# Model-Based Robust Filtering and Experimental Design for Stochastic Differential Equation Systems

Guang Zhao, Xiaoning Qian, *Senior Member, IEEE*, Byung-Jun Yoon, *Senior Member, IEEE*, Francis J. Alexander, and Edward R. Dougherty, *Fellow, IEEE*

***Abstract*—We derive robust linear filtering and experimental design for systems governed by stochastic differential equations (SDEs) under model uncertainty. Given a model of signal and observation processes, an optimal linear filter is found by solving the Wiener-Hopf equation; with model uncertainty, it is desirable to derive a corresponding robust filter. This article assumes that the physical process is modeled via a SDE system with unknown parameters; the signals are degraded by blurring and additive noise. Due to time-dependent stochasticity in SDE systems, the system is nonstationary; and the resulting Wiener-Hopf equation is difficult to solve in closed form. Hence, we discretize the problem to obtain a matrix system to carry out the overall procedure. We further derive an intrinsically Bayesian robust (IBR) linear filter together with an optimal experimental design framework to determine the importance of SDE parameter(s). We apply the theory to an SDE-based pharmacokinetic two-compartment model to estimate drug concentration levels.**

***Index Terms*—Robust filtering, stochastic differential equation (SDE), model uncertainty, mean objective cost of uncertainty (MOCU), optimal experimental design (OED).**

## I. Introduction

IT IS common practice in signal processing to begin with a stochastic-process model (signal plus noise), a covariance (or power spectra) model, or a state-observation model, as with Kalman filtering. However, in a physical context, the signal model may be derived from a physical model, which can be a parameterized mathematical system. Hence, the properties of the signal, and of the resulting filter, depend on the physical model, and the signal parameters are expressed in terms of the parameters of the physical model. If there is uncertainty with regard to some parameters in the physical model, this uncertainty is propagated to the signal model, for instance, uncertainty in the covariance matrix. The key factor for this paper is that, if the uncertainty in the physical model arises from lack of scientific knowledge and the uncertainty is characterized by a prior distribution governing the uncertain (random) parameters, thereby characterizing our scientific understanding of the uncertainty, then that prior distribution continues to govern the uncertain parameters in the signal model. In summary, both the signal model and its uncertainty are dictated by the physical model, and not hypothesized independently.

We focus on optimally filtering signals generated by stochastic differential equations (SDEs) when some parameters of the SDEs are uncertain. Given an SDE, the desired (random) signal satisfies the SDE and is derived from it. Its parameters are from the SDE, and to the extent that the latter ones are uncertain, the signal is uncertain.

We apply intrinsically optimal Bayesian (IBR) filtering to the signal. An IBR filter is optimal relative to both the standard mean-square-error (MSE) for linear filtering (which leads to the Wiener-Hopf integral equation [1], [2]) and the uncertainty, that is, the prior distribution on the parameters of the SDE. With model uncertainty, the ordinary Wiener-Hopf equation is replaced by the effective Wiener-Hopf equation, which incorporates the expectation of the correlation functions across the uncertainty class.

When originally applied in [3], stationarity was assumed, thereby leading to the IBR Wiener filter expressed in terms of effective power spectra. Although a general continuous-time non-stationary effective Wiener-Hopf equation is also presented in [3], methods of solving non-stationary setups are not in discussed. Here, because the signal is nonstationary due to the SDE model in which we are operating, an IBR linear filter will have to be derived directly from the Wiener-Hopf equation, which must be done numerically. Different approaches have been proposed to approximate the nonstationary optimal linear filter. Here we discretize the effective Wiener-Hopf integral equation to obtain an approximate solution. Note that since there are no new observations, there is no state-observation pair and recursive filtering does not apply.

An IBR filter is robust in the sense that it performs best on average across the uncertainty class; however, it is not optimal relative to the true model. Since the true model is unknown, we

Manuscript received September 15, 2019; revised March 2, 2020, April 30, 2020, and May 17, 2020; accepted May 31, 2020. Date of publication June 10, 2020; date of current version June 30, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Abd-Krim Seghouane. The work of X. Qian was supported in part by the National Science Foundation (NSF) Awards 1553281and 1835690. The work of B.-J. Yoon was supported in part by the NSF Award 1835690. The work of E. R. Dougherty and F. J. Alexander was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical Multifaceted Integrated Capability Centers program under Award DE-SC0019303. *(Corresponding author: Guang Zhao.)*

Guang Zhao, Xiaoning Qian, and Edward R. Dougherty are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: guangzhao@tamu.edu; xqian@ece.tamu.edu; edward@ece.tamu.edu).

Byung-Jun Yoon is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA, and also with the Brookhaven National Laboratory, Upton, NY 11973 USA (e-mail: bjyoon@ece.tamu.edu).

Francis J. Alexander is with the Brookhaven National Laboratory, Upton, NY 11973 USA (e-mail: falexander@bnl.gov).

Digital Object Identifier 10.1109/TSP.2020.3001384

quantify the cost of uncertainty relative to the MSE by averaging the loss of performance across the uncertainty class. The *mean objective cost of uncertainty* (MOCU) is the average increase in error across the uncertainty class arising from using an IBR filter rather than individual optimal filters for each model in the uncertainty class [4]. The MOCU provides a cost of the uncertainty relative to the objective. A key aspect of the work is to apply MOCU-based optimal experimental design [5] in the framework of signal models derived from physical models: determine the unknown parameter in the SDE physical system model whose experimental assessment optimally reduces the expected (residual) MOCU when the obtained parameter value is put into the model and a new IBR filter is derived. The procedure can be done iteratively to yield sequential experimental design. We would like to emphasize that here we have two types of data. One is the observations of the signals generated from the SDE model, which we aim to estimate using our IBR filter; the other is the experimental assessments of the parameters in the SDE model, which help reduce the model uncertainty and MOCU by experimental design.

The general idea is to tie physical modeling, optimal signal processing, and experimental design (here in the SDE framework). We will demonstrate aspects of the procedure via a synthetic example. Then we will apply it to an SDE-based pharmacokinetic two-compartment model that differentiates the body into a central compartment (plasma) and a peripheral compartment (tissues), and describes the relationship between the drug concentrations in the central and peripheral compartments, and the measurement of the drug concentration in the central compartment.

## II. Background

In this section, we briefly review the background for IBR operators and optimal experimental design in the IBR context.

### A. IBR Filtering

Optimal operator design involves a mathematical model for the physical system, a class of operators, and an optimization problem defined by a cost function:

$$\psi_{\text{opt}} = \arg\min_{\psi\in\mathcal{F}} C(\psi), \tag{1}$$

where $\mathcal{F}$ is the operator class and $C(\psi)$ is the cost of applying an operator $\psi$. With model uncertainty, the true model is assumed to belong to an *uncertainty class* of models parameterized by a vector $\theta \in \Theta$. We define an *intrinsically Bayesian robust (IBR) operator* by

$$\psi^{\Theta}_{\text{IBR}} = \arg\min_{\psi\in\mathcal{F}} \mathbb{E}_{\Theta}[C_\theta(\psi)], \tag{2}$$

where each $\theta \in \Theta$ corresponds to a model and the *prior probability distribution* $\pi(\theta)$ quantifies our prior knowledge regarding the physical system. Note that the expectation is with respect to $\pi(\theta)$ on the uncertainty class $\Theta$ and $C_\theta(\psi)$ denotes the corresponding cost of applying $\psi$ to the model $\theta$ [3], [4]. If there is no prior knowledge beyond the uncertainty class itself, then the prior can be taken to be uniform and $\pi(\theta)$ is noninformative. An IBR operator is robust in the sense that on average it performs well over the uncertainty class $\Theta$.

When there is a data sample $S$, the prior can be updated to a posterior distribution $\pi^*(\theta) = \pi(\theta|S)$, and (2) then defines an *optimal Bayesian operator (OBO)* $\psi^{\Theta}_{\text{OBO}}$ [6], [7]. An IBR operator is an OBO with $S = \varnothing$, namely, when there is no data but only prior knowledge constraining the model $\theta \in \Theta$.

In the signal filtering problem, the operators mentioned above are just filters. Signal filtering involves a joint random process $(X(t), Y(s))$, $t \in T, s \in S$, and optimal filtering involves estimating the signal $Y(s)$ at time $s$ via a filter $\psi$ given observations $\{X(t)\}_{t\in T}$. A filter $\psi \in \mathcal{F}$ is a mapping on the space $\mathcal{S}$ of possible observed signals and a cost function takes the form $C(Y(s), \hat{Y}(s))$, with $\hat{Y}(s) = \psi(X)(s)$. For fixed $s \in S$, an optimal filter is defined by (1) with $C(\psi) = C(Y(s), \psi(X)(s))$. With uncertainty, there is an *uncertainty class* $\{(X_\theta(t), Y_\theta(s)), t \in T, s \in S, \theta \in \Theta\}$. An IBR filter, or optimal Bayesian filter, is defined by (2) with $C_\theta(\psi) = C_\theta(Y_\theta(s), \psi(X_\theta)(s))$ [3], [6].

Finding IBR filters involves developing a theory by which (2) can be solved – in a similar way as (1) is solved except that the *effective characteristics* pertaining to the full uncertainty class are used rather than the characteristics of a single signal model. An observation-signal pair $(X(t), Y(s))$ is *solvable* under the function class $\mathcal{F}$ and cost $C$ if there exists a solution to (1) under the processes. An observation-signal pair $(X_\Theta(t), Y_\Theta(s))$ is referred to as an *effective process* under the function class $\mathcal{F}$, uncertainty class $\Theta$, and costs $C$ and $C_\theta$ if for all $\psi \in \mathcal{F}$,

$$\mathbb{E}_\Theta[C_\theta(Y_\theta(s), \psi(X_\theta)(s))] = C(Y_\Theta(s), \psi(X_\Theta)(s)). \tag{3}$$

If there exists a solvable effective process $(X_\Theta(t), Y_\Theta(s))$ with the optimal filter $\psi_\Theta$, then $\psi^{\Theta}_{\text{IBR}} = \psi_\Theta$ [3].

