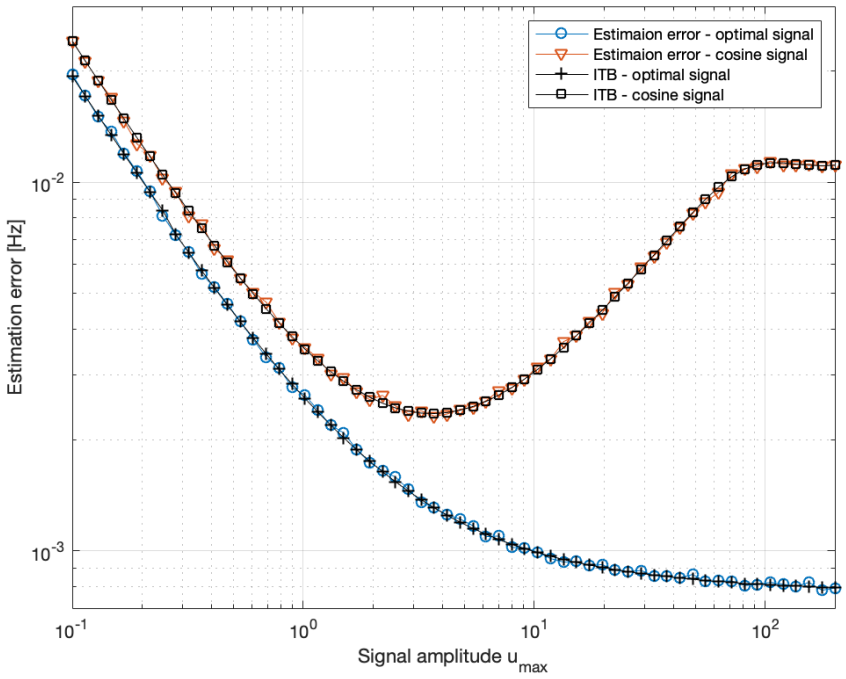

Figure 7. The estimation error of the Larmor frequency $f_L = \frac{\theta}{2\pi T_2}$ and the Information-Theoretic Bound (ITB), (20) as a function of the maximum admissible signal amplitude $u_{\max}$. The errors were computed using the MAP estimator (10). Both the errors and the ITB were calculated for two cases: (i) the optimal input signal and (ii) for the the harmonic input $u(t) = 0.5u_{\max}(1 + \cos(m_\theta t))$. Results are based on a Monte Carlo simulation with 2000 repetitions. The prior was Gaussian with mean Larmor frequency $f_L = 10$ kHz, and with its initial uncertainty $\sigma_{f_L} = 10$ Hz.

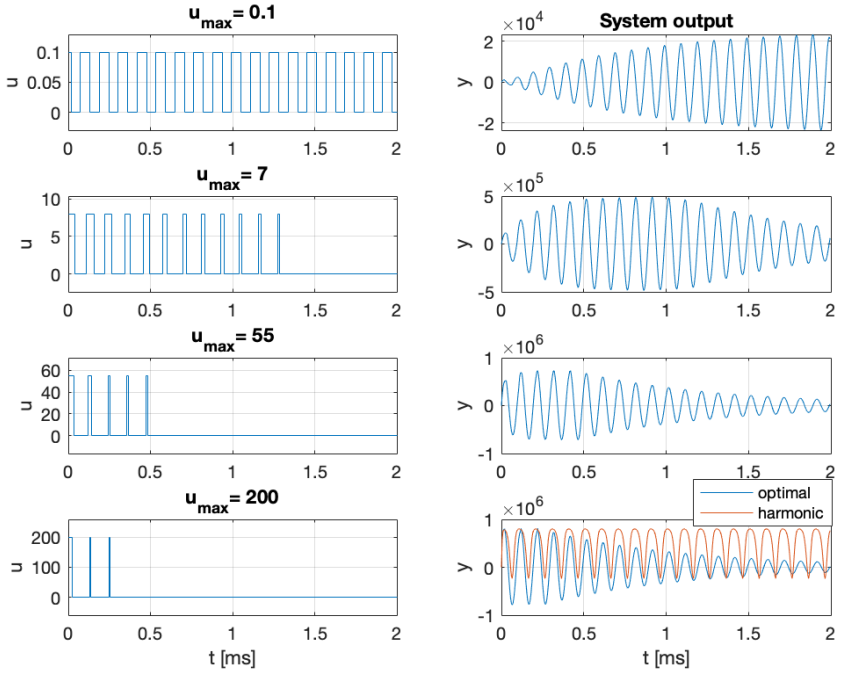

Figure 8. The optimal signals with small medium and large amplitude and the corresponding system outputs. The figure in the lower right panel also shows the system output for the harmonic signal $u(t) = 0.5u_{max}(1 + \cos(m_\theta t))$. The prior was Gaussian with mean Larmor frequency $f_L = 10$ kHz, and with its initial uncertainty $\sigma_{f_L} = 10$ Hz.

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

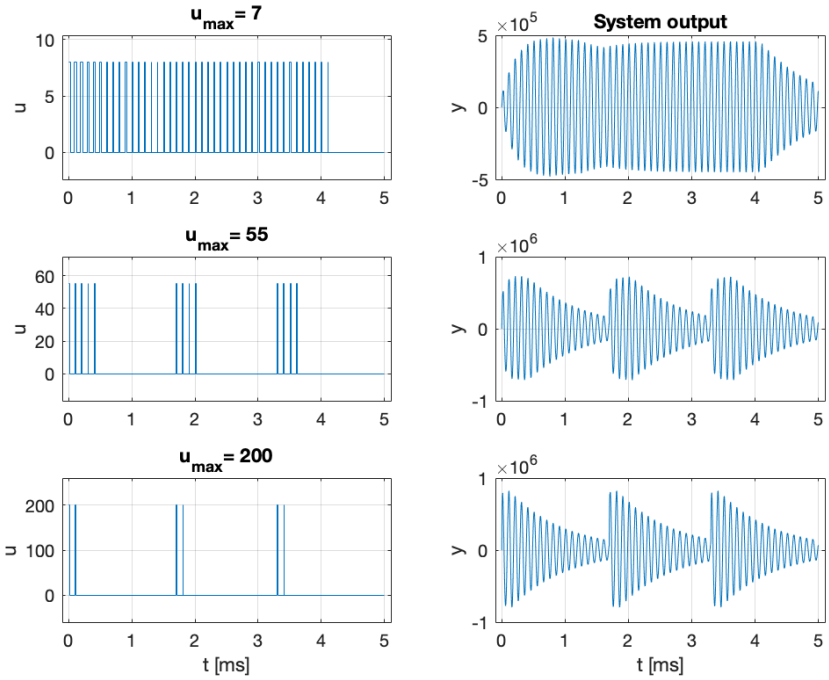

Figure 9. Optimal input signals of small, medium, and large amplitudes with corresponding system outputs for a 5 ms experiment. The prior was Gaussian with mean Larmor frequency $f_L = 10$ kHz, and with its initial uncertainty $\sigma_{f_L} = 10$ Hz.

