Journal of Process Control 23 (2013) 613–629

Contents lists available at SciVerse ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Optimal experiment design under process noise using Riccati differential equations

D. Telen[a], B. Houska[b], F. Logist[a], E. Van Derlinden[a], M. Diehl[c], J. Van Impe[a,*]

[a] BioTeC & OPTEC, Chemical Engineering Department, Katholieke Universiteit Leuven, W. de Croylaan 46, 3001 Leuven, Belgium
[b] Centre for Process Systems Engineering & Department of Chemical Engineering, Imperial College London, Roderic Hill Building, Room C509, South Kensington Campus, London SW7 2AZ, United Kingdom
[c] SCD & OPTEC, Electrical Engineering Department, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 20 March 2012
Received in revised form
19 November 2012
Accepted 19 November 2012

Keywords:
Optimal experiment design
Process noise
Variance–covariance matrix
Biochemical processes
Dynamic optimization

ABSTRACT

In this paper, we present a numerical method for optimal experiment design of nonlinear dynamic processes. Here, we suggest to optimize an approximation of the predicted variance–covariance matrix of the parameter estimates, which can be computed as the solution of a Riccati differential equation. In contrast to existing approaches, the proposed method allows us to take process noise into account and requires less derivative states to be computed compared to the traditional Fisher information matrix based approach. This process noise is assumed to be a time-varying random disturbance which is not known at the time when the experiment is designed. We illustrate the technique by solving an optimal experiment design problem for a fed-batch bioreactor benchmark case study. Here, we concentrate on how the optimal input design and associated accuracy of the parameter identification is influenced when process noise is present.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Nonlinear differential equation models are valuable tools for the analysis, design, control and optimization of dynamic processes. After a relevant model structure has been chosen, the model has to be calibrated before it can be used in practice as, e.g., discussed in [43]. Hence, the model parameters have to be estimated as accurately as possible in order to match the model predictions with the experimental data. The problem, however, is that performing experiments can be cost and/or labor intensive, e.g., manpower has to be diverted, samples have to be taken and subsequently analyzed. Optimal design of experiments for parameter estimation can help to systematically develop a limited set of experiments which will contain the highest amount of information and from which the model parameters can be estimated with the highest accuracy.

The research on optimal experiment design has been initialized by the pioneering work of Fisher [14]. Originally, experiment design techniques have been developed for static linear and nonlinear models as investigated by Kiefer and Wolfowitz [22] as well as Box and Lucas [10]. Extensions to linear dynamic systems have been proposed in the work of Mehra [29] and Gevers and Ljung [16]. Moreover, for an in depth discussion on recent developments in this field, we refer to the developments of Hjalmarsson [18] and Rojas et al. [34] and the references therein. With respect to nonlinear dynamic systems, the work by Espie and Machietto [13] is reported to be the first. A recommendable textbook on optimal experiment design has been written by Pukelsheim [33]. Finally, a more recent overview of developments in this field can be found in the work of Franceschini and Macchietto [15].

Optimal experiment design problems for nonlinear dynamic models require the selection of time-varying inputs which maximize the information content of an experiment. Typically, a scalar function of the Fisher information matrix is used to quantify the information content. As a result these problems lead to a specific and challenging class of dynamic optimization or optimal control problems. Nowadays, optimal control problems are mostly solved by direct methods in which the original infinite dimensional problem is reformulated to a finite dimensional nonlinear program via discretization of the controls and/or states. Several direct approaches exist, e.g., single shooting [35], multiple shooting [8,25], and orthogonal collocation [5,6]. An important work that addresses the specific numerical aspects of optimal experiment design for nonlinear differential equation systems has been written by Bauer et al. [4]. Extensions with respect to robust

* Corresponding author.
E-mail address: jan.vanimpe@cit.kuleuven.be (J. Van Impe).

0959-1524/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jprocont.2012.11.005

optimal experiment design and efficient multiple shooting implementations have been developed by Körkel and co-workers in [23] and [24], respectively.

Most of the approaches for optimal experiments suggest to compute the Fisher information matrix as, e.g., proposed in [15]. However, in the current paper, we propose an alternative approach which is based on a direct computation of the approximate variance–covariance matrix using Riccati differential equations. This approach can be interpreted as a continuous extension of the method proposed in [4]. Moreover, in the current paper, it is shown that this approximate variance–covariance matrix is related to the inverse of the classic Fisher information matrix. The current Riccati formulation has several advantages. The most important one, and, hence, one of the main contribution of the current paper, is that we can take process noise into account. Here, process noise is considered to be an unmodeled random variable affecting the dynamics of the system. In addition, due to the proposed continuous formulation, these Riccati differential equations can easily be integrated in existing direct optimal control solvers such as the implementation which has been developed in [19]. Furthermore, the Riccati differential equation based formulation exhibits a particular structure which can be efficiently exploited to speed up computations.

The second main contribution of the current paper is that we apply the proposed technique to the design of an optimal time-varying feed profile for a benchmark fed-batch bioreactor case-study which has originally been proposed in [3,37]. Similar reactor models have been considered in [30,31,1,12,17] as well as [36].

The current paper is structured as follows. In Section 2 the problem of optimal experiment design is formulated mathematically. We first discuss the classic computation of the Fisher information matrix. Additionally, we discuss the problems which arise when including process noise in the formalation of optimal experiment design problems. To conclude this section we discuss a selection of different design criteria for optimal experiment design. In Section 3 we compare different methods for computing the variance–covariance matrix with a Riccati differential equation based technique. Additionally, we discuss the computation of the derivatives of the variance–covariance matrix. To conclude this section, the formulation of optimal experiment design problems using the method is explained. We illustrate the Riccati differential equation based method in Section 4 for the dynamic fed-batch bioreactor benchmark model. Finally, we conclude the paper in Section 5.

2. Mathematical problem formulation

Let us consider the following uncertain dynamic system:

$$\forall t \in [0,T]: \quad \dot{y}(t) = g(y(t), p, u(t), w(t)) \quad \text{with} \quad y(0) = y_0. \tag{1}$$

Here, $y(t) \in \mathbb{R}^{n_y}$ denotes the state, $p \in \mathbb{R}^p$ an unknown but time-invariant parameter, $u(t) \in \mathbb{R}^{n_u}$ a control input which we want to optimize, and $w(t) \in \mathbb{R}^{n_w}$ an unknown and time-varying input. All these variables enter the right-hand side function g in a possibly nonlinear way. Note that for the theoretical considerations in this paper, we will use an alternative notation which stacks the parameters to the states. More precisely, we define $x(t) := (y(t)^\top, p^\top)^\top$ such that

$$\dot{x}(t) = f(x(t), u(t), w(t)) := (g(y(t), p, u(t), w(t))^\top, 0^\top)^\top \quad \text{with} \quad x(0) = (y_0^\top, p^\top)^\top. \tag{2}$$

Most of the traditional formulations of optimal experiment design analyze how accurately the unknown parameter p can be measured with the experiment at hand. However, similar to [4], we suggest in the current paper a more general framework which allows optionally analyzing not only the joint information about the parameter p but also the joint information about the state $y(t)$ and parameter p. This motivates us to collect these variables in one vector $x(t)$ aiming at a more compact notation. Thus, the unknowns which we want to estimate are now the initial value x_0 and the time varying input w. Finally, we plan to take measurements of the form $\eta(t) := z(t) + v(t) \in \mathbb{R}^{n_z}$, where the output relation $z(t) = H(x(t))$ may in general be nonlinear, too. Here, $v(t)$ denotes the measurement error.