Robust filter design goes back to the late 1970s, with robust Wiener filtering involving minimax optimality in regard to uncertain power spectra [8]–[11]. Robust design was extended to nonlinear filters and placed into a Bayesian framework by assuming a prior probability distribution governing the uncertainty class, the aim being to find a filter with minimal expected error across the uncertainty class [12]. IBR filters are fully optimal under this framework.

Other robust filters include a minimax estimator ($\tau$-robust) associated with $\tau$-divergence space [13], a minimax estimator under covariance uncertainty with the given eigenvector matrix and bounded eigenvalues [14], a minimax estimator with an uncertain model matrix [15], and a distributed estimation formulation with model uncertainties [16].

Although we are not using recursive filters, for the sake of completeness we mention some robust Kalman filters. Adaptive Kalman filters simultaneously estimate the noise covariances along with the state estimation [17], [18]. Finite-impulse-response analogues have also been proposed [19], [20]. A regularized least-squares framework has been employed in which unknown parameters embody the deviation of the model parameters from their nominal values [21]. Another approach penalizes sensitivity of estimation relative to modeling error [22]. It has also been extended to the situation in which the observation

can be randomly lost [23]. Last but not least, robust Kalman filtering has been addressed in the IBR framework [24], [25].

### B. Experimental Design

While an IBR operator is optimal over the uncertainty class, it is likely to be suboptimal relative to the true model. This performance loss is the cost of uncertainty. For any $\theta \in \Theta$ and operator family $\mathcal{F}$, the *objective cost of uncertainty* relative to $\theta$ is $C_\theta(\psi_{\mathrm{IBR}}^{\Theta}) - C_\theta(\psi_\theta)$. The *mean objective cost of uncertainty* (MOCU) [4] is the expectation of this cost over all possible models:

$$\mathrm{M}_{\mathcal{F}}(\Theta) = \mathbb{E}_{\Theta}[C_\theta(\psi_{\mathrm{IBR}}^{\Theta}) - C_\theta(\psi_\theta)]. \tag{4}$$

While we have defined MOCU for an IBR operator relative to the prior, it can also be defined for an OBO relative to the posterior.

MOCU is used to choose experiments to optimally reduce the model uncertainty relevant to the operational objective. For example, given $k$ experiments $T_1, ..., T_k$, where experiment $T_i$ exactly determines the uncertain parameter $\theta_i$ in $\theta = (\theta_1, \theta_2, ..., \theta_k)$, the issue for experimental design is which experiment to conduct first. Let $\theta|\bar{\theta}_i = \theta|(\theta_i = \bar{\theta}_i)$ be the *conditional uncertainty vector* composed of all uncertain parameters other than $\theta_i$ with $\theta_i = \bar{\theta}_i$. $\Theta|\bar{\theta}_i = \{\theta|\bar{\theta}_i : \theta \in \Theta\}$ is the *reduced uncertainty class* given $\theta_i = \bar{\theta}_i$. The IBR operator for $\Theta|\bar{\theta}_i$ is denoted $\psi_{\mathrm{IBR}}^{\Theta|\bar{\theta}_i}$ and is called the *reduced IBR operator* relative to $\bar{\theta}_i$.

If the experiment $T_i$ obtains the model parameter value $\bar{\theta}_i$, then the *remaining MOCU* given $\theta_i = \bar{\theta}_i$ is

$$\mathrm{M}_{\mathcal{F}}(\Theta|\bar{\theta}_i) = \mathbb{E}_{\Theta|\bar{\theta}_i}[C_{\theta|\bar{\theta}_i}(\psi_{\mathrm{IBR}}^{\Theta|\bar{\theta}_i}) - C_{\theta|\bar{\theta}_i}(\psi_{\theta|\bar{\theta}_i})], \tag{5}$$

where the expectation is relative to the conditional distribution $\pi(\theta|\bar{\theta}_i)$. The remaining MOCU is the MOCU for $\psi_{\mathrm{IBR}}^{\Theta|\bar{\theta}_i}$ relative to $\Theta|\bar{\theta}_i$.

Treating the remaining MOCU as a function of $\theta_i$ and taking the expectation with respect to $\pi(\theta_i)$ yields the expected remaining MOCU, given parameter $\theta_i$,

$$\mathbb{E}_{\theta_i}[\mathrm{M}_{\mathcal{F}}(\Theta|\theta_i)] = \mathbb{E}_{\theta_i}[\mathbb{E}_{\Theta|\theta_i}[C_{\theta|\theta_i}(\psi_{\mathrm{IBR}}^{\Theta|\theta_i}) - C_{\theta|\theta_i}(\psi_{\theta|\theta_i})]], \tag{6}$$

which is called the *experimental design value* and denoted by $\mathrm{D}(\theta_i)$. An *optimal experiment* $T_{i^*}$ is defined by

$$i^* = \arg\min_{i=1,...,k} \mathrm{D}(\theta_i) = \arg\min_{i=1,...,k} \mathrm{R}(\theta_i), \tag{7}$$

where

$$\mathrm{R}(\theta_i) = \mathbb{E}_{\theta_i}[\mathbb{E}_{\Theta|\theta_i}[C_{\theta|\theta_i}(\psi_{\mathrm{IBR}}^{\Theta|\theta_i})]] \tag{8}$$

is called the *residual IBR cost* for $T_i$, and $\theta_{i^*}$ is called the *primary parameter* [26]. The resulting $T_{i^*}$ is the experiment that is expected to minimize the model uncertainty pertaining the cost. Experiments can be chosen in a greedy sequential manner by repeating the process for the remaining unknown parameters, or by using dynamical programming. This sequential experimental design procedure is illustrated in Fig. 1.

Note that in the discussion above, we assume that the experiment $T_i$ can determine $\theta_i$ exactly. The strategy can be easily extended to more general cases with imprecise experiments [27].

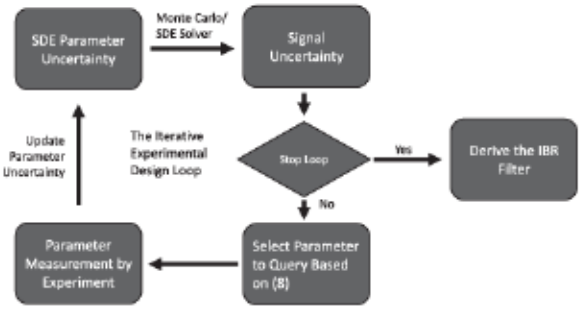



Fig. 1. MOCU-based experimental design loop for robust filtering.

When the value $\bar{\theta}_i$ obtained from the experiment $T_i$ is imprecise with distribution $p(\bar{\theta}_i|\theta_i)$, the piror $\pi(\theta)$ can be accordingly updated to a posterior distribution $\pi(\theta|\bar{\theta}_i)$. Then the IBR filter $\psi_{\mathrm{IBR}}^{\Theta|\bar{\theta}_i}$ defined by (2) is optimal with respect to the posterior $\pi(\theta|\bar{\theta}_i)$, and the residual IBR cost is still in the same form as (8).

## III. IBR Linear Filter for Nonstationary Signals

Consider an uncertain signal model $(X_\theta, Y_\theta), \theta \in \Theta$, with the MSE cost function and the class of linear functions

$$\mathcal{F} = \left\{ \psi : \psi(X)(s) = \int_T g(s,t)X(t)dt \right\}. \tag{9}$$

The solvable class $\Phi$ consists of all process pairs $(X, Y)$ such that $\psi(X)(s)$ has a finite second moment for any $g(s,t)$ and there exists $\widehat{g}(s,t)$ for which the Wiener-Hopf equation is satisfied:

$$R_{YX}(s,t) = \int_T \widehat{g}(s,u)R_X(u,t)du, \tag{10}$$

where $R_X(u,t)$ and $R_{YX}(s,t)$ are autocorrelation and cross-correlation functions, respectively.

With the uncertain signal model, we now define the *effective correlation functions* by $R_{\Theta,Y}(s,v) = \mathbb{E}_\Theta[R_{Y_\theta}(s,v)]$, $R_{\Theta,X}(t,u) = \mathbb{E}_\Theta[R_{X_\theta}(t,u)]$, and $R_{\Theta,YX}(s,t) = \mathbb{E}_\Theta[R_{Y_\theta X_\theta}(s,t)]$. As an autocorrelation function, $R_{X_\theta}(t,u)$ is conjugate symmetric and nonnegative definite for all $\theta \in \Theta$. $R_{\Theta,X}(t,u)$ has the same properties and is therefore also a valid autocorrelation function. It is straightforward to show that (3) is satisfied. If $(X_\Theta, Y_\Theta) \in \Phi$, meaning that the Wiener-Hopf equation relative to $(X_\Theta, Y_\Theta)$ is satisfied, then $(X_\Theta, Y_\Theta)$ is an effective process for the uncertainty class $\Theta$ and an IBR linear filter is given by the solution, $\widehat{g}(s,t)$, to the *effective Wiener-Hopf equation* [3]:

$$R_{\Theta,YX}(s,t) = \int_T \widehat{g}(s,u)R_{\Theta,X}(u,t)du. \tag{11}$$

All basic equations hold with characteristics replaced by effective characteristics $R_{\Theta,Y}$, $R_{\Theta,X}$, and $R_{\Theta,YX}$.