7. Discussion and Conclusions

This paper has developed a Bayesian framework for optimal input signal design in the identification of quasi-linear stochastic dynamical systems. By an Information-Theoretic Lower Bound on estimation error and its connection to the Bayesian Cramér–Rao Bound, we showed that maximizing mutual information provides a principled alternative to Fisher-information–based criteria. The proposed method relies on the maximization of the MI lower bound (30), which produces a tractable surrogate objective for both finite parameter sets and parameter spaces of continuum cardinality. A key contribution is the algorithmic reduction of the dimension of covariance matrices required for inversion by a factor of N, making the method feasible for long-term experiments.

The comparison with the average D-optimal design highlights the practical benefits of the Bayesian approach. While classical methods are computationally efficient, they require complex differentiations to evaluate parameter sensitivities and may yield suboptimal results when parameter uncertainty is large or when the system exhibits significant non-linearities. In contrast, the proposed Bayesian method requires only the system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{G}$, together with the prior distributions of the parameter and initial conditions, without the need to calculate derivatives of prediction errors. This makes the method applicable to a much broader class of systems, while also enabling it to handle large initial parameter uncertainty.

The method also has certain limitations. The lower bound of the MI involves exponential terms that can vanish when the pairwise distance factors $d_{i,j}(\boldsymbol{U})$ are large, which can cause numerical problems. However, this drawback can be mitigated by appropriate scaling of the optimization problem. If we consider the simplified optimization problem (36), with only two candidate parameter values, these numerical problems never occur. The discrete approximation of the MI (29) is a potential source of problems and the weights and nodes in (38) should be carefully selected to achieve sufficient approximation accuracy. The third limitation arises from the fact that the maximized criterion is only a lower bound on the MI and is generally not tight. Consequently, there certainly exists a class of problems for which maximizing this lower bound is inefficient and may generate signals that are far from optimality in the sense of maximizing the MI (19)

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

In all analyzed examples, the proposed Bayesian approach, although approximate, generated signals no worse, or even better (see Figures 2 and 4), than the classical methods.

The second example illustrates that a non-Gaussian prior distribution leads to increased errors in the average D-optimal method. For a Gaussian prior distribution, it was confirmed that both the average D-optimal and the proposed Bayesian method yield identical results. This observation underscores the sensitivity of the classical approach to the form of the prior distribution and highlights the necessity of employing estimation techniques that are robust to non-Gaussian priors. In the third example, the D-optimal method produces results almost identical to those of the Bayesian approach. To explain this, note that in this problem the prior distribution of the parameter θ is relatively narrow. Then the function $d_{1,2}(\boldsymbol{U})$, which we minimize in this task, is approximately proportional to the sensitivity of the output to changes in θ. Thus, $d_{1,2}(\boldsymbol{U})$ can be interpreted as a quantity proportional to the Fisher information. Consequently, the resulting input signals and the corresponding estimation errors are nearly identical.

The study of atomic sensor models further demonstrates the practical relevance of the approach. The optimal signals in these examples are always better than the harmonic signal with a frequency equal to the expected natural frequency of the oscillator. The fourth example, a seemingly minor modification of the oscillator from the third example, shows that the dependence of the system matrices on the control signal is significant and leads to completely different optimal signals. In the analyzed examples, the MAP estimator achieves an Information-Theoretic Lower Bound (20), but this is not always the case and, depending on the task, there are better estimators. Unfortunately, finding them is difficult.

Since the method produces the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{x}_k|\boldsymbol{Y}_k)$, it can be easily converted to a sequential Bayesian Adaptive Design (BAD) algorithm [5,6]. Then the optimal strategy is a functional of the posterior, that is, $\boldsymbol{u}_k = \boldsymbol{\phi}_k(p(\boldsymbol{\theta}|\boldsymbol{Y}_k), \boldsymbol{m}_k(\boldsymbol{\theta}), \boldsymbol{S}_k(\boldsymbol{\theta}))$. In the simplest case, when the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$ do not depend on $\boldsymbol{u}_k$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta_1}, \boldsymbol{\theta}_2\}$, the optimal strategy $\boldsymbol{\phi}_k$ can be determined by maximizing (30) on the trajectories of the system (71)-(75). This problem is deterministic, therefore, relatively simple, and can be solved using deterministic optimal control methods.

From a broader perspective, quasi-linear systems arise naturally in quantum mechanics, chemical engineering, and thermal processes, making the proposed method widely applicable. In conclusion, this work provides both theoretical justification and practical tools for Bayesian input design in quasi-linear stochastic systems. By bridging information-theoretic principles with efficient computational methods, it establishes a foundation for robust experimental design in a wide range of applications. The results reported here should stimulate further research at the intersection of Bayesian inference, control, and the identification of non-linear systems.

Author Contributions: Article concept, proofs of Theorem 1 and Lemmas 2, 3, 4, Appendices A, B and C, MATLAB codes and the implementation of Bayesian and classical signal selection methods, development of all examples, comparison with classical methods, all formulas derivations, text writing: Piotr Bania. Determination of the optimal spectrum and signal generation using the MOOSE-2 toolbox in Sect. 6.1, 6.3, verification of the correctness of formulas describing discrete systems in Sect. 6.1, 6.3, 6.4 and verification of the proofs of Lemmas 3,4: Anna Wójcik. Text proofreading, literature review, introduction, discussion and conclusions: Piotr Bania and Anna Wójcik.

Funding: This research was supported by the statutory subsidy of the AGH University of Science and Technology, No. 16.16.120.773 and by the Initiative of Excellence – Research University (IDUB) program.

Data Availability Statement: The MATLAB codes, in particular the functions for calculating the lower bound (31) and the $d_{i,j}$ in (32), (36), are available on GitHub repository.