In the following, we are interested in the case that the uncertainty functions w and v are stochastic white noise signals which are bounded almost surely and for which positive semi-definite matrices $V(t)$ and $W(t)$ are given such that

$$\begin{aligned} &\mathbb{E}\{w(t)\} = 0, \quad \mathbb{E}\left\{w(t)w(t')\right\} = W(t)\delta(t-t'), \\ &\mathbb{E}\{v(t)\} = 0, \quad \mathbb{E}\left\{v(t)v(t')\right\} = V(t)\delta(t-t') \end{aligned} \tag{3}$$

for all $t, t' \in \mathbb{R}$. The initial value $x(0) \in \mathcal{N}(\eta_0, Q_0)$ is assumed to be a random variable with Gaussian distribution, given variance–covariance matrix $Q_0 \in \mathbb{R}^{n_x \times n_x}$, and given expectation $\eta_0 \in \mathbb{R}^{n_x}$.

Let us assume for a moment that we have already taken measurements $\eta(t)$ for $t \in [0,T]$ using a given control input u. In this situation, we are interested in solving a nonlinear parameter estimation problem in order to compute the maximum likelihood estimate of the state. As we consider nonlinear dynamic systems, we do not have an explicit expression for this maximum likelihood estimate. Consequently, we have to compute the state estimate by solving a nonlinear optimal control problem of the form

$$(x^*, w^*) := \underset{x(\cdot), w(\cdot)}{\operatorname{argmin}} \quad J[x(\cdot), w(\cdot)] \quad \text{s.t.} \quad \dot{x}(t) = f(x(t), u(t), w(t))$$

with $t \in [0,T]$. Here, the objective J is a least-squares term

$$J[x(\cdot), w(\cdot)] := \left\|x(0) - \eta_0\right\|^2_{Q_0^{-1}} + \int_0^T \left\|w(t)\right\|^2_{W(t)^{-1}} \mathrm{d}t + \int_0^T \left\|H(x(t)) - \eta(t)\right\|^2_{V(t)^{-1}} \mathrm{d}t,$$

while (x^*, w^*) denotes the maximum likehood estimate for the state and noise, respectively. In this context and also in the following sections we use the notation

$$\|x\|^2_Y := x^\top Y x$$

in order to denote weighted Euclidean norms for positive definite weighting matrices Y. In the field of optimal experiment design we are interested in computing and optimizing the variance–covariance matrix of the maximum likelihood estimate $x^*(T)$ for the final state without knowing the measurements η yet. As this in general very difficult, we rely in this paper on linear approximation techniques which are briefly reviewed in the following section.

2.1. Variance–covariance approximation without process noise

One of the most broadly used formulations of optimal experiment design problems for parameter estimation [15] is based on the (approximate) Fisher information matrix $F_p[t, u]$ of the experiment. In the case, $F_p[t, u]$ is a functional of the control input u. For the case that we do not have any disturbances, i.e., for $w(t) = 0$ and $W(t) = 0$, and if the initial values y_0 for the states are known, it is defined as

$$F_p[t, u] := Q_p^{-1} + \int_0^t \frac{\partial H(x(\tau))}{\partial p}^\top V(\tau)^{-1} \frac{\partial H(x(\tau))}{\partial p} \, \mathrm{d}\tau. \tag{4}$$

Here, $Q_p \in \mathbb{S}_{++}{}^{n_p}$ is the a priori variance–covariance of the estimate for p which is for simplicity of presentation assumed to be positive definite. $F_p[t, u]$ is a symmetric and positive definite $n_p \times n_p$-matrix. Moreover, we have the monotonicity relation $F_p[t, u] \preceq F_p[t', u]$ for all $t, t' \in \mathbb{R}_+$ with $t' > t$ as the information about p can only increase when collecting more measurements.

Note that the above formulation of the Fisher information matrix can be extended to incorporate information on the initial values. The formulation of the Fisher information is then

$$F[t, u] := Q_0^{-1} + \int_0^t \frac{\partial H(x(\tau))}{\partial x_0}^\top V(\tau)^{-1} \frac{\partial H(x(\tau))}{\partial x_0} \, \mathrm{d}\tau. \tag{5}$$

Note that $F[t, u]$ is again a symmetric and positive semi-definite matrix but with different dimensions, i.e., $n_x \times n_x$, as $x_0 \in \mathbb{R}^{n_x}$ collects all unknown initial values which we want to estimate. In the following, we introduce the short hands

$$C(\tau) := \frac{\partial H(x(\tau))}{\partial x} \quad \text{and} \quad S(\tau) := \frac{\partial x(\tau)}{\partial x_0},$$

where $S(t)$ can also be computed from a variational differential equation of the form

$$\dot{S}(\tau) = A(\tau)S(\tau) \quad \text{with} \quad S(0) = I \quad \text{and} \quad A(\tau) := \frac{\partial f(x(\tau), u(\tau), 0)}{\partial x}.$$

Now, the Fisher information matrix can also be written as

$$F[t, u] = Q_0^{-1} + \int_0^t S(\tau)^\top C(\tau)^\top V(\tau)^{-1} C(\tau) S(\tau) \, \mathrm{d}\tau.$$

In the remainder of this paper we will use the latter formulation for the Fisher information matrix. Under the assumption of unbiased and uncorrelated Gaussian noise, the (approximate) Fisher information matrix (FIM) has an interesting property: its inverse $F[t, u]^{-1}$ approximates the parameter estimation variance–covariance matrix [26].

2.2. Variance-covariance approximation including process noise

For the case that we expect w to affect the experiment, we have to take this noise in the computations into account. In other words, we have to estimate both the unknown initial state x_0 and the function w. Unfortunately, the Fisher information matrix becomes in this case infinite dimensional, as w is a function in t. In principle, we could discretize w such that the Fisher information matrix can be constructed, but this would have the disadvantage that the size of the Fisher information matrix can become very large, as its dimension depends on the discretization accuracy. Thus, there arises the question how we can deal with this case numerically. In practice, we are typically not interested in optimizing the accuracy of the estimate for w. This is due to the fact that w is random anyway, while the system parameters are not supposed to change if we repeat the experiment. Thus, we propose in this paper to concentrate on the computation of the variance–covariance matrix of the estimated states only.