In the nonstationary case, the integral-form Wiener-Hopf equation can be difficult to solve in closed form, and numerical approximations are employed. The authors in [28] proposed a time-frequency formulation of the nonstationary linear filter, which can be approximately valid for underspread cases. It is

also possible to approximately approach it by solving discrete-time Wiener-Hopf equations [29], [30]. Here we use the discrete-time approach to approximate the continuous time signal and observation with signal vector $\boldsymbol{Y} = \{Y(s_i)\}$ at $N$ discrete time points $s_i, i \leq N$ and observation vector $\boldsymbol{X} = \{X(t_j)\}$ at $M$ discrete time points $t_j, j \leq M$. The integral form of the Wiener-Hopf equation then turns into the following matrix form:

$$\boldsymbol{R}_{YX} = \widehat{\boldsymbol{G}}\boldsymbol{R}_X, \tag{12}$$

where $\boldsymbol{R}_X = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$, $\boldsymbol{R}_{YX} = \mathbb{E}[\boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}}]$ are the autocorrelation and cross-correlation of the matrix form, respectively; and $\widehat{\boldsymbol{G}}$ is the matrix-form optimal filter. Similarly, the effective Wiener-Hopf equation in the matrix form can be written as:

$$\boldsymbol{R}_{\Theta,YX} = \widehat{\boldsymbol{G}}_{\Theta}\boldsymbol{R}_{\Theta,X}, \tag{13}$$

and the solution is

$$\widehat{\boldsymbol{G}}_{\Theta} = \boldsymbol{R}_{\Theta,YX}[\boldsymbol{R}_{\Theta,X}]^{+}, \tag{14}$$

where the superscript $+$ denotes the pseudoinverse. The error covariance matrix of $\hat{\boldsymbol{Y}}_{\mathrm{IBR}} = \widehat{\boldsymbol{G}}_{\Theta}\boldsymbol{X}_{\theta}$ is

$$\begin{aligned}
&\mathbb{E}_{\Theta}[\mathbb{E}[(\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta})(\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta})^{\mathrm{T}}]] \\
&= \mathbb{E}_{\Theta}[\mathbb{E}[(\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta})\hat{\boldsymbol{Y}}_{\mathrm{IBR}}^{\mathrm{T}}]] - \mathbb{E}_{\Theta}[\mathbb{E}[(\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta})\boldsymbol{Y}_{\theta}^{\mathrm{T}}]] \\
&= -\mathbb{E}_{\Theta}[\mathbb{E}[(\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta})\boldsymbol{Y}_{\theta}^{\mathrm{T}}]] \\
&= \mathbb{E}_{\Theta}[\mathbb{E}[\boldsymbol{Y}_{\theta}\boldsymbol{Y}_{\theta}^{\mathrm{T}}]] - \mathbb{E}_{\Theta}[\mathbb{E}[\hat{\boldsymbol{Y}}_{\mathrm{IBR}}\boldsymbol{Y}_{\theta}^{\mathrm{T}}]] \\
&= \boldsymbol{R}_{\Theta,Y} - \boldsymbol{R}_{\Theta,YX}[\boldsymbol{R}_{\Theta,X}]^{+}\boldsymbol{R}_{\Theta,YX}^{\mathrm{T}},
\end{aligned} \tag{15}$$

where the second equality holds because $\hat{\boldsymbol{Y}}_{\mathrm{IBR}} - \boldsymbol{Y}_{\theta}$ is orthogonal to any linear combination of $\boldsymbol{Y}_{\theta}$. Especially, $\hat{\boldsymbol{Y}}_{\mathrm{IBR}}$ as an IBR filter achieves the Bayesian optimality [31]. The last equality follows from (14). The MSE of the IBR filter is just the trace of the error covariance matrix.

### A. MOCU for the Discrete Wiener-Hopf Equation

With the derived IBR Wiener filter, we can quantify the model uncertainty in the MOCU framework relative to the IBR filter:

$$\begin{aligned}
M_{\mathcal{F}}(\Theta) &= \mathbb{E}_{\Theta}[C_{\theta}(\widehat{\boldsymbol{G}}_{\Theta}) - C_{\theta}(\widehat{\boldsymbol{G}}_{\theta})] \\
&= \mathbb{E}_{\Theta}[C_{\theta}(\widehat{\boldsymbol{G}}_{\Theta})] - \mathbb{E}_{\Theta}[C_{\theta}(\widehat{\boldsymbol{G}}_{\theta})] \\
&= \mathrm{tr}(\boldsymbol{R}_{\Theta,Y} - \boldsymbol{R}_{\Theta,YX}[\boldsymbol{R}_{\Theta,X}]^{+}\boldsymbol{R}_{\Theta,YX}^{\mathrm{T}}) \\
&\quad - \mathbb{E}_{\Theta}[\mathrm{tr}(\boldsymbol{R}_{\theta,Y} - \boldsymbol{R}_{\theta,YX}[\boldsymbol{R}_{\theta,X}]^{+}\boldsymbol{R}_{\theta,YX}^{\mathrm{T}})] \\
&= -\mathrm{tr}(\boldsymbol{R}_{\Theta,YX}[\boldsymbol{R}_{\Theta,X}]^{+}\boldsymbol{R}_{\Theta,YX}^{\mathrm{T}}) \\
&\quad + \mathbb{E}_{\Theta}[\mathrm{tr}(\boldsymbol{R}_{\theta,YX}[\boldsymbol{R}_{\theta,X}]^{+}\boldsymbol{R}_{\theta,YX}^{\mathrm{T}})].
\end{aligned} \tag{16}$$

Experimental design for IBR Wiener filtering involves minimizing the IBR residual cost: $i^* =$

$$\underset{i\in 1,\ldots,k}{\operatorname{argmin}}\, \mathbb{E}_{\bar{\theta}_i}[\mathrm{tr}(\boldsymbol{R}_{\Theta|\bar{\theta}_i,Y} - \boldsymbol{R}_{\Theta|\bar{\theta}_i,YX}[\boldsymbol{R}_{\Theta|\bar{\theta}_i,X}]^{+}\boldsymbol{R}_{\Theta|\bar{\theta}_i,YX}^{\mathrm{T}})] \tag{17}$$

$$= \underset{i\in 1,\ldots,k}{\operatorname{argmax}}\, \mathbb{E}_{\bar{\theta}_i}[\mathrm{tr}(\boldsymbol{R}_{\Theta|\bar{\theta}_i,YX}[\boldsymbol{R}_{\Theta|\bar{\theta}_i,X}]^{+}\boldsymbol{R}_{\Theta|\bar{\theta}_i,YX}^{\mathrm{T}})], \tag{18}$$

where $\boldsymbol{R}_{\Theta|\bar{\theta}_i,Y} = \mathbb{E}_{\Theta|\bar{\theta}_i}[\boldsymbol{R}_{Y_\theta}]$, $\boldsymbol{R}_{\Theta|\bar{\theta}_i,X} = \mathbb{E}_{\Theta|\bar{\theta}_i}[\boldsymbol{R}_{X_\theta}]$, $\boldsymbol{R}_{\Theta|\bar{\theta}_i,YX} = \mathbb{E}_{\Theta|\bar{\theta}_i}[\boldsymbol{R}_{Y_\theta X_\theta}]$, and the second equation holds for the reason that

$$\mathbb{E}_{\bar{\theta}_i}[\boldsymbol{R}_{\Theta|\bar{\theta}_i,Y}] = \mathbb{E}_{\bar{\theta}_i}[\mathbb{E}_{\Theta|\bar{\theta}_i}[\boldsymbol{R}_{Y_\theta}]] = \mathbb{E}_{\Theta}[\boldsymbol{R}_{Y_\theta}] = \boldsymbol{R}_{\Theta,Y} \tag{19}$$

is unrelated to the index $i$. As shown in (17) and (18), the IBR residual cost is only a function of the auto- and cross-correlations, in the form of the MSE of the linear IBR filter. Therefore there is no need to re-derive the filter during the experimental design procedure.

## IV. IBR Linear Filter and Experimental Design With Stochastic Differential Equations

Stochastic differential equations (SDEs) are widely applied for stochastic process modeling in areas such as pharmacology [32], population biology [33], [34] and mathematical finance [35]. In addition to the differential equations governing the processes under study, SDEs include diffusion processes to model potential random effects disturbing the processes of interest. Usually the diffusion process is a Wiener process. Assume that the $n$-dimensional random process under study, $\boldsymbol{Y}(t) \in \mathcal{Y} \subseteq \mathbb{R}^n$, is defined within the time interval $t \in [0, T]$; and the corresponding SDE is driven by an $m$-dimensional Wiener process $\boldsymbol{W}(t)$. Then the typical form of an Itô SDE is [36]:

$$\mathrm{d}\boldsymbol{Y}(t) = \boldsymbol{f}(t, \boldsymbol{Y}(t))\mathrm{d}t + \boldsymbol{g}(t, \boldsymbol{Y}(t))\mathrm{d}\boldsymbol{W}(t), \tag{20}$$

where $\boldsymbol{f} : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{g} : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are the drift vector and diffusion matrix, respectively.