Acknowledgments: The authors gratefully acknowledge Morgan Mitchell, Jan Kołodyński, Klaudia Dilcher, Aleksandra Sierant, Julia Amorós-Binefa and Diana Méndez-Ávalos for insightful discussions on quantum control and atomic magnetometers.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript: FIM - Fisher Information Matrix; CRB - Cramér-Rao Bound; BCRB - Bayesian Cramér-Rao Bound; ITB - Information-Theoretic Bound; DOE - Design of Experiment; SDE - Stochastic Differential Equation. The norm of the vector $x \in \mathbb{R}^n$ is denoted by $|x|$. For any square matrix $\mathbf{Q}$, the quadratic form $\boldsymbol{x}^{\mathsf{T}}\mathbf{Q}\boldsymbol{x}$ is denoted by $|\boldsymbol{x}|^2_{\mathbf{Q}}$. The trace and determinant of the matrix $\mathbf{A}$ are denoted by $\mathrm{tr}\mathbf{A}$, $|\mathbf{A}|$ or $\det(\mathbf{A})$. The set of symmetric, positive definite matrices of dimension n is denoted by $\mathbf{S}^+(n)$. The symbol $\mathrm{col}(\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n)$ denotes the column vector. $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{m}, \mathbf{S})$ means that $\boldsymbol{\xi}$ has a normal distribution with mean $\boldsymbol{m}$ and covariance $\mathbf{S}$. The density of the Gaussian variable is denoted by $\mathcal{N}(\boldsymbol{x}, \boldsymbol{m}, \mathbf{S}) = 2\pi^{-\frac{n}{2}}|\mathbf{S}|^{-\frac{1}{2}}\exp(-0.5(\boldsymbol{x}-\boldsymbol{m})^{\mathsf{T}}\mathbf{S}^{-1}(\boldsymbol{x}-\boldsymbol{m}))$.

Appendix A. Proofs

Proof of Theorem 1. Let $\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U}) = \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{U})}(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{U})$ be the Minimum Mean Squared Error estimator of $\boldsymbol{\theta}$. The conditional covariance of $\hat{\boldsymbol{\theta}}_M$ is given by $\mathbf{C}(\boldsymbol{Y},\boldsymbol{U}) = \int p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{U})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U}))(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U}))^{\mathsf{T}}d\boldsymbol{\theta}$. Since Gaussian distribution maximizes entropy over all distributions with the same covariance, it can be proved, (see [22][Thm. 8.6.5, p. 255]) that

$$H_{\theta|Y}(\boldsymbol{U}) = \mathbb{E}(-\ln p(\boldsymbol{\theta}|\boldsymbol{Y},\boldsymbol{U})) \leqslant \tfrac{1}{2}\mathbb{E}\ln((2\pi e)^{n_\theta}|\mathbf{C}(\boldsymbol{Y},\boldsymbol{U})|). \tag{A1}$$

Any covariance matrix $\mathbf{C}$ satisfies the inequality $|\mathbf{C}| \leqslant (n_\theta^{-1}\mathrm{tr}(\mathbf{C}))^{n_\theta}$ (see [22], Thm. 17.9.4, p. 680). Hence

$$\ln|\mathbf{C}(\boldsymbol{Y},\boldsymbol{U})| \leqslant n_\theta \ln n_\theta^{-1}\mathrm{tr}\mathbf{C}(\boldsymbol{Y},\boldsymbol{U}). \tag{A2}$$

Taking into account (19) and Eqs. (A1-A2) we have

$$H_\theta - I_{\theta;Y}(\boldsymbol{U}) = H_{\theta|Y}(\boldsymbol{U}) = \mathbb{E}(-\ln p(\theta|\boldsymbol{Y},\boldsymbol{U})) \leqslant$$
$$\tfrac{1}{2}\mathbb{E}\ln((2\pi e)^{n_\theta}|\mathbf{C}(\boldsymbol{Y},\boldsymbol{U})|) \leqslant \tfrac{n_\theta}{2}\mathbb{E}\ln(2\pi e n_\theta^{-1}\mathrm{tr}\mathbf{C}(\boldsymbol{Y},\boldsymbol{U})).$$

By concavity of the logarithm and from Jensen's inequality

$$H_\theta - I_{\theta;Y}(\boldsymbol{U}) \leqslant \tfrac{n_\theta}{2}\ln(2\pi e n_\theta^{-1}\mathbb{E}\mathrm{tr}\mathbf{C}(\boldsymbol{Y},\boldsymbol{U})).$$

Using the equality $\mathbb{E}\mathrm{tr}\mathbf{C}(\boldsymbol{Y},\boldsymbol{U}) = \mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U})|^2$, yields

$$H_\theta - I_{\theta;Y}(\boldsymbol{U}) \leqslant \tfrac{n_\theta}{2}\ln(2\pi e n_\theta^{-1}\mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U})|^2).$$

Since $\hat{\boldsymbol{\theta}}_M$ is MMSE then $\mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_M(\boldsymbol{Y},\boldsymbol{U})|^2 \leqslant \mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})|^2$ and

$$H_\theta - I_{\theta;Y}(\boldsymbol{U}) \leqslant \tfrac{n_\theta}{2}\ln(2\pi e n_\theta^{-1}\mathbb{E}|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\boldsymbol{Y},\boldsymbol{U})|^2),$$

which is equivalent to the first inequality in (20). The proof of the Efroimovitch inequality, that is, the second inequality in (20), is given in [23][Cor. 3, Ch. 2.2, p. 16]. □

Proof of Lemma 3. The proof of (59), (63)-(68) is well documented in the literature and can be found in [19] and [45][Thm. 12.3, p. 187]. However, for completeness and the convenience of the reader, we reproduce it here in its entirety. Let us denote $\boldsymbol{X}_k = \mathrm{col}(x_0,\ldots,x_k)$, $\boldsymbol{Y}_k = \mathrm{col}(x_0,\ldots,x_k)$. We begin by recalling the filtering equations. If $\boldsymbol{\theta}$ is fixed parameter, then the solution of equation (45) is a Gauss-Markov process with transition density

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1},\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_k, \mathbf{A}_{k-1}\boldsymbol{x}_{k-1}+\mathbf{B}_{k-1}, \mathbf{A}_{k-1}\mathbf{S}_{k-1}\mathbf{A}_{k-1}^{\mathsf{T}}+\mathbf{G}_{k-1}\mathbf{G}_{k-1}^{\mathsf{T}}), \tag{A3}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where we used the notation of Sect.4 and we will omit the arguments $\boldsymbol{U}$ and $\boldsymbol{u}_k$, in all the formulas below. It follows from (46), that the density of the observations $\boldsymbol{y}_k$, conditioned on $\boldsymbol{X}_k$, $\boldsymbol{Y}_{k-1}$, $\boldsymbol{\theta}$, has the form:

$$p(\boldsymbol{y}_k|\boldsymbol{X}_k, \boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}_k, \mathbf{C}\boldsymbol{x}_k, \mathbf{S}_v(\boldsymbol{\theta})). \tag{A4}$$