For this aim, we analyze the solution $x^*(T)$ of the least-squares problem in an uncertainty affine approximation: let us introduce the short hands

$$A(t) := \frac{\partial f(\bar{x}(t), u(t), 0)}{\partial x}, \quad B(t) := \frac{\partial f(\bar{x}(t), u(t), 0)}{\partial w}, \quad C(t) := \frac{\partial H(\bar{x}(t))}{\partial x},$$
$$d(t) := f(\bar{x}(t), u(t), 0), \quad e(t) := H(\bar{x}(t)).$$

Here, the function $\bar{x}$ is assumed to satisfy the nominal differential equation

$$\frac{\mathrm{d}\,\bar{x}(t)}{\mathrm{d}t} = f(\bar{x}(t), u(t), 0) \quad \text{with} \quad \bar{x}(0) = \eta_0.$$

The main idea of the approximation strategy is to compute a first order Taylor expansion of the solution of the original least-squares optimization problem (4) with respect to the auxiliary random variables ξ_0, ω, and ν. For this aim, we solve the linearized least-squares problem

$$\xi_{\mathrm{lin}}[u, \xi_0, \omega, \nu] := \underset{x}{\operatorname{argmin}} \min_{w} \quad J_{\mathrm{lin}}[x, w] \quad \text{s.t.} \quad \dot{x}(t) = A(t)x(t) + B(t)w(t) + d(t) \tag{6}$$

with $t \in [0, T]$. In this context, the objective functional J_{lin} is defined to be a quadratic Gauss–Newton approximation of the original nonlinear least-squares objective J. In other words, we define

$$J_{\text{lin}}[x, w] := \left\|x(0) - \xi_0\right\|^2_{Q_0^{-1}} + \int_0^T \left(\left\|w(t) - \omega(t)\right\|^2_{W(t)^{-1}} + \left\|C(t)x(t) + e(t) - \nu(t)\right\|^2_{V(t)^{-1}} \right) \mathrm{d}t.$$

Note that the function $\xi_{\text{lin}}[u, \xi_0, \omega, \nu]$ is well-defined and unique as we assume that the weighting matrices Q_0, $W(t)$, and $V(t)$ are all positive definite such that the optimization problem (6) is strictly convex.

The main advantage of the linear approximation is that we know that the random variable $\xi_{\text{lin}}[u, \xi_0, \omega, \nu](t)$ has for all $t \in [0, T]$ a Gaussian probability density $\rho_{\text{lin}}[t, u]$ as long as the auxiliary random variables ξ_0, ω, and ν satisfy

$$\mathbb{E}\{\omega(t)\} = 0, \quad \mathbb{E}\left\{\omega(t)\omega(t')\right\} = W(t)\delta(t - t'),$$
$$\mathbb{E}\{\nu(t)\} = 0, \quad \mathbb{E}\left\{\nu(t)\nu(t')\right\} = V(t)\delta(t - t')$$

for all $t, t' \in \mathbb{R}$ and $\xi_0 \in \mathcal{N}(\eta_0, Q_0)$. More precisely, the function $\rho_{\text{lin}}[t, u]$ can explicitly be written as

$$\forall x \in \mathbb{R}^{n_x} : \ \rho_{\text{lin}}[t, u](x) = (2\pi)^{-n_x/2} \det(Q[t, u])^{-1/2} \exp\left(-\frac{1}{2} \left(x - x^*_{\text{lin}}\right)^\top Q[t, u]^{-1} \left(x - x^*_{\text{lin}}\right) \right),$$

where we use the short hands

$$x^*_{\text{lin}}(t) := \xi_{\text{lin}}[u, \eta_0, 0, 0](t) \quad \text{and} \quad Q[t, u] := \mathbb{E}\{\xi_{\text{lin}}[u, \xi_0, \omega, \nu](t)\xi_{\text{lin}}[u, \xi_0, \omega, \nu](t)^\top\}. \tag{7}$$

As in the previous section, the variance–covariance matrix $Q[t, u]$ can be regarded as a functional in the control input u. Moreover, for the special case that we have no process noise, i.e., for $W(t) = 0$, we have the relation $Q[t, u] = S(t)F(t, u)^{-1}S(t)^\top$ [33] cf. Lemma 3.1. Being at this point, we have not yet discussed the details of how to compute $Q[t, u]$ numerically. These details will be elaborated in Section 3.

2.3. Optimal experiment design problems

The aim is to optimize a given quality measure $\Phi(Q[T, u])$ of the (approximate) variance–covariance matrix of the state estimate at time T. The optimal experiment design problem of interest can be summarized as

$$\min_u \quad \Phi(Q[T, u]) \quad \text{s.t.} \quad u(t) \in \mathbb{U}(t) \quad \text{forall} \quad t \in [0, T], \tag{8}$$

where $\mathbb{U}(t) \subseteq \mathbb{R}^{n_u}$ is the set of feasible control inputs at time t. Some popular choices for scalar design criteria are listed below [33,42,43,15]:

- D-optimal designs minimize the geometric mean of the eigenvalues of $Q[T, u]$, i.e., we choose $\Phi(Q[T, u]) := \text{Det}(Q[T, u])$. For the case that we have no process noise, we may also compute a scaled version of the Fisher information matrix and maximize $\text{Det}\left(S(T)^{-\top}F[T, u]S(T)^{-1}\right)$. This is equivalent as we have $\text{Det}(Q) = 1/\text{Det}\left(Q^{-1}\right)$ for any invertible matrix Q. Note that D-optimal experiment design has the advantage that the objective is scaling invariant [15], i.e., if we rescale the states, we end up with the same experiment design.
- E-optimal designs aim at minimizing the maximum eigenvalue of $Q[T, u]$, i.e., we choose $\Phi(Q[T, u]) := \lambda_{\max}(Q[T, u])$. Similar to the D-optimal design, we may alternatively maximize the minimum eigenvalue of $S(T)^{-\top}F[T, u]S(T)^{-1}$.
- A-optimal designs minimize the quadratic mean of the parameter estimation errors, i.e., we choose $\Phi(Q[T, u]) := \text{Tr}(Q[T, u])$. In contrast to D- and E-optimal design there is no "simple" formula for computing the A-optimal experiment design objective without inverting $F[T, u]$. For some application it has been reported that maximizing the heuristic objective $\text{Tr}(F[T, u])$ – or a suitably scaled version of the trace of the Fisher information matrix – can lead to an acceptable design [29]. However, we shall see that for the example in this paper a maximization of $\text{Tr}(F[T, u])$ can even lead to design which is worse than the design which is obtained by minimizing the term $\text{Tr}(Q[T, u])$.

Sometimes it is useful to scale the matrix $Q[t, u]$. For example, the scaled A-criterion can be written as $\Phi(Q[t, u]) = \text{Tr}\left(\Sigma Q[t, u]\Sigma^\top\right)$, where Σ is a given positive semi-definite scaling matrix. When we choose

$$\Sigma = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}, \tag{9}$$

we aim at optimizing those elements of $Q[t, u]$ which are associated with the parameter variance–covariance. This allows interpreting the formulation in a similar fashion as the current practice of optimal experiment design for parameter estimation in nonlinear dynamic models see [15] if we fix y_0.