If $\boldsymbol{f}$ and $\boldsymbol{g}$ are in the linear form shown in (21), the solutions of the corresponding SDEs can be Gaussian processes. Assume the functions $\boldsymbol{f}$ and $\boldsymbol{g}$ are given by

$$\begin{aligned}
\boldsymbol{f}(t, \boldsymbol{Y}(t)) &= \boldsymbol{A}(t)\boldsymbol{Y}(t) + \boldsymbol{a}(t), \\
\boldsymbol{g}(t, \boldsymbol{Y}(t)) &= \boldsymbol{B}(t),
\end{aligned} \tag{21}$$

where $\boldsymbol{A}(t)$ and $\boldsymbol{B}(t)$ are matrices of size $n \times n$ and $n \times m$, respectively, and $\boldsymbol{a}(t)$ is a vector of size $n$. The resulting SDE takes the form

$$\mathrm{d}\boldsymbol{Y}(t) = (\boldsymbol{A}(t)\boldsymbol{Y}(t) + \boldsymbol{a}(t))\mathrm{d}t + \boldsymbol{B}(t)\mathrm{d}\boldsymbol{W}(t), \boldsymbol{Y}(0) = c. \tag{22}$$

The initial-valued SDE has a unique solution if and only if the initial condition $c$ is either a constant or a Gaussian distributed random variable. The mean and auto-correlation of the Gaussian process are given by

$$\boldsymbol{m}(t_i) = \boldsymbol{\Phi}(t_i)(\mathbb{E}[c] + \int_0^{t_i} \boldsymbol{\Phi}(s)^{-1}\boldsymbol{a}(s)\mathrm{d}s) \tag{23}$$

and

$$\begin{aligned}
\boldsymbol{\Psi}(t_i, t_j) = \boldsymbol{\Phi}(t_i)\Bigg(&\mathbb{E}\left[(c - \mathbb{E}[c])(c - \mathbb{E}[c])^{\mathrm{T}}\right] \\
&+ \int_0^{t_i} \boldsymbol{\Phi}(u)^{-1}\boldsymbol{B}(u)\boldsymbol{B}(u)^{\mathrm{T}}(\boldsymbol{\Phi}(u)^{-1})^{\mathrm{T}}\mathrm{d}u\Bigg)\boldsymbol{\Phi}(t_j)^{\mathrm{T}},
\end{aligned} \tag{24}$$

where $0 \le t_i \le t_j \le T$ and $\Phi(t)$ is the fundamental matrix of the ordinary differential equation

$$\mathrm{d}\boldsymbol{Y}(t) = \boldsymbol{A}(t)\boldsymbol{Y}(t)\mathrm{d}t. \tag{25}$$

When there is no closed-form solution, approximate numerical solutions of SDEs can be obtained by the Euler-Maruyama method [37]: Partition the interval $[0, T]$ into $N$ equal subintervals of the width $\Delta t = T/N$: $0 = t_0 < t_1 < \cdots < t_N = T$. Then the numerical solution to the SDE is computed recursively by the difference equation:

$$\boldsymbol{Y}_{n+1} = \boldsymbol{Y}_n + \boldsymbol{f}(t_n, \boldsymbol{Y}_n)\Delta t + g(t_n, \boldsymbol{Y}_n)\Delta \boldsymbol{W}_n, \tag{26}$$

where $\Delta \boldsymbol{W}_n = \boldsymbol{W}_{t_{n+1}} - \boldsymbol{W}_{t_n}$ is a Gaussian random vector with independent components and the variance of each component is $\Delta t$. Monte Carlo discrete samples of $\boldsymbol{Y}(t)$ can be generated according to (26), based on which we can estimate the stochastic characteristics.

In this paper, we consider IBR filtering and optimal experimental design for the stochastic signal $\boldsymbol{Y}(t)$ described by an SDE with a vector $\theta = (\theta_1, \theta_2, ..., \theta_k)$ of uncertain parameters, so that $\boldsymbol{Y}(t)$ satisfies the SDE

$$\mathrm{d}\boldsymbol{Y}(t) = \boldsymbol{f}(t, \boldsymbol{Y}(t); \theta)\mathrm{d}t + g(t, \boldsymbol{Y}(t); \theta)\mathrm{d}\boldsymbol{W}(t). \tag{27}$$

The model uncertainty can be characterized by $\pi(\theta)$, the prior distribution of $\theta$. Denote the observation of $\boldsymbol{Y}(t)$ as $\boldsymbol{X}(t)$. Assume the observation procedure follows a linear observation model:

$$\boldsymbol{X}(t) = \int_0^T \boldsymbol{Y}(s)h(s,t)\mathrm{d}s + n(t), \tag{28}$$

where $h(s,t)$ is the blurring function and $n(t)$ is white noise. We derive the IBR linear filter to estimate $\boldsymbol{Y}(t)$ from $\boldsymbol{X}(t)$. The function class $\mathcal{F}$ is defined by (9) and the MSE is used as the cost function. Among the experiments that can exactly determine one of the uncertain parameters, we aim to predict the one minimizing the design value defined in (6).

## V. Computational Complexity Analysis

Here we analyze the complexity of optimal experimental design for SDE model-based filtering considered in this paper. Assume the dimensions of the signal vector $\boldsymbol{Y}$ and observation vector $\boldsymbol{X}$ are $N_y$ and $N_x$, respectively. Note that $N_y$ and $N_x$ are equal to the multiplication of the number of discrete time points for discrete approximation and the channel numbers of the corresponding signal and observation processes. In addition, we assume that we have $k$ uncertain parameters in the SDE system and therefore there are $k$ possible experiments to specify each parameter for our experimental design setup, which requires solving the optimization problem in (18) over $k$ parameters.

The objective function in (18) involves the computation of the expectation over $\bar{\theta}_i$, which can be calculated by Monte Carlo (MC) integration. Assume we sample $M_1$ samples of $\theta_i^{(j)}, j \le M_1$. Given each $\theta_i^{(j)}$, if we have closed-form effective correlation matrices in (18), we just need to calculate the matrix multiplication and the trace given effective correlation matrices inside the expectation. First, computing the pseudoinverse $[\boldsymbol{R}_{\Theta|\theta_i,X}]^+$ has cubic complexity $O((N_x)^3)$. The matrix multiplication to derive $A = \boldsymbol{R}_{\Theta|\bar{\theta}_i,YX}[\boldsymbol{R}_{\Theta|\bar{\theta}_i,X}]^+$ has complexity of $O(N_y(N_x)^2)$ and calculating the trace $tr(A\boldsymbol{R}^{\mathrm{T}}_{\Theta|\bar{\theta}_i,YX})$ has the complexity $O(N_xN_y)$. The complexity of the matrix calculations for each sample is $O(N_y(N_x)^2 + (N_x)^3)$, and therefore the complexity of optimal experimental design by solving (18) is $O(kM_1(N_y(N_x)^2 + (N_x)^3))$.

In practice, there is typically no closed-form solution to the underlying SDE system modeling the signal process, hence there is no closed-form expression for effective correlation matrices. In such cases, we would also need to estimate the effective correlation matrices in (18) by MC sampling in addition to the matrix calculations analyzed above. Assume we generate $M_2$ samples of $(\theta|\theta_i^{(j)}, \boldsymbol{Y}^{(j)}, \boldsymbol{X}^{(j)}), j \le M_2$, where $\boldsymbol{Y}^{(j)}$ can be generated by (26) and $\boldsymbol{X}^{(j)}$ by (28). Due to the Markovian property of (26), the complexities of sampling $\boldsymbol{Y}^{(j)}$ and $\boldsymbol{X}^{(j)}$ are all linear. The effective cross-correlation is estimated by:

$$\boldsymbol{R}_{\Theta|\bar{\theta}_i,YX} = \frac{1}{M_2}\sum_{j=1}^{M_2} \boldsymbol{Y}^{(j)}(\boldsymbol{X}^{(j)})^{\mathrm{T}}, \tag{29}$$

with the complexity $O(M_2N_yN_x)$. Similarly, the complexity of estimating the effective auto-correlation of $\boldsymbol{X}$ is $O(M_2N_x^2)$. With these, the complexity of optimal experimental design is

$$O(kM_1[N_y(N_x)^2 + (N_x)^3 + M_2(N_x^2 + N_yN_x)]). \tag{30}$$

Note that the MC integration procedure for effective correlation matrix estimation has to sample the uncertainty class of all $k$ parameters and $M_2$ can grow exponentially with $k$.

## VI. Synthetic Experiments

To demonstrate the performance of the proposed robust filtering and optimal experimental design methods, we first consider a synthetic example, which assumes that the original signal $\boldsymbol{Y}(t)$ is generated by an SDE of the form in (22). Assume $\boldsymbol{Y}(t)$ is a two-channel signal and the parameters of the corresponding SDE are given by

$$\begin{aligned}
\mathbf{A}(t) &= \frac{\theta_1}{100}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\
\mathbf{a}(t) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\
\mathbf{B}(t) &= 0.1\begin{pmatrix} 1 & \theta_2 \\ \theta_2 & 1 \end{pmatrix}, \\
\boldsymbol{Y}(0) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix},
\end{aligned} \tag{31}$$

where $\theta = (\theta_1, \theta_2)$ is the uncertain parameter vector, $\boldsymbol{Y}(t)$ is defined within the time interval $[0, T = 100]$.

$\boldsymbol{X}(t)$ is the observation of $\boldsymbol{Y}(t)$, which is corrupted by a blurring function $h(t)$ with additive noise $\boldsymbol{N}(t)$:

$$\boldsymbol{X}(t) = \int_0^T h(t-s)\boldsymbol{Y}(s)ds + \boldsymbol{N}(t), \tag{32}$$

where

$$h(t) = \frac{1}{B}(\mathrm{sgn}(t) - \mathrm{sgn}(t-B)), \tag{33}$$

with $B = 10$, is a scalar function, so the blurring effect is the same for both channels. The variances of additive noise for both channels are the same, $\sigma^2 = 0.01$.