According to the assumptions given at the beginning of Sect. 4, the initial distribution of $\boldsymbol{x}_0$ is given by:

$$p(\boldsymbol{x}_0|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{m}_0^-(\boldsymbol{\theta}), \mathbf{S}_0^-(\boldsymbol{\theta})), \tag{A5}$$

where $\boldsymbol{m}_0^-$, $\mathbf{S}_0^-$ are smooth functions and $\mathbf{S}_0^-(\boldsymbol{\theta}) > 0$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. To find $p(\boldsymbol{Y}_k|\boldsymbol{\theta})$ we proceed as follows:

$$\begin{aligned}
p(\boldsymbol{x}_k, \boldsymbol{Y}_k|\boldsymbol{\theta}) &= \int p(\boldsymbol{y}_k, \boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1}|\boldsymbol{\theta}) d\boldsymbol{x}_{k-1} = \\
&= \int p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1}|\boldsymbol{\theta}) d\boldsymbol{x}_{k-1} = \\
&= p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{k-1}, \boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_{k-1} = \\
&= p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_{k-1} p(\boldsymbol{Y}_{k-1}|\boldsymbol{\theta}) = \\
&= p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{Y}_{k-1}|\boldsymbol{\theta}),
\end{aligned} \tag{A6}$$

where

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_{k-1}|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_{k-1}, \tag{A7}$$

is the so called predictive distribution or prediction step. Integration of both sides of (A6) over $\boldsymbol{x}_k$ yields

$$p(\boldsymbol{Y}_k|\boldsymbol{\theta}) = p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{Y}_{k-1}|\boldsymbol{\theta}), \tag{A8}$$

where

$$p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_k, \tag{A9}$$

is the predictive distribution of $\boldsymbol{y}_k$. Dividing (A6) by (A8), gives the correction step:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k, \boldsymbol{\theta}) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta})}. \tag{A10}$$

Summarizing the above considerations, we have the following algorithm.
1) Set the initial conditions:

$$p(\boldsymbol{x}_0|\boldsymbol{Y}_{-1}, \boldsymbol{\theta}) = p(\boldsymbol{x}_0|\boldsymbol{\theta}), p(\boldsymbol{Y}_{-1}|\boldsymbol{\theta}) = 1. \tag{A11}$$

2) For $k = 0, 1, \ldots, N$, calculate:

$$p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_k, \tag{A12}$$

$$p(\boldsymbol{Y}_k|\boldsymbol{\theta}) = p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{Y}_{k-1}|\boldsymbol{\theta}), \tag{A13}$$

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k, \boldsymbol{\theta}) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta})}{p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_k}, \tag{A14}$$

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{Y}_k, \boldsymbol{\theta}) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\boldsymbol{Y}_k, \boldsymbol{\theta}) d\boldsymbol{x}_k. \tag{A15}$$

By substituting (A3), (A4), (A5) into (A11)-(A15), and after somewhat tedious calculations, we obtain:

$$p(\boldsymbol{x}_k|\boldsymbol{Y}_k, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x}_k, \boldsymbol{m}_k(\boldsymbol{\theta}), S_k(\boldsymbol{\theta})), \tag{A16}$$

$$p(\boldsymbol{y}_k|\boldsymbol{Y}_{k-1}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}_k, \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta}), \boldsymbol{\Sigma}_k(\boldsymbol{\theta})), \tag{A17}$$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

where $\boldsymbol{m}_k(\boldsymbol{\theta})$, $\boldsymbol{S}_k(\boldsymbol{\theta})$, $\boldsymbol{m}_k^-(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_k(\boldsymbol{\theta})$, $k = 0, 1, \ldots$, are given by the Kalman filtering equations (63)-(68). Then, by using the recursive formula (A13) and (A17), we get:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{k=0}^{N} \mathcal{N}(\boldsymbol{y}_k, \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta}), \boldsymbol{\Sigma}_k(\boldsymbol{\theta})), \tag{A18}$$

where $\boldsymbol{Y} = \mathrm{col}(\boldsymbol{y}_0, \ldots, \boldsymbol{y}_N)$. On the other side, accordingly to (55)-(58), we have:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{Y}, \boldsymbol{F}(\boldsymbol{\theta}), \mathbf{S}(\boldsymbol{\theta})) = \prod_{k=0}^{N} \mathcal{N}(\boldsymbol{y}_k, \mathbf{C}\boldsymbol{m}_k^-(\boldsymbol{\theta}), \boldsymbol{\Sigma}_k(\boldsymbol{\theta})). \tag{A19}$$

Taking the logarithm of both sides and calculating its expected value we get:

$$\begin{aligned} &\int p(\boldsymbol{Y}|\boldsymbol{\theta})(-\ln p(\boldsymbol{Y}|\boldsymbol{\theta}))d\boldsymbol{Y} = \\ &= \tfrac{1}{2}\ln\Big((2\pi e)^{n_y(N+1)}|\mathbf{S}(\boldsymbol{\theta})|\Big) = \tfrac{1}{2}\ln\left((2\pi e)^{n_y(N+1)}\prod_{k=0}^{N}|\boldsymbol{\Sigma}_k(\boldsymbol{\theta})|\right). \end{aligned} \tag{A20}$$

Hence $|\mathbf{S}| = \prod_{k=0}^{N}|\boldsymbol{\Sigma}_k|$, which proves (61). Now, taking the logarithm of (A19), we have:

$$\tfrac{1}{2}\ln|\mathbf{S}| + \tfrac{1}{2}|\boldsymbol{Y} - \boldsymbol{F}|^2_{\mathbf{S}^{-1}} = \frac{1}{2}\ln\left(\prod_{k=0}^{N}|\boldsymbol{\Sigma}_k|\right) + \tfrac{1}{2}\sum_{k=0}^{N}|\boldsymbol{y}_k - \mathbf{C}\boldsymbol{m}_k^-|^2_{\boldsymbol{\Sigma}_k^{-1}}, \tag{A21}$$

where the arguments has been omitted for convenience. Since $|\mathbf{S}| = \prod_{k=0}^{N}|\boldsymbol{\Sigma}_k|$, we get (60). Putting (60) and (61) into (9) gives (62). □