A detailed discussion of other criteria can be found in [43]. In the following section, we shall see that even for the case that we have no process noise, a direct computation of the variance–covariance approximation $Q[T, u]$ can be cheaper than computing the Fisher information matrix $F[T, u]$. Thus, we propose in this paper to always use the design criteria which are based on $Q[T, u]$ rather than those based on $F[T, u]$.

Finally, we note that before performing the actual experiment, the optimal solution for $Q[T, u]$ should be assessed carefully in terms of its size. If some eigenvalues of $Q[T, u]$ are very large and cannot be improved further by optimization, taking the experiment might not be worthwhile, as some of the parameters cannot be estimated with sufficient accuracy. In this case, additional measurements have to be taken repeating the optimization of the experiment design for a different set of sensors or for a longer duration T.

3. Computation of the variance–covariance matrix

In order to numerically solve the linearized least-squares optimal control problem (6) there are basically two methods possible: a direct or an indirect method. Let us first concentrate on the direct approach where we first discretize the problem (6) and then solve the associated finite dimensional optimization problem. In the context of optimal experiment design this direct method has originally been proposed by Bauer et al. [4] and Körkel et al. [23]. In Section 3.1 we give a brief overview of their strategy, while Section 3.3 suggests an alternative computation based on an indirect method.

3.1. Computation of the variance–covariance matrix via a direct method

In order to discretize the infinite dimensional optimal control problem (6), we choose a piecewise constant discretization of the time-varying input

$$w(t) \approx \sum_{i=0}^{N-1} q_{i+1} I_{[t_i, t_{i+1})}(t),$$

where $I_{[a,b)}(t)$ is equal to 1 if $t \in [a, b)$ and equal to 0 otherwise. The time sequence $0 = t_0 < t_1 < \cdots < t_N = T$ can for example be equidistant. Moreover, we denote with $s_i \approx x(t_i)$ the approximation of the state at time t_i. Now, we can replace the problem (6) with its discrete version [8] of the form

$$\begin{aligned} \min_{s,q} \quad & \left\|s_0 - \xi_0\right\|^2_{Q_0^{-1}} + \sum_{i=1}^{N} h \left(\left\|q_i - \omega_i\right\|^2_{W(t_i)^{-1}} + \left\|C(t_i)s_i + e(t_i) - \nu_i\right\|^2_{V(t_i)^{-1}} \right) \\ \text{s.t.} \quad & s_{i+1} = S_x^i s_i + S_q^i q_{i+1} + r_i \qquad \text{forall} \quad i \in \{0, \ldots, N-1\}. \end{aligned} \tag{10}$$

Here, $S_x^i := G_x^i(t_{i+1})$, $S_q^i := G_q^i(t_{i+1})$, and $r^i := g_i(t_{i+1})$ can be computed from the associated variational differential equation

$$\dot{G}_x^i(t) = A(t)G_x^i(t), \quad \dot{G}_q^i(t) = A(t)G_q^i(t) + B(t), \quad \dot{g}_i(t) = A(t)g_i(t) + d(t) \tag{11}$$

using the initial values $G_x^i(t_i) = I$, $G_q^i(t_i) = 0$, and $g_i(t_i) = 0$. Finally, we summarize the discrete optimization variable in one vector

$$l := (s_0^\top, \ldots, s_{N+1}^\top, q_1^\top, \ldots, q_N^\top)^\top.$$

Using this notation, the discrete optimization problem (10) can be written in the more compact notation

$$\min_l \left\|J_1 l + F_1[\xi_0, \omega, \nu]\right\|_2^2 \quad \text{s.t.} \quad J_2 l + F_2 = 0, \tag{12}$$

where $F_1[\xi_0, \omega, \nu], F_2, J_1$, and J_2 are defined in such a way that the optimization problems (10) and (12) are equivalent. Note that the solution of the discretized least-squares optimization problem is an approximation of the exact solution $x^*_{\text{lin}}(t)$ which has been defined in (7). Let us define a projection matrix P_i such that we have

$$s_i = \begin{pmatrix} P_i \\ 0 \end{pmatrix}^\top l.$$

Now, the discrete approximation of $\xi_{\text{lin}}[u, \xi_0, \omega, \nu](t_i)$ can be written as

$$\xi_{\text{lin}}[u, \xi_0, \omega, \nu](t_i) = \begin{pmatrix} P_i \\ 0 \end{pmatrix}^\top \begin{pmatrix} J_1^\top J_1 & J_2^\top \\ J_2 & \end{pmatrix}^{-1} \begin{pmatrix} J_1^\top F_1[\xi_0, \omega, \nu] \\ F_2 \end{pmatrix} + \mathbf{O}(h). \tag{13}$$

Note that this approximation becomes exact if the mesh size h converges to zero, i.e., if we consider the limit for $h := \max_i \left|t_{i+1} - t_i\right| \to 0$. The discretization error is of order $O(h)$, i.e., it converges linearly to 0 if the mesh size h is refined. In addition, we are interested in the associated discrete approximation of the variance–covariance function $Q[\cdot, u]$ which has also been defined in (7). This approximation can be written as $Q[t_i, u] = Q_i + \mathbf{O}(h)$, where we define

$$Q_i := \begin{pmatrix} P_i \\ 0 \end{pmatrix}^\top \begin{pmatrix} J_1^\top J_1 & J_2^\top \\ J_2 & \end{pmatrix}^{-1} \begin{pmatrix} J_1^\top J_1 & 0 \\ 0 & \end{pmatrix} \begin{pmatrix} J_1^\top J_1 & J_2^\top \\ J_2 & \end{pmatrix}^{-1} \begin{pmatrix} P_i \\ 0 \end{pmatrix}. \tag{14}$$

Note that the above discrete approximation for $Q[t, u]$ has been used in [4] in the context of optimal experiment design. This matrix expression (14) can be further simplified [9,24] to

$$Q_i := \begin{pmatrix} P_i \\ 0 \end{pmatrix}^\top \begin{pmatrix} J_1^\top J_1 & J_2^\top \\ J_2 & \end{pmatrix}^{-1} \begin{pmatrix} P_i \\ 0 \end{pmatrix}. \tag{15}$$

This expression can be obtained by using the following property $KJ_1^T J_1 K = K$ with

$$\begin{pmatrix} J_1^\top J_1 & J_2^\top \\ J_2 & 0 \end{pmatrix}^{-1} := \begin{pmatrix} K & L \\ L^\top & M \end{pmatrix}.$$

The property can easily be checked by writing the upper left block of the inverse of the KKT matrix by employing the Schur-complement or Null-space projection. In order to verify Eq. (14) we expand the expression for $F_1[\xi_0, \omega, \nu]$ which reads

$$F_1[\xi_0, \omega, \nu] = \left(Q_0^{-1/2}\xi_0,\ \sqrt{h}W(t_1)^{-1/2}\omega_1, \ldots,\ \sqrt{h}W(t_N)^{-1/2}\omega_N,\ \sqrt{h}V(t_1)^{-1/2}\nu_1, \ldots,\ \sqrt{h}V(t_N)^{-1/2}\nu_N \right).$$

As the discrete random variables have the variance–covariance terms $\mathbb{E}\{(\xi_0 - \eta_0)(\xi_0 - \eta_0)^\top\} = Q_0$, $\mathbb{E}\{\omega_i \omega_i^\top\} = \frac{1}{h}W(t_i)$, as well as $\mathbb{E}\{\nu_i \nu_i^\top\} = \frac{1}{h}V(t_i)$ for all $i \in \{1, \ldots, N\}$, we find

$$\mathbb{E}\left\{ F_1[\xi_0, \omega, \nu] F_1[\xi_0, \omega, \nu]^\top \right\} = I.$$

Using this relation in combination with Eq. (13) yields

$$Q[t_i, u] = \mathbb{E}\left\{ \xi_{\text{lin}}[u, \xi_0, \omega, \nu](t_i) \xi_{\text{lin}}[u, \xi_0, \omega, \nu](t_i)^\top \right\} = Q_i + \mathbf{O}(h)$$

for all $i \in \{0, \ldots, N\}$ which explains the connection between continuous time definitions (7) and their discrete-time approximations (13) and (14).