As mentioned earlier, the SDE has a unique solution as a Gaussian process. Therefore, we can obtain a closed-form expression of correlations between $\boldsymbol{Y}(t)$ and $\boldsymbol{X}(t)$. Let's begin with the fundamental matrix of (25):

$$\Phi(t) = \begin{pmatrix} e^{\theta_1 t} & 0 \\ 0 & e^{\theta_1 t} \end{pmatrix}, \tag{34}$$

with $\mathrm{d}\Phi(t)/\mathrm{d}t = \boldsymbol{A}(t)\Phi(t)$. The auto-correlation of $\boldsymbol{Y}(t)$ can be calculated by (24):

$$\begin{aligned}
&\boldsymbol{R}_Y(t_i, t_j) \\
&= \Phi(t_i)\left(\int_{t_0}^{t_i} \Phi(u)^{-1}\mathrm{B}(u)\mathrm{B}(u)^{\mathrm{T}}(\Phi(u)^{-1})^{\mathrm{T}}\,\mathrm{d}u\right)\Phi(t_j)^{\mathrm{T}} \\
&= \frac{1}{2\theta_1}\{e^{\theta_1(t_i+t_j)} - e^{\theta_1(t_j-t_i)}\}\begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} \\
&= r_Y(t_i, t_j)\begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix},
\end{aligned} \tag{35}$$

where

$$r_Y(t_i, t_j) = \frac{1}{2\theta_1}\{e^{\theta_1(t_i+t_j)} - e^{\theta_1(t_j-t_i)}\}. \tag{36}$$

Equation (35) holds for $t_j \geq t_i \geq 0$, and we have $\boldsymbol{R}_Y(t_i, t_j) = \boldsymbol{R}_Y(t_j, t_i)$. Based on the observation model in (32), the cross-correlation is

$$\begin{aligned}
&\boldsymbol{R}_{YX}(t_i, t_j) \\
&\quad = \int_0^T h(t_j - s)\boldsymbol{R}_Y(t_i, s)\mathrm{d}s \\
&\quad = \int_0^T h(t_j - s)r_Y(t_i, s)\mathrm{d}s\begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} \\
&\quad = r_{YX}(t_i, t_j)\begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix},
\end{aligned} \tag{37}$$

with

$$r_{YX}(t_i, t_j) = \int_0^T h(t_j - s)r_Y(t_i, s)\mathrm{d}s. \tag{38}$$

The auto-correlation of $\boldsymbol{X}(t)$ is

$$\begin{aligned}
&\boldsymbol{R}_X(t_i, t_j) \\
&= \int_0^T\int_0^T h(t_i - s)\boldsymbol{R}_Y(s, u)h(t_j - u)\mathrm{d}s\mathrm{d}u + \sigma^2\delta(t_i - t_j)I_2 \\
&= \int_0^T\int_0^T h(t_i - s)r_Y(s, u)h(t_j - u)\mathrm{d}s\mathrm{d}u \cdot \\
&\quad \begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} + \sigma^2\delta(t_i - t_j)I_2 \\
&= r_X(t_i, t_j)\begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} + \sigma^2\delta(t_i - t_j)I_2,
\end{aligned} \tag{39}$$

with

$$r_X(t_i, t_j) = \int_0^T\int_0^T h(t_i - s)r_Y(s, u)h(t_j - u)\mathrm{d}s\mathrm{d}u. \tag{40}$$

The integrals in (38) and (40) can be calculated directly and have piecewise closed-form expressions depending on the value relationships between $t_i$, $t_j$, $B$ and $T$.

As noted previously, we consider the discrete filtering problem. The signals from the SDE system are sampled at discrete time points $t = 0, 1, 2, ..., 100$. We denote the flattened discrete time vectors of $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ as:

$$\boldsymbol{X}^N = (X_0^1, \ldots, X_N^1, X_1^2, \ldots, X_N^2)^{\mathrm{T}} \tag{41}$$

and

$$\boldsymbol{Y}^N = (Y_0^1, \ldots, Y_N^1, Y_1^2, \ldots, Y_N^2)^{\mathrm{T}}, \tag{42}$$

where $Y_j^i$ and $X_j^i$ indicate the signal and observation values at the $i$-th channel and time $j$, respectively, for $i = 1, 2$ and $0 \leq j \leq N$.

Then the matrix forms of correlations are

$$\boldsymbol{R}_{\theta,YX}^N = \begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} \otimes \boldsymbol{r}_{YX}^N, \tag{43}$$

$$\boldsymbol{R}_{\theta,X}^N = \begin{pmatrix} 1+\theta_2^2 & 2\theta_2 \\ 2\theta_2 & 1+\theta_2^2 \end{pmatrix} \otimes \boldsymbol{r}_X^N + \sigma^2 I_{2N}, \tag{44}$$

where $\otimes$ indicates the Kronecker product, and

$$\begin{aligned}
\boldsymbol{r}_{YX}^N &= \begin{pmatrix} r_{YX}(t_0, t_0) & \cdots & r_{YX}(t_0, t_N) \\ \vdots & \ddots & \vdots \\ r_{YX}(t_N, t_0) & \cdots & r_{YX}(t_N, t_N) \end{pmatrix}, \\
\boldsymbol{r}_X^N &= \begin{pmatrix} r_X(t_0, t_0) & \cdots & r_X(t_0, t_N) \\ \vdots & \ddots & \vdots \\ r_X(t_N, t_0) & \cdots & r_X(t_N, t_N) \end{pmatrix}
\end{aligned} \tag{45}$$

are corresponding matrix forms of $r_{YX}(t_i, t_j)$ and $r_X(t_i, t_j)$.

### A. IBR Filter Performance

To examine the performance of the IBR filter $\widehat{G}_\Theta = \boldsymbol{R}_{\Theta,YX}^N[\boldsymbol{R}_{\Theta,X}^N]^+$, fix $\theta_1 = 1$ and let $\theta_2$ be uniformly distributed over the interval $(-1, 1)$. We then have a closed-form expression for the expectation over $\theta$:

$$\boldsymbol{R}_{\Theta,YX}^N = \begin{pmatrix} 4/3 & 0 \\ 0 & 4/3 \end{pmatrix} \otimes \boldsymbol{r}_{YX}^N, \tag{46}$$

$$\boldsymbol{R}_{\Theta,X}^N = \begin{pmatrix} 4/3 & 0 \\ 0 & 4/3 \end{pmatrix} \otimes \boldsymbol{r}_X^N + \sigma^2 I_{2N}. \tag{47}$$

Note that the effective correlation doesn't correspond to any specific value of $\theta_2$. Inserting (46) and (47) to (14) yields the matrix-form IBR filter $\widehat{G}_\Theta$.

To show the performance of the IBR filter, we compare it with the optimal filter for $\theta_2 = 0.8$ and the recently developed $\tau$-robust filter which is robust with bounded $\tau$-divergence and is the optimal filter based on the nominal statistics with respect to $\theta_2 = 0$ [13]. The result for applying the three filters on the observation of the signal generated by the SDE with $\theta_2 = 0.8$ is shown in Fig. 2. Note that the IBR filter has a performance

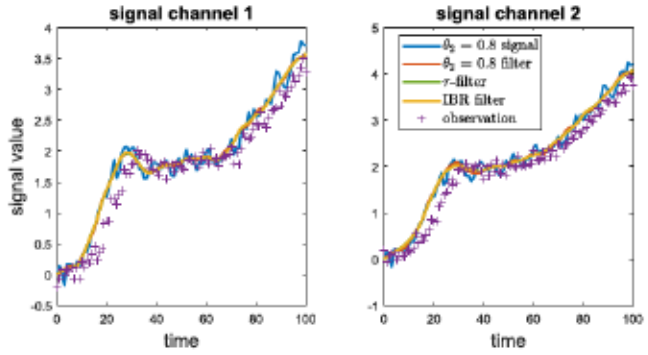



Fig. 2. One sample from the SDE (22), (31) with $\theta_2 = 0.8$. Left and right sub-figures show the signals from the first and the second channels, respectively. The original signals are in blue. The filtered signals based on the optimal filter for $\theta_2 = 0.8$ are in red. The filtered signals using the $\tau$-robust filter are in green. The IBR filtered signals are in yellow. The corrupted observations are in purple crosses.

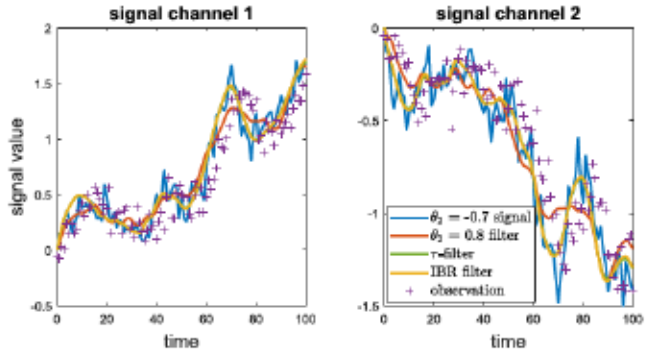



Fig. 3. One sample simulated from the SDE (22), (31) with $\theta_2 = -0.7$. Left and right sub-figures show the signals from the first and the second channels, respectively. The original signals are in blue. The filtered signals based on the optimal Wiener filter for $\theta_2 = 0.8$ are in red. The filtered signals using the $\tau$-robust filter are in green. The IBR filtered signals are in yellow. The corrupted observations are in purple crosses.

(Mean Square Error (MSE) = 2.6015) fairly close to the filter that is optimal for $\theta_2 = 0.8$ (MSE = 2.5269) and performs better than the $\tau$-robust filter (MSE = 2.6121).

Next we applied the same filters on the observation of the signal generated by the SDE with $\theta_2 = -0.7$, the result being shown in Fig. 3. Here the IBR filter still maintains relatively good performance (MSE = 2.2360), followed by the $\tau$-robust filter with MSE = 2.2416, but the filter optimal for $\theta_2 = 0.8$ shows a significantly degraded performance (MSE = 4.4331) due to the model mismatch.