Proof of lemma 4. Let $\boldsymbol{\Theta} = \{\theta_1, ..., \theta_r\}, \boldsymbol{\theta}_i \in \mathbb{R}^{n_\theta}$. We are interested in calculation of the quantity

$$d_{i,j}(\boldsymbol{U}) = \tfrac{1}{8}\boldsymbol{\Delta}_{i,j}^{\mathsf{T}}\Big(\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)\Big)^{-1}\boldsymbol{\Delta}_{i,j} + \tfrac{1}{2}\ln|\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)| - \tfrac{1}{4}\ln(|\mathbf{S}_i||\mathbf{S}_j|), \tag{A22}$$

where

$$\boldsymbol{\Delta}_{i,j} = \boldsymbol{F}(\boldsymbol{\theta}_i, \boldsymbol{U}) - \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}), \mathbf{S}_i = \mathbf{S}(\boldsymbol{\theta}_i, \boldsymbol{U}), \mathbf{S}_j = \mathbf{S}(\boldsymbol{\theta}_j, \boldsymbol{U}) \tag{A23}$$

and $\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{U})$, $\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{U})$, are defined by Eqs. (55-58) of Sect. 4. Let us define

$$\boldsymbol{Y}^{(i)} = \boldsymbol{F}(\boldsymbol{\theta}_i, \boldsymbol{U}) + \boldsymbol{Z}^{(i)}, \boldsymbol{Y}^{(j)} = \boldsymbol{F}(\boldsymbol{\theta}_j, \boldsymbol{U}) + \boldsymbol{Z}^{(j)}, \tag{A24}$$

$$\tilde{\boldsymbol{Y}} = \tfrac{1}{\sqrt{2}}\Big(\boldsymbol{Y}^{(i)} - \boldsymbol{Y}^{(j)}\Big) = \tfrac{1}{\sqrt{2}}\Big(\boldsymbol{\Delta_{i,j}} + \boldsymbol{Z}^{(i)} - \boldsymbol{Z}^{(j)}\Big), \tag{A25}$$

where $\boldsymbol{Z}^{(i)} \sim \mathcal{N}(0, \mathbf{S}_i)$, $\boldsymbol{Z}^{(j)} \sim \mathcal{N}(0, \mathbf{S}_j)$. The density of variable $\tilde{\boldsymbol{Y}}$, given $\tilde{\boldsymbol{\theta}} = \mathrm{col}(\boldsymbol{\theta_i}, \boldsymbol{\theta}_j)$, has the form

$$p(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}}) = \mathcal{N}\Big(\tilde{\boldsymbol{Y}}, \tfrac{1}{\sqrt{2}}\boldsymbol{\Delta_{i,j}}, \tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)\Big). \tag{A26}$$

Now, let us consider the following two systems:

$$\boldsymbol{x}_{k+1}^{(i)} = \mathbf{A}(\boldsymbol{\theta}_i, \boldsymbol{u}_k)\boldsymbol{x}_k^{(i)} + \mathbf{B}(\boldsymbol{\theta}_i, \boldsymbol{u}_k) + \mathbf{G}(\boldsymbol{\theta}_i, \boldsymbol{u}_k)\boldsymbol{w}_k^{(i)}, \boldsymbol{y}_k^{(i)} = \mathbf{C}\boldsymbol{x}_k^{(i)} + \boldsymbol{v}_k^{(i)}, \tag{A27}$$

$$\boldsymbol{x}_{k+1}^{(j)} = \mathbf{A}(\boldsymbol{\theta}_j, \boldsymbol{u}_k)\boldsymbol{x}_k^{(j)} + \mathbf{B}(\boldsymbol{\theta}_j, \boldsymbol{u}_k) + \mathbf{G}(\boldsymbol{\theta}_j, \boldsymbol{u}_k)\boldsymbol{w}_k^{(j)}, \boldsymbol{y}_k^{(j)} = \mathbf{C}\boldsymbol{x}_k^{(j)} + \boldsymbol{v}_k^{(j)}, \tag{A28}$$

where $\boldsymbol{w}_k^{(i)}, \boldsymbol{w}_k^{(j)} \sim \mathcal{N}(0, \mathbb{I})$ and $\boldsymbol{v}_k^{(i)}, \boldsymbol{v}_k^{(j)} \sim \mathcal{N}(0, \mathbf{S}_v)$ are mutually independent. The components of the vectors $\boldsymbol{Y}^{(i)}$ and $\boldsymbol{Y}^{(i)}$ in (A24) correspond to the outputs of the systems (A27) and (A28), that is $\boldsymbol{Y}^{(i)} =$

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

$\mathrm{col}(\boldsymbol{y}_0^{(i)},\ldots,\boldsymbol{y}_N^{(i)})$, $\boldsymbol{Y}^{(j)} = \mathrm{col}(\boldsymbol{y}_0^{(j)},\ldots,\boldsymbol{y}_N^{(j)})$. Then on the basis of (A25), we get $\tilde{\boldsymbol{Y}} = \mathrm{col}(\tilde{\boldsymbol{y}}_0,\ldots,\tilde{\boldsymbol{y}}_0)$, where

$$\tilde{\boldsymbol{y}}_k = \tfrac{1}{\sqrt{2}}\left(\boldsymbol{y}_k^{(i)} - \boldsymbol{y}_k^{(j)}\right) = \tfrac{1}{\sqrt{2}}\left(\mathbf{C}\left(\boldsymbol{x}_k^{(i)} - \boldsymbol{x}_k^{(j)}\right) + \left(\boldsymbol{v}_k^{(i)} - \boldsymbol{v}_k^{(j)}\right)\right). \tag{A29}$$

Defining the matrices $\tilde{\mathbf{A}}_k$, $\tilde{\mathbf{B}}_k$, $\tilde{\mathbf{G}}_k$, $\tilde{\mathbf{C}}$, according to (69) and (70), and taking into account that $\frac{1}{\sqrt{2}}\left(\boldsymbol{v}_k^{(i)} - \boldsymbol{v}_k^{(j)}\right) \sim \mathcal{N}(0,\mathbf{S}_v)$, we can replace Eqs. (A27-A29) with a single, $2n$-dimensional system with n_y outputs:

$$\tilde{\boldsymbol{x}}_{k+1} = \tilde{\mathbf{A}}_k(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{x}}_k + \tilde{\mathbf{B}}_k(\tilde{\boldsymbol{\theta}}) + \tilde{\mathbf{G}}_k(\tilde{\boldsymbol{\theta}})\tilde{\boldsymbol{w}}_k, \tilde{\boldsymbol{y}}_k = \tilde{\mathbf{C}}\tilde{\boldsymbol{x}}_k + \boldsymbol{v}_k, \tag{A30}$$