3.2. The cost of computing the variance–covariance matrix with a direct approach

When the formulation (15) is used to compute the objective value, the algorithm depends crucially on the cost of computing the matrices Q_i. This computational cost depends mainly on the cost of computing J_1 and J_2 as well as on the cost of evaluating the expression (15):

- The cost of computing the non-zero entries of the matrix J_1 is mainly determined by the cost of computing the derivative matrices $C(t_i) = \frac{\partial H(\overline{\eta}(t))}{\partial x}$ of the function H. Depending on the dimension of h and x forward or backward automatic differentiation can be used. However, the cost of computing the matrices $C(t_i)$ is in practice often negligible in comparison to the cost of computing the matrix J_2. Here, the blocks S_x^i and S_q^i can be computed by propagating the variational differential equations (11) for G_x^i and G_q^i in forward mode. This requires $n_x + n_w$ forward directions. However, if n_w is larger than zero it is usually advisable to compute these matrices by backward automatic differentiation which requires only n_x backward directions.
- For the computation of the inverse matrix which occurs in the expression (15) there are basically two options. Firstly, we can use a condensing approach which exploits the sparsity of the matrix J_2 by block Gauss elimination. The details of this condensing approach can for example be found in [7]. If we introduce the notation $\overline{n} := \max\{n_x, n_y, n_w\}$, we need

$$\mathbf{O}\left(N^3 n_w^2 \overline{n}\right) + \mathbf{O}\left(N^2 n_w \overline{n}^2\right) + \mathbf{O}\left(N \overline{n}^3\right) \tag{16}$$

flops to compute the matrix Q_N with the condensing approach. Note that in most of the practically relevant situations we have $n_y < n_x$ and $n_w < n_x$ such that we may substitute $\overline{n} = n_x$. The second option is to use a discrete Riccati recursion [20] for the computation of the sequence

$$\begin{aligned} P_i &:= S_x^i Q_i {S_x^i}^\top + \frac{1}{h} S_q^i W(t_{i+1}) {S_q^i}^\top, \\ Q_{i+1} &= P_i - h P_i C(t_{i+1})^\top \left[h C(t_{i+1}) P_i C(t_{i+1})^\top + V(t_{i+1}) \right]^{-1} C(t_{i+1}) P_i \end{aligned} \tag{17}$$

for $i \in \{0, \ldots, N-1\}$, which exploits sparsity in a different manner. If this recursion is implemented, we need in general only $\mathbf{O}\left(N \overline{n}^3\right)$ flops to compute Q_N. Thus, if we have $n_w = 0$, both strategies have the same complexity. However, if the integer $N^2 n_w$ is not small compared to n_x the Riccati recursion is more efficient than the condensing strategy. Note that the recursion (17) for the matrices Q_i can in principle be obtained by expanding the block matrix representation of the matrices J_1 and J_2 which can be substituted in Eq. (14). However, a more intuitive and direct derivation of (17) can be obtained by interpreting the current matrix $Q_i \doteq \text{VAR}(x(t_i))$ as the linear approximation of the variance covariance matrix of the state $x(t_i)$ at time t_i. In this interpretation, the matrix P_i corresponds to a forward propagation of the matrix Q_i from time t_i to time t_{i+1} without taking the next measurement into account yet. Thus, we find

$$\text{VAR}(x(t_{i+1})) \doteq \text{VAR}\left(S_x^i x(t_i) + S_q^i \omega_{i+1} + r_i \right) \doteq S_x^i Q_i {S_x^i}^\top + \frac{1}{h} S_q^i W(t_{i+1}) {S_q^i}^\top =: P_i.$$

Once the next measurement at time t_{i+1} arrives the incoming information, given in form of the Fisher information matrix $hC(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1})$, must be added to the forwarded information P_i^{-1}. Thus, the information update at time t_{i+1} can be written as

$$Q_{i+1}^{-1} = P_i^{-1} + hC(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1}).$$

Taking the inverse on both sides and applying the Sherman–Morrison–Woodbury formula yields the discrete Riccati recursion (17).

3.3. Computation of the variance–covariance matrix via an indirect method

An alternative numerical strategy for computing the matrix $Q[T,u]$ is to use an indirect approach for solving the linearized least-squares optimal control problem (6). Here, we first compute the optimal solution by using Pontryagin's maximum principle [32]. It is well-known [21] that an application of this strategy yields a Riccati differential equation for the function Q, which can be written as

$$\begin{aligned}\dot{Q}[t,u] &= A(t)Q[t,u] + Q[t,u]A(t)^\top + B(t)W(t)B(t)^\top - Q[t,u]C(t)^\top V(t)^{-1}C(t)Q[t,u]\\ Q[0,u] &= Q_0.\end{aligned} \tag{18}$$

As an alternative to the derivation with Pontryagin's maximum principle or the Hamilton–Jacobi–Bellman equation, the above Riccati differential equation can also be obtained by starting with the discrete recursion (17) and computing the limit for $h \to 0$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}h}Q_{i+1}^{-1}\Big|_{h=0} &= \frac{\mathrm{d}}{\mathrm{d}h}\left[P_i^{-1} + hC(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1})\right]^{-1}\Big|_{h=0} = P_i\left(\frac{\mathrm{d}}{\mathrm{d}h}P_i^{-1} + C(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1})\right)P_i\Big|_{h=0}\\ &= P_i\left(P_i^{-1}\left[\frac{\mathrm{d}}{\mathrm{d}h}P_i\right]P_i^{-1}\right)P_i\Big|_{h=0} + Q_iC(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1})Q_i = \frac{\mathrm{d}}{\mathrm{d}h}P_i\Big|_{h=0} + Q_iC(t_{i+1})V(t_{i+1})^{-1}C(t_{i+1})Q_i,\end{aligned}$$

where we can use

$$\frac{\mathrm{d}}{\mathrm{d}h}P_i\Big|_{h=0} = A(t_i)Q_i + Q_iA(t_i)^\top + B(t_i)W(t_{i+1})B(t_i)^\top$$

as well as $t_{i+1} \to t_i =: t$ for $h \to 0$ such that Eq. (18) is obtained in the limit. In this sense, a computation of the variance–covariance-matrix $Q[t,u]$ via the above differential equation is very similar to the direct approach in combination with the discrete Riccati recursion. However, the continuous formulation has the advantage that the Riccati differential equation can very easily be added in a standard optimal control solver for continuous time dynamics. Thus, this has the advantage that we can employ existing numerical integration and optimization methods which will automatically take care of an accurate discretization. Moreover, we show in the following section how we can exploit the structure of the Riccati equation within derivative based optimal control algorithms.