### B. Optimal Experimental Design

The optimal experimental design problem is to determine which one of the two parameters, $\theta_1$ or $\theta_2$, should be specified first to minimize the cost due to uncertainty. Taking MSE for signal filtering as the cost, the cost function for experimental design is the residual IBR cost of two parameters, expressed as:

$$\mathrm{R}(\theta_1) = \mathbb{E}_{\theta_1}[\mathbb{E}_{\Theta|\theta_1}[C_{\Theta|\theta_1}(\widehat{G}_{\Theta|\theta_1})]], \tag{48}$$

$$\mathrm{R}(\theta_2) = \mathbb{E}_{\theta_2}[\mathbb{E}_{\Theta|\theta_2}[C_{\Theta|\theta_2}(\widehat{G}_{\Theta|\theta_2})]]. \tag{49}$$

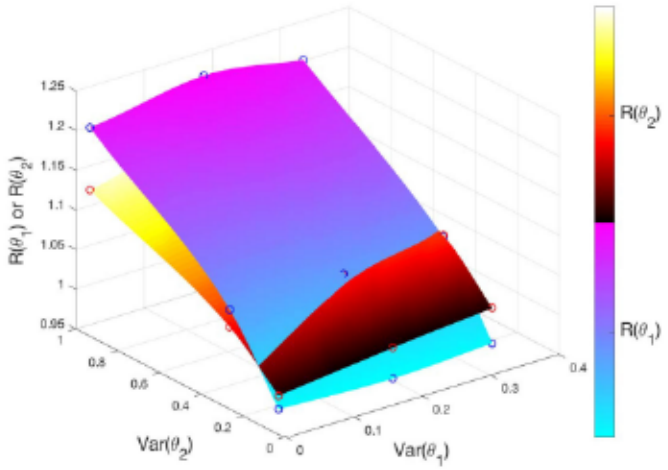



Fig. 4. Comparison of Residual IBR cost of $\theta_1$ and $\theta_2$. The parameter with less Residual IBR cost is suggested to be specified by next experiment. Red circles and blue circles are precise calculations of $R(\theta_1)$ and $R(\theta_2)$, respectively, and the surfaces are obtained by cubic spline interpolation.

We assume $\theta_1$ and $\theta_2$ are independent. $\theta_2$ is distributed over the interval $(-1, 1)$ as:

$$\theta_2 = 2\epsilon - 1, \quad \epsilon \sim Beta(\beta, \beta), \tag{50}$$

with $\beta$ the distribution parameter. $\theta_1$ is uniformly distributed as

$$\theta_1 \sim U(5 - L/2, \quad 5 + L/2), \tag{51}$$

with distribution parameter $L$.

In our simulations, we set three different values for $\beta = 0.5, 1.5, 5$ and $L = 0.5, 1.5, 2$, so that we have 9 combinations of distribution hyperparameters. The residual IBR cost is calculated by Monte Carlo sampling. For $\mathrm{R}(\theta_1)$, for each given pair of distribution parameters, 200 sample pairs of $\theta_1$ are taken for Monte Carlo computation, and for each $\theta_1$, the inner term has a closed-form expression as in (17):

$$\mathbb{E}_{\Theta|\theta_1}[C_{\Theta|\theta_1}(\widehat{G}_{\Theta|\theta_1})] = \mathbb{E}_{\theta_1}[\mathrm{tr}(\boldsymbol{R}_{\Theta|\theta_1,Y} - \boldsymbol{R}_{\Theta|\theta_1,YX}[\boldsymbol{R}_{\Theta|\theta_1,X}]^{+}\boldsymbol{R}'_{\Theta|\theta_1,YX})].$$

$\mathrm{R}(\theta_2)$ is calculated similarly by Monte Carlo sampling. We just need to substitute $\theta_1$ with $\theta_2$ in the above expression to calculate $\mathbb{E}_{\Theta|\theta_2}[C_{\Theta|\theta_2}(\widehat{G}_{\Theta|\theta_2})]$. The residual IBR costs are shown in Fig. 4. The variances of the two parameters are $\mathrm{Var}(\theta_2) = \frac{1}{2\beta+1}$ and $\mathrm{Var}(\theta_1) = \frac{L^2}{12}$.

From the figure we can see how the variances of the uncertain parameters influence the IBR residuals. The variance of $\theta_2$ has a higher influence on the IBR residuals than the variance of $\theta_1$ does. As $\mathrm{Var}(\theta_2)$ increases, both IBR residuals increase as a larger variance introduces more uncertainty in the model. But when $\mathrm{Var}(\theta_2)$ is large, $\mathrm{R}(\theta_2)$ is smaller than $\mathrm{R}(\theta_1)$, because estimating $\theta_2$ can reduce the uncertainty (or the cost thereof) more than estimating $\theta_1$, pertaining to the filtering performance in this case. For small $\mathrm{Var}(\theta_2)$, we have the opposite conclusion.

To further illustrate the strength of the MOCU-based experimental design, here we perform experiments with a more complicated uncertainty class and compare its performance with

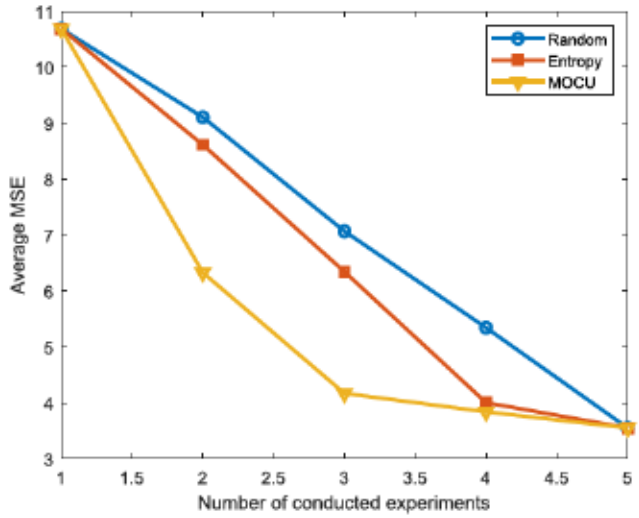



Fig. 5. The average performance of sequential experimental design with different strategies. In each setup, the MSE is obtained after conducting each experiment in a sequence of experiments for the SDE signal and observation model with four unknown parameters.

both entropy-based [38] and random sequential experimental design.

Assume that we have four uncertain parameters, $\theta = (\theta_1, \theta_2, B, \sigma)$ in the model described by (31)–(33). These uncertain model parameters follow independent uniform distributions. For each parameter, we assume an experiment can be performed to obtain its value. In addition we assume all the parameter measurements have Gaussian errors. We perform a sequential experimental design to decide which model parameter to measure in each iteration so that we can most effectively improve the filtering performance within a relatively small number of iterations. For this experimental design problem, we compare the MOCU-based strategy described by (18) with both entropy-based strategy and random strategy. The entropy-based strategy chooses the experiment to measure the parameter with the largest Shannon entropy; and the random strategy simply chooses one out of the uncertain parameters in a random fashion. To compare the different strategies in different cases, we set three different groups of parameter distributions for sequential experimental design: (1) $\theta_1 \sim U(3,6), \theta_2 \sim U(-2,2)$, $B \sim U(8,10.5)$, $\sigma \sim U(0.01,1.2)$; (2) $\theta_1 \sim U(3,6)$, $\theta_2 \sim U(-1.4,1.4)$, $B \sim U(8,10.5)$, $\sigma \sim U(0.01,1.2)$; (3) $\theta_1 \sim U(3.7,6)$, $\theta_2 \sim U(-1,1)$, $B \sim U(8,10.5)$, $\sigma \sim U(0.01,2)$. For all the cases, the parameters have Gaussian measurement error with a variance $\sigma_\epsilon^2 = 0.05$. In each cases, we randomly generate 100 groups of parameters, and perform sequential experimental design following the three strategies. After each iteration, we calculate the remaining MSE of the corresponding IBR filter to quantify the remaining uncertainty. Fig. 5 provides the change of the average MSE with the number of experiment iterations for these three experimental design strategies. As expected, our MOCU-based strategy consistently identifies the most critical uncertain parameter, whose measurement leads to the maximum reduction of the MSE with our filtering objective in design. As a result, after two experiments, when two parameters have been determined, the performance of our MOCU-based strategy has almost reached the level obtained when there is no uncertainty remaining (after four experiments), whereas for both entropy and random design there remains significant uncertainty after two experiments, meaning that they have not identified the best two parameters to estimate.

## VII. Pharmacokinetics Model

In this section, we illustrate the IBR filter and experimental design for a pharmacokinetic two-compartment model based on a SDE system [39]. Differentiating the body into a central compartment (plasma) and a peripheral compartment (tissues), the two-compartment model describes the relationship between the drug concentration in the central compartment $Y_1(t)$, the drug concentration in the peripheral compartment $Y_2(t)$, and the measurement $x_t$ of the drug concentration in the central compartment. The transit of the drug throughout the body is described by the SDE shown below:

$$\begin{aligned} \mathrm{d}Y_1(t) &= (k_{21}Y_2(t) - k_{12}Y_1(t) - k_{10}Y_1(t))\,\mathrm{d}t + \sigma_1 \mathrm{d}\,W_1(t), \\ \mathrm{d}Y_2(t) &= (k_{12}Y_1(t) - k_{21}Y_2(t))\,\mathrm{d}t + \sigma_2\,\mathrm{d}\,W_2(t), \\ X(t) &= Y_1(t) + \epsilon, \qquad \epsilon \sim N(0, \sigma_\epsilon^2), \end{aligned} \tag{52}$$

where $W_1(t)$, $W_2(t)$ are independent Wiener processes, and $k_{10}$, $k_{12}$, and $k_{21}$ are individual rate constants (parameters) possessing the joint prior distribution

$$\theta = (k_{10},\, k_{12},\, k_{21})^{\mathrm{T}} \sim N(\mu, \Omega).$$

Following the case example in [32], we set the statistics of the prior as: $\mu = (0.2, 0.5, 0.25)^{\mathrm{T}}$, and $\Omega$ a diagonal matrix with $diag(\Omega) = (0.01^2, 0.1^2, 0.02^2)^{\mathrm{T}}$. Other parameters are set to $\sigma_\epsilon^2 = 0.04$, $\sigma_1 = 0.1$, $\sigma_2 = 0.1$. The initial condition of $Y_1(t)$ and $Y_2(t)$ are set to be 10 and 0, respectively, which corresponds to the case of Intravenous injection: the pharmacy is initially injected to the plasma and then diffuse to the tissue. After finding the IBR filter using the preceding theory, we consider its performance, and then turn to the problem of specifying in what order to determine the individual rate constants to optimally reduce the MSE of estimating $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t))^{\mathrm{T}}$. We consider the discrete case with sampling points from 0 to 10 by an increment of 0.01.