where $\tilde{\boldsymbol{x}}_k = \mathrm{col}(\boldsymbol{x}_k^{(i)},\boldsymbol{x}_k^{(j)})$, $\tilde{\boldsymbol{w}}_k = \mathrm{col}(\boldsymbol{w}_k^{(i)},\boldsymbol{w}_k^{(j)})$ and $\boldsymbol{v}_k \sim \mathcal{N}(0,\mathbf{S}_v)$. Proceeding analogously to the proof of Lemma 3, we infer that the conditional density of variable $\tilde{\boldsymbol{Y}}$ is given by:

$$p(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}}) = \prod_{k=0}^{N}\mathcal{N}(\tilde{\boldsymbol{y}}_k, \tilde{\mathbf{C}}\tilde{\boldsymbol{m}}_k^-(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\Sigma}}_k(\tilde{\boldsymbol{\theta}})), \tag{A31}$$

where $\tilde{\boldsymbol{m}}_k^-(\tilde{\boldsymbol{\theta}})$, $\tilde{\boldsymbol{\Sigma}}_k(\tilde{\boldsymbol{\theta}})$, are calculated recursively by the Kalman filter equations

$$\tilde{\boldsymbol{\Sigma}}_k = \mathbf{S}_v + \tilde{\mathbf{C}}\tilde{\mathbf{S}}_k^-\tilde{\mathbf{C}}^\mathsf{T}, \tag{A32}$$
$$\tilde{\mathbf{L}}_k = \tilde{\mathbf{S}}_k^-\tilde{\mathbf{C}}^\mathsf{T}\tilde{\boldsymbol{\Sigma}}_k^{-1}, \tag{A33}$$
$$\tilde{\boldsymbol{m}}_k = (\mathbb{I} - \tilde{\mathbf{L}}_k\tilde{\mathbf{C}})\tilde{\boldsymbol{m}}_k^- + \tilde{\mathbf{L}}_k\tilde{\boldsymbol{y}}_k, \tag{A34}$$
$$\tilde{\mathbf{S}}_k = \tilde{\mathbf{S}}_k^- - \tilde{\mathbf{L}}_k\tilde{\boldsymbol{\Sigma}}_k\tilde{\mathbf{L}}_k^\mathsf{T}, \tag{A35}$$
$$\tilde{\boldsymbol{m}}_{k+1}^- = \tilde{\mathbf{A}}_k\tilde{\boldsymbol{m}}_k + \tilde{\mathbf{B}}_k, \tag{A36}$$
$$\tilde{\mathbf{S}}_{k+1}^- = \tilde{\mathbf{A}}_k\tilde{\mathbf{S}}_k\tilde{\mathbf{A}}_k^\mathsf{T} + \tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^\mathsf{T}, k = 0,1...,N, \tag{A37}$$

with initial conditions (76). Comparing (A26) and (A31) gives:

$$p(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}}) = \mathcal{N}\left(\tilde{\boldsymbol{Y}}, \tfrac{1}{\sqrt{2}}\Delta_{i,j}, \tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)\right) = \prod_{k=0}^{N}\mathcal{N}(\tilde{\boldsymbol{y}}_k, \tilde{\mathbf{C}}\tilde{\boldsymbol{m}}_k^-(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\Sigma}}_k(\tilde{\boldsymbol{\theta}})). \tag{A38}$$

Taking the logarithm of both sides and calculating its expected value we get:

$$\begin{aligned}&\int p(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}})(-\ln p(\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{\theta}}))d\tilde{\boldsymbol{Y}} = \\ &= \tfrac{1}{2}\ln\left((2\pi e)^{n_y(N+1)}|\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)|\right) = \tfrac{1}{2}\ln\left((2\pi e)^{n_y(N+1)}\prod_{k=0}^{N}|\tilde{\boldsymbol{\Sigma}}_k(\tilde{\boldsymbol{\theta}})|\right).\end{aligned} \tag{A39}$$

Hence

$$\ln|\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)| = \sum_{k=0}^{N}\ln|\tilde{\boldsymbol{\Sigma}}_k|. \tag{A40}$$

Taking the logarithm of (A38) and using (A40) gives:

$$\tfrac{1}{2}(\tilde{\boldsymbol{Y}} - \tfrac{1}{\sqrt{2}}\boldsymbol{\Delta}_{i,j})^\mathsf{T}\left(\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)\right)^{-1}(\tilde{\boldsymbol{Y}} - \tfrac{1}{\sqrt{2}}\boldsymbol{\Delta}_{i,j}) = \tfrac{1}{2}\sum_{k=0}^{N}|\tilde{\boldsymbol{y}}_k - \tilde{\mathbf{C}}\tilde{\boldsymbol{m}}_k^-|^2_{\tilde{\boldsymbol{\Sigma}}_k^{-1}}. \tag{A41}$$

Finally, since $\tilde{\boldsymbol{Y}} = \mathrm{col}(\tilde{\boldsymbol{y}}_0,\ldots,\tilde{\boldsymbol{y}}_0)$, then substituting $\tilde{\boldsymbol{Y}} = 0$, $\tilde{\boldsymbol{y}}_k = 0$ in (A41), (A34), we conclude that:

$$\tfrac{1}{8}\Delta_{i,j}^\mathsf{T}\left(\tfrac{1}{2}(\mathbf{S}_i + \mathbf{S}_j)\right)^{-1}\Delta_{i,j} = \tfrac{1}{4}\sum_{k=0}^{N}|\tilde{\mathbf{C}}\tilde{\boldsymbol{m}}_k^-|^2_{\tilde{\boldsymbol{\Sigma}}_k^{-1}}. \tag{A42}$$

where $\tilde{\boldsymbol{m}}_k^-$, $\tilde{\boldsymbol{\Sigma}}_k$ fulfils equations (71)-(75). The last term in (A22) is calculated according to Lemma 3. □

Peer-reviewed version available at Entropy 2025, 27, 1041; doi:https://doi.org/10.3390/e27101041