Lemma 3.1. *For the special case that we have no time-varying uncertainty, i.e., for the case $W(t) = 0$, Eq. (17) simplifies to*

$$\begin{aligned}\dot{Q}[t,u] &= A(t)Q[t,u] + Q[t,u]A(t)^\top - Q[t,u]C(t)^\top V(t)^{-1}C(t)Q[t,u]\\ Q[0,u] &= Q_0.\end{aligned} \tag{19}$$

In this case, we have the inverse relation $Q[t,u] = S(t)F[t,u]^{-1}S(t)^\top$ between the Fisher information matrix $F[t,u]$ and the variance covariance matrix $Q[t,u]$.

Proof. Note that the Fisher information matrix satisfies by definition a differential equation of the form

$$\begin{aligned}\dot{F}[t,u] &= S(t)^\top C(t)^\top V(t)^{-1}C(t)S(t)\\ F[0,u] &= Q_0^{-1}.\end{aligned}$$

Now, we compute

$$\frac{d}{dt}\left[S(t)F[t,u]^{-1}S(t)^\top\right] = A(t)\left[S(t)F[t,u]^{-1}S(t)^\top\right] + \left[S(t)F[t,u]^{-1}S(t)^\top\right]A(t)^\top + S(t)\left[\frac{d}{dt}F[t,u]^{-1}\right]S(t)^\top,$$

where

$$\frac{d}{dt}F[t,u]^{-1} = -F[t,u]^{-1}\dot{F}[t,u]F[t,u]^{-1} = -F[t,u]^{-1}S(\tau)^\top C(\tau)^\top V(\tau)^{-1}C(\tau)S(\tau)F[t,u]^{-1}.$$

This yields

$$\begin{aligned}\frac{d}{dt}\left[S(t)F[t,u]^{-1}S(t)^\top\right] &= A(t)\left[S(t)F[t,u]^{-1}S(t)^\top\right] + \left[S(t)F[t,u]^{-1}S(t)^\top\right]A(t)^\top\\ &\quad + \left[S(t)F[t,u]^{-1}S(t)^\top\right]C(t)^\top V(t)^{-1}C(t)\left[S(t)F[t,u]^{-1}S(t)^\top\right].\end{aligned}$$

In other words, the term $S(t)F[t,u]^{-1}S(t)^\top$ and the function $Q[t,u]$ satisfy the same differential equation and they have the same initial value $S(0)F[0,u]^{-1}S(0)^\top = Q_0 = Q(0)$. Consequently, we must have $Q[t,u] = S(t)F[t,u]^{-1}S(t)^\top$, as the solution of Lipschitz continuous differential equations is unique. □

Note that an implementation of the Riccati differential equation (18) requires besides the integration of the nominal differential equation $\frac{n_x(n_x+1)}{2}$ additional states, as we can exploit symmetry. In contrast to this, a direct computation of the Fisher information matrix requires us to solve first the variational differential equation for $S(t)$ needing n_x^2 states and second we need another $\frac{n_x(n_x+1)}{2}$ additional states for the evaluation of the integral term in Eq. (5). Note that if y_0 is fixed, then the Fisher information approach is the cheapest.

3.4. Derivatives of the variance–covariance matrix

As it was outlined in Section 2, we are in the field of optimal experiment design interested in optimizing a term of the form $\Phi(Q[T,u])$. For most of the optimal control algorithms it is important to compute derivatives efficiently. In the case that we work with multiple-shooting or collocation algorithms, we are interested in computing the derivative of $Q[t,u]$ with respect to the initial matrix Q_0 which is typically

the most expensive block as we are dealing with a matrix derivative. If we compute the corresponding term $\frac{\partial Q[t,u]}{\partial Q_0}$ without exploiting any structure we have to introduce a variational differential equation with $\mathbf{O}(n_x^4)$ states. Fortunately, a deeper analysis shows that the Riccati differential equation has an important structure which can be exploited. In fact, it is possible to compute the derivative $\frac{\partial Q[t,u]}{\partial Q_0}$ by propagating a differential equation with $\mathbf{O}(n_x^2)$ states only. The corresponding strategy is summarized in the following theorem:

Theorem 3.1. *Let us assume that Q_0 is given in such a way that the differential equation (18) admits a solution Q on the interval $[0, T]$. Then the differential equation*

$$\forall t \in [0,T]: \quad \dot{R}(t) = \left[A(t) + Q(t)C(t)^\top V(t)^{-1}C(t)\right] R(t) \quad \text{with} \quad R(0) = I \tag{20}$$

admits a solution $R : \mathbb{R} \to \mathbb{R}^{n_x \times n_x}$. Moreover, the derivative of $Q(t)$ with respect to the initial value Q_0 exists for all $t \in [0, T]$ and can be written as

$$\forall t \in [0,T]: \quad \frac{\partial Q(t)}{\partial Q_0} = R(t) \otimes R(t).$$

In this context, $\otimes$ denotes the Kronecker product.

Proof. Let us first compute the directional derivative $G(t) := \frac{\partial Q(t)}{\partial Q_0}\Delta$ for a given test direction $\Delta \in \mathbb{R}^{n\times n}$ with $\Delta = \Delta^\top$. A differentiation of the Riccati ODE (18) yields a differential equation for G:

$$\begin{aligned} &\forall t \in [0,T]: \quad \dot{G}(t) = A(t)G(t) + G(t)A(t)^\top + G(t)C(t)^\top V(t)^{-1}C(t)Q(t) + Q(t)C(t)^\top V(t)^{-1}C(t)G(t) \\ &G(0) = \Delta. \end{aligned} \tag{21}$$

In the next step, we show that G can be written as $G(t) = R(t)\Delta R(t)^\top$, where R satisfies the differential equation (20). For this aim, we expand for all $t \in [0, T]$:

$$\begin{aligned} \dot{G}(t) &= \dot{R}(t)\Delta R(t)^\top + R(t)\Delta\dot{R}(t)^\top = \left[A(t) + Q(t)C(t)^\top V(t)^{-1}C(t)\right] R(t)\Delta R(t)^\top + R(t)\Delta R(t)^\top \left[A(t) + Q(t)C(t)^\top V(t)^{-1}C(t)\right]^\top \\ &= A(t)G(t) + G(t)A(t)^\top + G(t)C(t)^\top V(t)^{-1}C(t)Q(t) + Q(t)C(t)^\top V(t)^{-1}C(t)G(t) \\ G(0) &= R(0)\Delta R(0)^\top = \Delta. \end{aligned} \tag{22}$$

Comparing the differential equations (21) and (22) yields a proof of the theorem. □

Remark 3.1. The idea behind the above theorem can also be used in combination with the discrete Riccati recursion (17). In this case, we introduce auxiliary iterates T_i which satisfy the discrete recursion

$$\forall i \in \{0, \ldots, N-1\}: \quad T_{i+1} = Q_{i+1}P_i^{-1}S_x^i R_i \quad \text{with} \quad T_0 = I.$$

Now, we have $\frac{\partial Q_i}{\partial Q_0} = R_i \otimes R_i$ for all $i \in \{0, \ldots, N\}$ which follows trivially by using the Sherman–Morrison–Woodbury formula in combination with (17) and other standard matrix differentiation rules. Analogous to the continuous case, the main point is also here that the recursion can be implemented by operating on the matrices R_i only while the tensor $\frac{\partial Q_i}{\partial Q_0}$ is never stored explicitly.