The SDE in (52) also follows the form of (22), with the matrices below:

$$\begin{aligned} \mathbf{A}(t) &= \begin{pmatrix} -k_{12} - k_{10} & k_{21} \\ k_{12} & -k_{21} \end{pmatrix}, \\ \mathbf{a}(t) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \mathbf{B}(t) &= \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \\ \boldsymbol{Y}(0) &= \begin{pmatrix} 10 \\ 0 \end{pmatrix}. \end{aligned} \tag{53}$$

Therefore, similar to the procedure in the synthetic example, we can calculate the IBR filter $\widehat{G}_\Theta$ through numerical integrals.

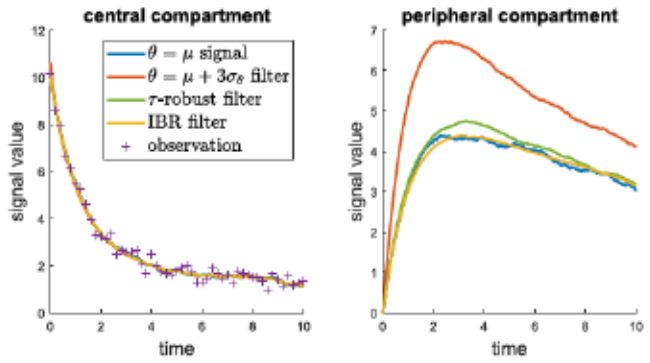



Fig. 6. One example based on the SDE with $\theta = \mu$. Left and right sub-figures show the drug concentration levels of central and peripheral compartments, respectively. Blue curves correspond to the actual signals; purple crosses indicate the measurements of concentration in the central compartment; red curves depict the estimation with the optimal filter for $\theta = \mu + 3\sigma_\theta$; green curves are the estimated signals with the $\tau$-robust filter; and yellow curves are filtered signals using the IBR filter. Only one out of every 20 measurements is visualized here to avoid curve cluttering.

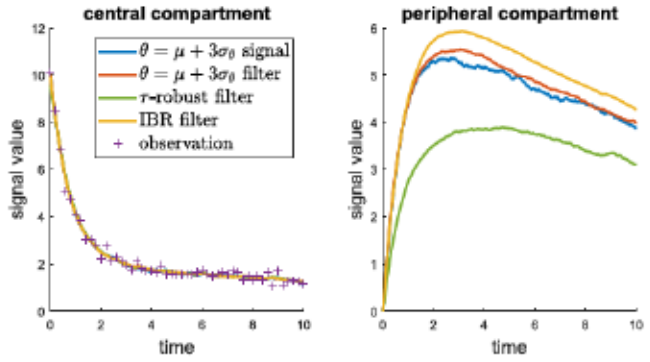



Fig. 7. One example based on the SDE with $\theta = \mu + 3\sigma_\theta$. Left and right sub-figures show the drug concentration levels of central and peripheral compartments, respectively. Blue curves correspond to the actual signals; purple crosses indicate the measurements of concentration in the central compartment; red curves depict the estimation with the optimal filter for $\theta = \mu + 3\sigma_\theta$; green curves are the estimated signals with the $\tau$-robust filter; and yellow curves are filtered signals using the IBR filter. Only one out of every 20 measurements is visualized here to avoid curve cluttering.

We compare the performance of the IBR filter with two other filters: the Wiener filters given specific values of the parameters $\theta = \mu + 3\sigma_\theta$ with $\sigma_\theta = (0.01, 0.1, 0.02)^{\mathrm{T}}$ the vector of standard deviations of parameters and the tau-robust filter. The comparison of the filtering performance on signals generated with $\theta = \mu$ and $\theta = \mu + 3\sigma_\theta$ are shown in Figs. 6 and 7, respectively. Since there is no direct observation in the peripheral compartment, the main source of the estimation error is in the peripheral compartment. Observation from the peripheral compartment shows that the IBR filter performs fairly well in both cases, while the Wiener filter with $\theta = \mu + 3\sigma_\theta$ performs well only in the case with matched parameters as expected. We notice that the $\tau$-robust filter does not show its robustness on signals generated with $\theta = \mu + 3\sigma_\theta$, probably because in the setting of this pharmacokinetics model, the uncertain parameters follow unbounded Gaussian distributions, while the $\tau$-robust filter is proposed under the bounded $\tau$-divergence assumption.

Then we perform sequential experimental design by calculating the design values of parameters $k_{10}, k_{12}$ and $k_{21}$. Suppose our experimental budget can afford to perform two experiments to determine two of the unknown parameters, and we want to sequentially minimize the MOCU with the remaining uncertainty. For the first experiment, the IBR residual costs are $R^1(k_{10}) = 51.3$, $R^1(k_{12}) = 45.2$ and $R^1(k_{21}) = 36.4$. So the first experiment should determine $k_{21}$. Following the first experiment, the true value of $k_{21}$ is put into the model and the design values are calculated based on the updated model. We randomly sample 10 values of $k_{21}$ as the result of the first experiment, and then calculate the IBR residual costs $R^2(k_{10}|k_{21})$ and $R^2(k_{12}|k_{21})$. All 10 random cases show that $k_{12}$ should be determined in the second experiment, and the average design values are $\mathbb{E}_{k_{21}}[D^2(k_{10}|k_{21})] = 33.4$ and $\mathbb{E}_{k_{21}}[D^2(k_{12}|k_{21})] = 14.2$. Although the choice of the second experiment in our example is the same for all sampled values of the primary parameter, in general, the choice of the second experiment depends upon the value of the primary parameter, so that the choice of the second experiment can vary depending on the determined value of the primary parameter. In this example, the estimation error of $Y_2(t)$ dominates the full estimation cost, since $Y_2(t)$ is not observed directly, and the estimation of $Y_2(t)$ is based on its correlation with $Y_1(t)$. Therefore, $k_{10}$ is less important than the other parameters, since it is conditionally independent of $Y_2(t)$ given $Y_1(t)$. Our calculation confirms this observation, preferring the estimation of $k_{12}$ over that of $k_{10}$.

## VIII. Conclusion

The mathematics of the IBR/MOCU paradigm depends on the physical model, the operational objective, the cost function, and the operator form. Applied to classification, uncertainty resides in the feature-label distributions; applied to Markov chains, it resides in the transition probability matrix; and applied to differential or stochastic differential equations, it resides in the physical constants of the equations, or perhaps even in the differential operators themselves.

In fact, this is an oversimplification. In the case of classification, it could be that the signals being classified are generated via an SDE, and we wish to classify a signal based upon a sampling of its time trajectory. This is precisely the situation in a study concerning optimal Bayesian classification of signals generated by an uncertain SDE model [40], although that study did not consider the experimental design problem. We raise this point to emphasize that uncertainty is propagated through the model and can manifest itself in the characteristics used in operator design. This means that, in Gaussian classification, the uncertainty class for the parameters induces an uncertainty class of mean vectors and covariance matrices, and in linear filtering, the parameter uncertainty class induces an uncertainty class of random-signal processes.

A subtle, but fundamental point arises concerning the distribution of the uncertainty vector. In the present study, we have propagated the prior distribution through the SDE so that the same distribution governs the uncertainty of signal processes. In the former classification study [40], no assumption was made on

the distribution of the parameter uncertainty class and it was then assumed that a normal-inverse-Wishart distribution governed the mean and covariance matrix of the uncertain Gaussian features constructed by sampling the signal trajectories. This was convenient because it allowed direct application of the theory of optimal Bayesian classification for Gaussian features [41], thereby resulting in a closed-form solution for the optimal Bayesian classifier. The convenience of the previous assumption comes at a significant price: if there is physical knowledge regarding the distribution of the uncertain parameters, it has been ignored. Thus, we believe that uncertainty propagation, as used in the present paper is more sound from a physical perspective, even if it leaves us with no hope of a convenient closed-form solution.

Although computational complexity did not impede us in the present paper, it can become a problem when there is high dimensionality, especially when the uncertainty class is large. Model reduction can be used to reduce the computations. For instance, a regulatory network model can be compressed by eliminating or combining nodes [42]. Model reduction remains an important research topic from a practical perspective, and to be effective, such reduction should be made in a way that maintains the structure most relevant to the objective – which makes it application dependent.

Finally, we note that MOCU-based experimental design can be generalized [43] in such a way that the original MOCU [4], as used herein, as well as both the knowledge gradient (KG) [44] and efficient global optimization (EGO) [45], are special cases.

## REFERENCES


[1] T. Kailath, "Lectures on Wiener and Kalman filtering," in *Lectures on Wiener and Kalman Filtering*. Berlin, Germany: Springer-Verlag, 1981, pp. 1–143.

[2] V. N. Fomin, *Optimal Filtering: Volume I: Filtering of Stochastic Processes*, vol. 457. Berlin, Germany: Springer-Verlag, 2012, pp. 111–119.

[3] L. A. Dalton and E. R. Dougherty, "Intrinsically optimal Bayesian robust filtering," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 657–670, Feb. 2014.

[4] B.-J. Yoon, X. Qian, and E. R. Dougherty, "Quantifying the objective cost of uncertainty in complex dynamical systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2256–2266, May 2013.

[5] R. Dehghannasiri, X. Qian, and E. R. Dougherty, "Optimal experimental design in the context of canonical expansions," *IET Signal Process.*, vol. 11, no. 8, pp. 942–951, 2017.

[6] X. Qian and E. R. Dougherty, "Bayesian regression with network prior: Optimal Bayesian filtering perspective," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6243–6253, Dec. 2016.

[7] E. R. Dougherty, *Optimal Signal Processing Under Uncertainty*. Bellingham, WA, USA: SPIE, 2018.