Appendix B. An example of the gap between ITB and BCRB

The difference between ITB (20) and BCRB (14) can be significant. To see this, let us consider the model

$$y = \theta + v, \tag{A43}$$

where $v \sim \mathcal{N}(0,1)$. The elementary calculation yields $J_D = 1$. If prior is Gaussian, that is, $p_0(\theta) = \mathcal{N}(\theta, m_\theta, \sigma_\theta^2)$, then $J_P = \sigma_\theta^{-2}$, $2(H_\theta - I_{\theta;y}) = \ln 2\pi e - \ln(\sigma_\theta^{-2} + 1)$ and both BCRB (14) and ITB (20) yield the same error estimate. Now, let us assume that

$$p_0(\theta) = \tfrac{1}{2}(\Phi(\alpha(1+\theta)) + \Phi(\alpha(1-\theta)) - 1), \tag{A44}$$

where $\alpha > 0$ is a parameter and $\Phi(t) = \int\limits_{-\infty}^{t} \mathcal{N}(t,0,1)dt$. The prior (A44) is an analytic function which, in the limit $\alpha \to \infty$, tends to the uniform distribution $U[-1,1]$. Since $H(y|\theta) = \frac{1}{2}\ln(2\pi e)$, $H(y) \leqslant \frac{1}{2}\ln(2\pi e(\mathrm{var}(\theta)+1))$, $\mathrm{var}(\theta) = \frac{1}{3} + O_1(\alpha^{-1})$, $H(\theta) = \ln 2 + O_2(\alpha^{-1})$ and $H_\theta - I_{\theta;y} = H(y|\theta) + H(\theta) - H(y)$, after elementary calculations, we get the ITB:

$$\mathbb{E}(\theta - \hat{\theta}(y))^2 \geqslant \frac{e^{2(H_\theta - I_{\theta;y})}}{2\pi e} \geqslant \frac{3\sqrt{2}}{8\pi e} + O(\alpha^{-1}). \tag{A45}$$

On the other hand, according to (12), we have:

$$J_P = \frac{\alpha^2}{2}\int\limits_{-\infty}^{\infty} \frac{(\mathcal{N}(\alpha(1+\theta),0,1) - \mathcal{N}(\alpha(1-\theta),0,1))^2}{\Phi(\alpha(1+\theta)) + \Phi(\alpha(1-\theta)) - 1} d\theta \geqslant \frac{\alpha}{2\sqrt{\pi}} \xrightarrow[\alpha\to\infty]{} \infty. \tag{A46}$$

Hence, BCRB (14) becomes trivial, but ITB still gives a reasonable error estimate. If the likelihood is non-Gaussian, a similar effect occurs. Therefore, BCRB generally underestimates the minimum possible estimation error.

Appendix C. Discretization of linear SDE

Consider continuous-time SDE

$$d\boldsymbol{x} = (\mathbf{A}_c\boldsymbol{x} + \mathbf{B}_c\boldsymbol{u})dt + \mathbf{G}_c d\boldsymbol{w}, \tag{A47}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$, $\boldsymbol{w}(t) \in \mathbb{R}^{n_w}$ is a vector of mutually independent standard Wiener processes. Let Δ denote the discretization period, and let $\boldsymbol{u}(t) = \boldsymbol{u}_k$, $t \in [t_k, t_{k+1}]$, $t_k = k\Delta$. Then process $\boldsymbol{x}_k = \boldsymbol{x}(t_k)$, fulfils the difference equation

$$\boldsymbol{x}_{k+1} = \mathbf{A}\boldsymbol{x}_k + \mathbf{B}u_k + \mathbf{G}\boldsymbol{w}_k, \tag{A48}$$

where $\boldsymbol{w}_k \sim \mathcal{N}(0, \mathbb{I}_{n_w})$ and

$$\mathbf{A} = e^{\mathbf{A}_c\Delta}, \mathbf{B} = \int\limits_0^\Delta e^{\mathbf{A}_c\tau}\mathbf{B}_c d\tau, \mathbf{D} = \mathbf{G}\mathbf{G}^\mathsf{T} = \int\limits_0^\Delta e^{\mathbf{A}_c\tau}\mathbf{G}_c\mathbf{G}_c^\mathsf{T}e^{\mathbf{A}_c^\mathsf{T}\tau}d\tau. \tag{A49}$$

If $\mathbf{D} > 0$, then $\mathbf{G}$ can be determined by the Cholesky factorization of $\mathbf{D}$. In the general case we use spectral decomposition $\mathbf{D} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\mathsf{T}$, and then $\mathbf{G} = \mathbf{Q}\mathbf{\Lambda}^{0.5}$.

References

1. Goodwin, G.C.; Payne, R.L. *Dynamic System Identification: Experiment Design and Data Analysis*; Academic Press: New York, 1977.
2. Ljung, L. *System Identification: Theory for the User*, second ed.; Prentice Hall PTR, 1999.
3. Söderström, T.; Stoica, P. *System Identification*; Prentice-Hall international series in systems and control engineering, Prentice-Hall, 1989.