3.5. Reformulation of optimal experiment design problems as standard optimal control problems

As mentioned above, the computation of the variance–covariance $Q[t, u]$ via the indirect approach using the Riccati differential equation has the advantage that the problem (8) can be reformulated into a standard optimal control problem for which existing solvers can be employed. More precisely, the above considerations have shown that the problem (8) is equivalent to an optimal control problem of the form

$$\begin{aligned} &\inf_{\bar{x}(\cdot), Q(\cdot), u(\cdot)} \Phi(Q[T,u]) \\ &\text{s.t.} \begin{cases} \dfrac{\partial Q[t,u]}{\partial t} = A(t)Q[t,u] + Q[t,u]A(t)^\top + B(t)W(t)B(t)^\top - Q[t,u]C(t)^\top V(t)^{-1}C(t)Q[t,u] \\ Q[0,u] = Q_0 \\ \dfrac{\partial \bar{x}(t)}{\partial t} = f(\bar{x}(t), u(t), 0) \\ \bar{x}(0) = \eta_0 \\ u(t) \in \mathbb{U}(t) \quad \text{forall } t \in [0,T]. \end{cases} \end{aligned} \tag{23}$$

Here, we recall that the functions $A(t) = \frac{\partial f(\bar{x}(t),u(t),0)}{\partial x}$, $B(t) := \frac{\partial f(\bar{x}(t),u(t),0)}{\partial w}$, and $C(t) := \frac{H(\bar{x}(t))}{\partial x}$ depend implicitly on $\bar{x}(t)$ and $u(t)$.

In order to solve problems of the form (23) numerically, we employ in this paper a multiple-shooting based SQP algorithm (cf. [7]) using the software `ACADOtoolkit` (cf. [19]). Here, the main difference to the optimal experiment design algorithm from Bauer et al. [4] is that the computation of the variance–covariance function Q is based on forward propagation of a Riccati differential equation. As this differential equation is in the implementation solved by an adaptive integrator we obtain a high accuracy in computing Q – although we might argue that this is not crucial in the sense that Q is anyhow only a linear approximation of the variance–covariance of the state estimate. However, there is another important aspect: the Riccati differential equation has a very fortunate numerical structure which can be exploited to obtain an efficient implementation, i.e., we can exploit Theorem 3.1 within an efficient multiple-shooting implementation.

4. Optimal experiment design for a fed-batch bioreactor

4.1. The dynamic process model

In order to benchmark the techniques for optimal experiment design, a well-mixed fed-batch bioreactor model (cf. [41]) is used as case study. The dynamic model equations are given by:

$$\frac{dC_s}{dt} = -\sigma C_x + \frac{u}{v}C_{s,in} - \frac{u}{v}C_s + w_1 \tag{24}$$

$$\frac{dC_x}{dt} = \mu C_x - \frac{u}{v}C_x + w_2 \tag{25}$$

$$\frac{d\mu_{max}}{dt} = 0 \tag{26}$$

$$\frac{dK_s}{dt} = 0 \tag{27}$$

$$\frac{dv}{dt} = u \tag{28}$$

in which C_s [g/L] is the concentration limiting substrate, C_x [g/L] the biomass concentration and v [L] the bioreactor volume. Note that we use the formulation where the unknown parameters are stacked as trivial differential equations, as we explained in Section 2. The function u [L/h] denotes the volumetric feed rate, containing a substrate concentration $C_{s,in}$. The specific growth rate studied in this case is of the monotonic *Monod* type. The corresponding algebraic relation is given by:

$$\mu = \mu_{max}\frac{C_s}{K_s + C_s}. \tag{29}$$

The substrate consumption rate is in the case study modeled by an affine dependence of the form:

$$\sigma = \frac{\mu}{Y_{X|S}} + m, \tag{30}$$

where $Y_{X|S}$ is the yield and m the maintenance factor. The parameter values are given in Table 1. The current best estimate for the parameters μ_{max} and K_s are represented as $\overline{\mu}_{max}$ and $\overline{K}_s$. The amount of biomass X and substrate S are given by: $C_x \cdot v = X$ and $C_s \cdot v = S$. The initial amount of biomass available is 10.5 g. The initial volume without substrate $v^*(0) = 7$ L. As the maximum volume $v_{max} = 10$ L and the total amount of limiting substrate available is 1500 g, the initial condition of the substrate and the volume, $S(0)$ and $v(0)$ are connected as follows: $S(0) + C_{s,in}(v_{max} - v(0)) = \alpha$. The feed rate u and the end time t_f are constrained by:

$$0 \le u \le 1\,\text{L/h} \tag{31}$$

$$5 \le t_f \le 40\,\text{h}. \tag{32}$$

The states C_s and C_x can be measured. Thus, the observation function is $H(t) = (C_s, C_x)^\top$. We assume that the measurements of the states are white noise signals satisfying the modeling assumption (3). The associated variance matrix is given as $V(t) = \text{diag}(\sigma^2_{C_s}, \sigma^2_{C_x}, 0, 0)^\top$. The initial variances of the states are

$$Q(0) = \text{diag}\left(10 \times \sigma^2_{C_s}, 10 \times \sigma^2_{C_x}, 0.05\left(\frac{1}{\text{h}}\right)^2, 0.5\left(\frac{\text{g}}{\text{L}}\right)^2\right)^\top. \tag{33}$$

The remaining non-diagonal components of the matrix $Q(0)$ are all 0. Note that the matrix Q is only a 4×4 matrix, as the differential state v is not affected by the uncertainty. The time-varying random disturbances influencing the process model are denoted by $w_1(t)$ and $w_2(t)$.

4.2. Optimal experiment design without process noise

In this section we design an optimal experiment for the fed-batch bioreactor. We first compare two different design criteria: the A-criterion, i.e., the trace of the variance–covariance matrix and the trace of the Fisher information matrix (TF). We only take the elements into account which are related to the uncertain parameters, μ_{max} and K_s. As we do not consider process noise, we have $w_1 = w_2 = 0$.