[8] V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecomm. Radio Eng.*, vol. 30, no. 3, pp. 58–64, 1976.

[9] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, no. 4-5, pp. 171–185, 1977.

[10] H. Poor, "On robust Wiener filtering," *IEEE Trans. Autom. Control*, vol. 25, no. 3, pp. 531–536, Jun. 1980.

[11] K. Vastola and H. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 316–327, Mar. 1984.

[12] A. M. Grigoryan and E. R. Dougherty, "Design and analysis of robust binary filters in the context of a prior distribution for the states of nature," *J. Math. Imag. Vis.*, vol. 11, no. 3, pp. 239–254, 1999.

[13] M. Zorzi, "On the robustness of the Bayes and Wiener estimators under model uncertainty," *Automatica*, vol. 83, pp. 133–140, 2017.

[14] Y. C. Eldar and N. Merhav, "Minimax MSE-ratio estimation with signal covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1335–1347, Apr. 2005.

[15] Y. C. Eldar, "Minimax estimation of deterministic parameters in linear models with a random model matrix," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 601–612, Feb. 2006.

[16] A. Sani and A. Vosoughi, "On distributed linear estimation with observation model uncertainties," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3212–3227, Jun. 2018.

[17] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Autom. Control*, vol. 15, no. 2, pp. 175–184, Apr. 1970.

[18] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 596–600, Mar. 2009.

[19] O. K. Kwon, W. H. Kwon, and K. S. Lee, "FIR filters and recursive forms for discrete-time state-space models," *Automatica*, vol. 25, no. 5, pp. 715–728, 1989.

[20] Y. S. Shmaliy, "Linear optimal FIR estimation of discrete time-invariant state-space models," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3086–3096, Jun. 2010.

[21] A. H. Sayed, "A framework for state-space estimation with uncertain models," *IEEE Trans. Automa. Control*, vol. 46, no. 7, pp. 998–1013, Jun. 2001.

[22] T. Zhou, "Sensitivity penalization based robust state estimation for uncertain linear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 1018–1024, Apr. 2010.

[23] T. Zhou, "Robust recursive state estimation with random measurement droppings," *IEEE Trans. Autom. Control*, vol. 61, no. 1, pp. 156–171, Jan. 2016.

[24] R. Dehghannasiri, M. S. Esfahani, and E. R. Dougherty, "Intrinsically Bayesian robust Kalman filter: An innovation process approach," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2531–2546, May 2017.

[25] R. Dehghannasiri, X. Qian, and E. Dougherty, "Bayesian robust Kalman smoother in the presence of unknown noise statistics," *EURASIP J. Adv. Signal Process.*, vol. 55, pp. 1982–1996, 2018.

[26] R. Dehghannasiri, B.-J. Yoon, and E. R. Dougherty, "Optimal experimental design for gene regulatory networks in the presence of uncertainty," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 4, pp. 938–950, Jul./Aug. 2015.

[27] D. N. Mohsenizadeh, R. Dehghannasiri, and E. R. Dougherty, "Optimal objective-based experimental design for uncertain dynamical gene networks with experimental error," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 218–230, Jan./Feb. 2018.

[28] H. Kirchauer, F. Hlawatsch, and W. Kozek, "Time-frequency formulation and design of nonstationary Wiener filters," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1995, vol. 3, pp. 1549–1552.

[29] Y. Sugiyama, "An algorithm for solving discrete-time Wiener-Hopf equations based upon Euclid's algorithm," *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 394–409, May 1986.

[30] M. Morii, M. Kasahara, and D. L. Whiting, "Efficient bit-serial multiplication and the discrete-time Wiener-Hopf equation over finite fields," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1177–1183, Nov. 1989.

[31] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[32] F.-R. Yan *et al.*, "Parameter estimation of population pharmacokinetic models with stochastic differential equations: Implementation of an estimation algorithm," *J. Probability Statist.*, vol. 2014, pp. 4–5, 2014.

[33] C. A. Braumann, "Itô versus Stratonovich calculus in random population growth," *Math. Biosciences*, vol. 206, no. 1, pp. 81–107, 2007.

[34] A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan, "A stochastic differential equation SIS epidemic model," *SIAM J. Appl. Math.*, vol. 71, no. 3, pp. 876–902, 2011.

[35] P. Malliavin and A. Thalmaier, *Stochastic Calculus of Variations in Mathematical Finance*. Berlin, Germany: Springer-Verlag, 2006.

[36] L. Arnold and A. V. Balakrishnan, "Stochastic differential equations theory and applications," *Bull. Amer. Math. Soc*, vol. 81, pp. 837–840, 1975.

[37] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, vol. 23. Berlin, Germany: Springer-Verlag, 2013.

[38] D. Coles and M. Prange, "Toward efficient computation of the expected relative entropy for nonlinear experimental design," *Inverse Problems*, vol. 28, no. 5, 2012, Art. no. 055019.

[39] L. Shargel, B. Andrew, and S. Wu-Pong, *Applied Biopharmaceutics and Pharmacokinetics*. New York, NY, USA: Appleton & Lange Stamford, 1999.

[40] A. Zollanvari and E. R. Dougherty, "Incorporating prior knowledge induced from stochastic differential equations in the classification of stochastic observations," *EURASIP J. Bioinf. Syst. Biol.*, vol. 2016, no. 1, pp. 1–14, 2016.

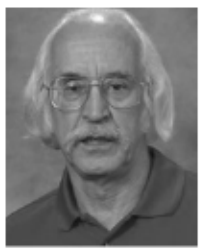
[41] L. A. Dalton and E. R. Dougherty, "Optimal classifiers with minimum expected error within a Bayesian framework-Part I: Discrete and Gaussian models," *Pattern Recognit.*, vol. 46, no. 5, pp. 1301–1314, 2013.
[42] R. Dehghannasiri, B.-J. Yoon, and E. R. Dougherty, "Efficient experimental design for uncertainty reduction in gene regulatory networks," *BMC Bioinf.*, vol. 16, no. 13, pp. 1–18, 2015.
[43] S. Boluki, X. Qian, and E. R. Dougherty, "Experimental design via generalized mean objective cost of uncertainty," *IEEE Access*, vol. 7, pp. 2223–2230, 2018.
[44] P. I. Frazier, W. B. Powell, and S. Dayanik, "A knowledge-gradient policy for sequential information collection," *SIAM J. Control Optim.*, vol. 47, no. 5, pp. 2410–2439, 2008.
[45] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

**Guang Zhao** received the B.S. degree in mechanical engineering from the Harbin Institute of Technology, Harbin, China, in 2011, and the M.S. degree in electrical engineering from Peking University, Beijing, China in 2015. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His main research interests are Bayesian optimization and Bayesian experimental design.

**Xiaoning Qian** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Yale University, New Haven, CT, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. He is affiliated with the TEES-AgriLife Center for Bioinformatics & Genomic Systems Engineering and the Center for Translational Environmental Health Research at Texas A&M University. His recent honors include the National Science Foundation CAREER Award, the Texas A&M Engineering Experiment Station (TEES) Faculty Fellow, and the Montague-Center for Teaching Excellence Scholar at Texas A&M University. His research interests include machine learning and Bayesian computation and their applications in materials science, computational network biology, genomic signal processing, and biomedical signal and image analysis.

**Byung-Jun Yoon** (Senior Member, IEEE) received the B.S.E. (*summa cum laude*) degree from Seoul National University, Seoul, Korea, in 1998, and the M.S. and Ph.D. degrees from the California Institute of Technology, Pasadena, in 2002 and 2007, respectively, all in electrical engineering. He joined the Department of Electrical and Computer Engineering, Texas A&M University, College Station, in 2008, where he is currently an Associate Professor. Since 2019, he holds a joint appointment at Brookhaven National Laboratory, Upton, NY, where he is a Scientist in Computational Science Initiative. His recent honors include the National Science Foundation (NSF) CAREER Award, the Best Paper Award at the 9th Asia Pacific Bioinformatics Conference (APBC), the Best Paper Award at the 12th Annual MCBIOS Conference, and the SLATE Teaching Excellence Award from the Texas A&M University System. His main research interests are in signal processing, machine learning, computational network biology, and bioinformatics.

**Francis J. Alexander** received Ph.D. degree in Physics from Rutgers University and a B.S. degree in physics and mathematics from The Ohio State University. He is the Deputy Director of the Computational Science Initiative at the U.S. Department of Energy (DOE)'s Brookhaven National Laboratory. Prior to that Alexander was at Los Alamos National Laboratory, where he was the acting Division Leader of the Computer, Computational, and Statistical Sciences (CCS) Division. During his more than 20 years at Los Alamos, he held several leadership roles, including as Deputy Leader of the CCS Divisions Information Sciences Group and leader of the Information Science and Technology Institute. Alexander first joined Los Alamos in 1991 as a Postdoctoral Researcher at the Center for Nonlinear Studies. He returned to Los Alamos in 1998 after doing postdoctoral work at the Institute for Scientific Computing Research at DOEs Lawrence Livermore National Laboratory and serving as a Research Assistant Professor at Boston Universitys Center for Computational Science. His current research interests include optimal design of experiments, computational physics, and nonequilibrium statistical mechanics. He has led many research projects and has published more than 50 papers in peer reviewed journals.



**Edward R. Dougherty** (Fellow, IEEE) received Ph.D. degree in mathematics from Rutgers University and an M.S. degree in Computer Science from Stevens Institute of Technology, and has been awarded the Doctor Honoris Causa by the Tampere University of Technology. He is a Distinguished Professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station, TX, where he holds the Robert M. Kennedy '26 Chair in Electrical Engineering. He is a fellow of both IEEE and SPIE, has received the SPIE President's Award, and served as the editor of the SPIE/IS&T Journal of Electronic Imaging. Prof. Dougherty is author of 19 books and author of more than 350 journal papers.