4. Pronzato, L. Optimal experimental design and some related control problems. *Automatica* **2008**, *44*, 303–325. https://doi.org/10.1016/j.automatica.2007.05.016.
5. Huan, X.; Jagalur, J.; Marzouk, Y. Optimal experimental design: Formulations and computations. *Acta Numerica* **2024**, *33*, 715–840. https://doi.org/10.1017/S0962492924000023.
6. Rainforth, T.; Foster, A.; Ivanova, D.R.; Bickford Smith, F. Modern Bayesian Experimental Design. *Statistical Science* **2024**, *39*, 100–114. https://doi.org/10.1214/23-STS915.
7. Fedorov, V.V.; Hackl, P. *Model-oriented design of experiments*; Vol. 125, Springer, 1997.
8. Lindley, D.V. On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics* **1956**, *27*, 986–1005.
9. Arimoto, S.; Kimura, H. Optimum input test signals for system identification - An information-theoretical approach. *International Journal of Systems Science* **1971**, *1*, 279–290. https://doi.org/10.1080/00207727108910011.
10. Chaloner, K.; Verdinelli, I. Bayesian Experimental Design: A Review. *Statistical Science* **1995**, *10*, 273–304.
11. Ryan, E.; Drovandi, C.; McGree, J.; Pettitt, A. A Review of Modern Computational Algorithms for Bayesian Optimal Design. *International Statistical Review* **2015**, *84*.
12. Kolchinsky, A.; Tracey, B.D. Estimating Mixture Entropy with Pairwise Distances. *Entropy* **2017**, *19*. https://doi.org/10.3390/e19070361.
13. Kolchinsky, A.; Tracey, B.D. Estimating Mixture Entropy with Pairwise Distances. *arXiv preprint arXiv:1706.02419* **2017**. v4, 22 August 2018.
14. Altafini, C.; Ticozzi, F. Modeling and Control of Quantum Systems: An Introduction. *IEEE Transactions on Automatic Control* **2012**, *57*, 1898–1917. https://doi.org/10.1109/TAC.2012.2184120.
15. Dong, D.; Petersen, I.R. Quantum control theory and applications: A survey. *IET Control Theory & Applications* **2010**, *4*, 2651–2671. https://doi.org/10.1049/iet-cta.2009.0508.
16. Friedly, J.C. *Dynamic Behavior of Processes*; Prentice-Hall International Series in the Physical and Chemical Engineering Sciences, Prentice-Hall: Englewood Cliffs, NJ, 1972; p. 590.
17. Lorenz, S.; Diederichs, E.; Telgmann, R.; Schütte, C. Discrimination of Dynamical System Models for Biological and Chemical Processes. *Journal of Computational Chemistry* **2007**, *28*, 1384–1399. https://doi.org/10.1002/jcc.20585.
18. Bania, P. Bayesian Input Design for Linear Dynamical Model Discrimination. *Entropy* **2019**, *21*, 351. https://doi.org/10.3390/e21040351.
19. Bania, P.; Baranowski, J. Field Kalman Filter and its approximation. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), 2016, pp. 2875–2880. https://doi.org/10.1109/CDC.2016.7798697.
20. Bania, P. Example for equivalence of dual and information based optimal control. *International Journal of Control* **2018**. https://doi.org/10.1080/00207179.2018.1436775.
21. Baranowski, J.; Bania, P.; Prasad, I.; Cong, T. Bayesian fault detection and isolation using Field Kalman Filter. *EURASIP Journal on Advances in Signal Processing* **2017**, *2017*. https://doi.org/10.1186/s13634-017-0502-0.
22. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2 ed.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2006.
23. Lee, K.Y. New Information Inequalities with Applications to Statistics. Phd thesis, University of California, Berkeley, 2022. EECS Department, UC Berkeley Technical Report.
24. Van Trees, H.L. *Detection, Estimation and Modulation Theory*; Vol. I, Wiley, 1968.
25. Efroimovich, S.Y. Information Contained in a Sequence of Observations. *Problemy Peredachi Informatsii* **1979**, *15*, 24–39.
26. Van Trees, H.L.; Bell, K.L., Eds. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*; Wiley-IEEE Press: Hoboken, NJ, 2007.
27. Jakowluk, W. *Optimal Input Signal Design in Control Systems Identification*; Oficyna Wydawnicza Politechniki Białostockiej: Białystok, Poland, 2024. Available at: https://pb.edu.pl/oficyna-wydawnicza/wp-content/uploads/sites/4/2024/06/Optimal-input-signal-design-in-control-systems-identification.pdf.
28. Jiménez-Martínez, R.; Kołodyński, J.; Troullinou, C.; Lucivero, V.G.; Kong, J.; Mitchell, M.W. Signal Tracking Beyond the Time Resolution of an Atomic Sensor by Kalman Filtering. *Phys. Rev. Lett.* **2018**, *120*, 040503. https://doi.org/10.1103/PhysRevLett.120.040503.
29. Troullinou, C.; Shah, V.; Lucivero, V.G.; Mitchell, M.W. Squeezed-Light Enhancement and Backaction Evasion in a High-Sensitivity Optically Pumped Magnetometer. *Phys. Rev. Lett.* **2021**, *127*, 193601. https://doi.org/10.1103/PhysRevLett.127.193601.

Peer-reviewed version available at *Entropy* **2025**, *27*, 1041; doi:https://doi.org/10.3390/e27101041

30. Bobrovsky, B. Z.; Mayer–Wolf, E.; Zakai, M. Some Classes of Global Cramér–Rao Bounds. *The Annals of Statistics* **1987**, *15*, 1421–1438. https://doi.org/10.1214/aos/1176350602.
31. Jeong, M.; Dytso, A.; Cardone, M. A Comprehensive Study on Ziv-Zakai Lower Bounds on the MMSE. *IEEE Transactions on Information Theory* **2025**, *71*, 3214–3236. https://doi.org/10.1109/TIT.2025.3541987.
32. Huber, M.F.; Bailey, T.; Durrant-Whyte, H.; Hanebeck, U.D. On entropy approximation for Gaussian mixture random vectors. In Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 181–188. https://doi.org/10.1109/MFI.2008.4648062.
33. Nocedal, J.; Wright, S.J. *Numerical Optimization*, 2nd ed.; Springer Series in Operations Research and Financial Engineering, Springer: New York, NY, 2006. https://doi.org/10.1007/978-0-387-40065-5.
34. Davis, P.J.; Rabinowitz, P. *Methods of Numerical Integration*, 2nd ed.; Academic Press: Orlando, FL, 1984.
35. Stroud, A.H. *Approximate Calculation of Multiple Integrals*; Prentice Hall: Englewood Cliffs, NJ, 1971.
36. Smolyak, S.A. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics Doklady* **1963**, *4*, 240–243.
37. Jansson, H. *Experiment Design with Applications in Identification for Control*; Royal Institute of Technology (KTH), 2004.
38. Annergren, M.; Larsson, C.A. MOOSE2—A toolbox for least-costly application-oriented input design. *SoftwareX* **2016**, *5*, 96–100. https://doi.org/10.1016/j.softx.2016.05.003.
39. Fabricant, A.; Novikova, I.; Bison, G. How to build a magnetometer with thermal atomic vapor: A tutorial. *New Journal of Physics* **2023**, *25*, 025001. https://doi.org/10.1088/1367-2630/acb840.
40. Budker, D.; Jackson Kimball, D.F., Eds. *Optical Magnetometry*; Cambridge University Press, 2013. https://doi.org/10.1017/CBO9780511846380.
41. Breuer, H.P.; Laine, E.M.; Piilo, J. Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems. *Phys. Rev. Lett.* **2009**, *103*, 210401. https://doi.org/10.1103/PhysRevLett.103.210401.
42. Shen, H.Z.; Shang, C.; Zhou, Y.H.; Yi, X.X. Unconventional single-photon blockade in non-Markovian systems. *Phys. Rev. A* **2018**, *98*, 023856. https://doi.org/10.1103/PhysRevA.98.023856.
43. Magrini, L.; Rosenzweig, P.; Bach, C.; Deutschmann-Olek, A.; Hofer, S.G.; Hong, S.; Kiesel, N.; Kugi, A.; Aspelmeyer, M. Real-time optimal quantum control of mechanical motion at room temperature. *Nature* **2021**, *595*, 373–377. https://doi.org/10.1038/s41586-021-03602-3.
44. Amorós-Binefa, J.; Kołodyński, J. Noisy Atomic Magnetometry with Kalman Filtering and Measurement-Based Feedback. *PRX Quantum* **2025**, *6*, 030331. https://doi.org/10.1103/k7nk-lrwd.
45. Särkkä, S. *Bayesian Filtering and Smoothing*; Vol. 3, *Institute of Mathematical Statistics Textbooks*, Cambridge University Press, 2013. https://doi.org/10.1017/CBO9781139344203.