In order to solve the optimization problems, a multiple-shooting approach is adopted. We use a piecewise constant control discretization with 30 intervals. The used integrator is a RK78 with absolute and relative tolerance set to 10^{-6}. The nonlinear program is solved by a SQP routine with tolerances set to 10^{-6}. The Fisher information matrix is computed by using the outlined approach from Section 2.1 while the approximation of the variance–covariance matrix is computed as explained in Section 3.3.

In Fig. 1 the states and controls of the optimal designed experiments are depicted using the trace of the Fisher information matrix and the A-criterion as objectives.

Table 1
Parameter values for the fed-batch bioreactor.

$\overline{\mu}_{max}$	0.1 1/h	$\overline{K}_s$	1 g/L
m	0.29 g/g	$Y_{X\|S}$	0.47 g/g
$\sigma^2_{C_s}$	1×10^{-2} g^2/L^2	$\sigma^2_{C_x}$	6.25×10^{-4} g^2/L^2

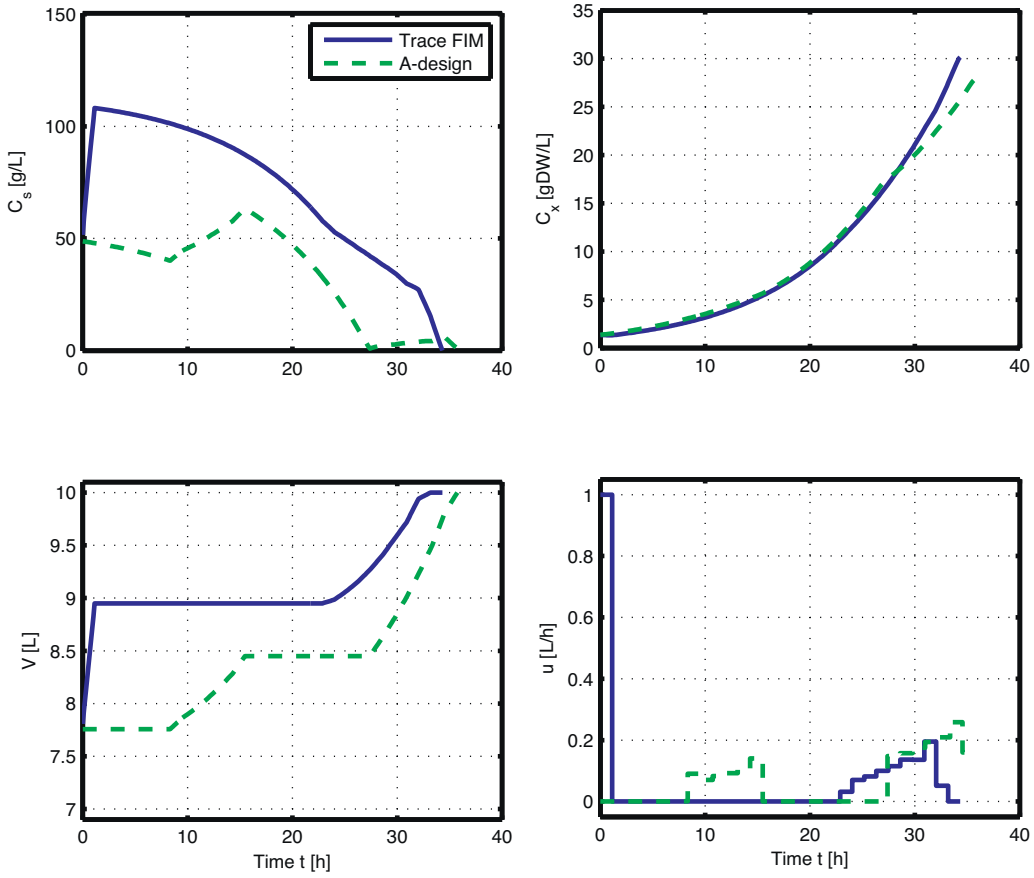

Fig. 1. Evolution of the states and the controls for the trace of the Fisher information and the A-criterion approach.

We observe that both experiments start with a similar amount of substrate concentration. The experiment of the trace of the Fisher information matrix approach starts with a feed rate u^*_{TF} set to its maximal value and is thus increasing the substrate concentration to above 100 g/L. The remainder of the experiment consists of a batch phase. After 25 h the feeding is resumed at a lower feed rate. The experiment obtained by the A-criterion approach, u^*_{A}, starts with a batch phase. The first feeding phase starts after 8 h. The substrate concentration subsequently increases but it remains below the concentration observed in the experiment obtained by the trace of the Fisher information matrix. After the first feeding phase, a batch phase follows up to 26 h after which a second feeding period takes place until the maximal volume of the reactor is reached. Note that the experiment obtained by the A-criterion approach results in slightly less biomass formed than the approach using the trace of the Fisher information matrix.

In order to analyze the difference between both approaches, we also look at the difference with respect to the obtained approximate variance–covariance matrices. The evolution of the diagonal elements of the approximate variance–covariance matrices is illustrated in Fig. 2. The diagonal elements Q_{33} and Q_{44} indicate the uncertainty of the parameters $\mu_{\max}$ and K_{s}. As indicated in Fig. 2, the uncertainty regarding these parameters are reduced throughout the experiments. The A-criterion approach reaches a smaller value for these states compared with the trace of the Fisher information matrix. From Fig. 2 we can infer that parameter K_{s} is the limiting factor in reducing the trace of the approximate variance–covariance matrix. We also note that the uncertainty with respect to the states of the model decreases continuously until additional substrate is fed which leads to an increase in the uncertainty.

To compare the uncertainty at the end of the experiment, we show the uncertainty ellipsoids related to the approximate variance–covariance matrices. From Fig. 3 we can conclude that the A-criterion approach leads to an uncertainty ellipsoid which is smaller than the one obtained by the trace of the Fisher information matrix approach. The uncertainty ellipsoid of the trace of the Fisher information matrix indicates that the experiment leads to an experiment where the parameter values are more correlated. This is indicated by the more tilted axes of the ellipsoid. This problem illustrates that using the approach with the Fisher information matrix can suffer some drawback compared with the variance–covariance case. This formulation was performed because practitioners want to avoid inverting the Fisher information matrix, our presented technique takes away this difficulty with respect to the A-criterion.

In addition to the comparison of the A-criterion and the trace of the Fisher information matrix we also compare the experiments obtained by the D- and E-criterion for both approaches.

The optimization problems are solved by a multiple-shooting approach. The controls are discretized in 30 piecewise constant parts. The integrator tolerances are decreased to 10^{-10} as the elements of the variance–covariance matrix need to be computed more accurately. The nonlinear program is solved with a KKT-tolerance set to 10^{-10}.

In Fig. 4, we depict the obtained states and controls for the obtained optimal experiments for the D- and E-criterion in both the Fisher information matrix framework as well as the variance–covariance matrix approach.

Note that for the D- and E-criterion the variance–covariance matrix approach resulted in the same experiment. In [39] it is observed that the E-criterion optimization profile results in a D-criterion value which is very close to the optimal D-criterion value. Similar observations

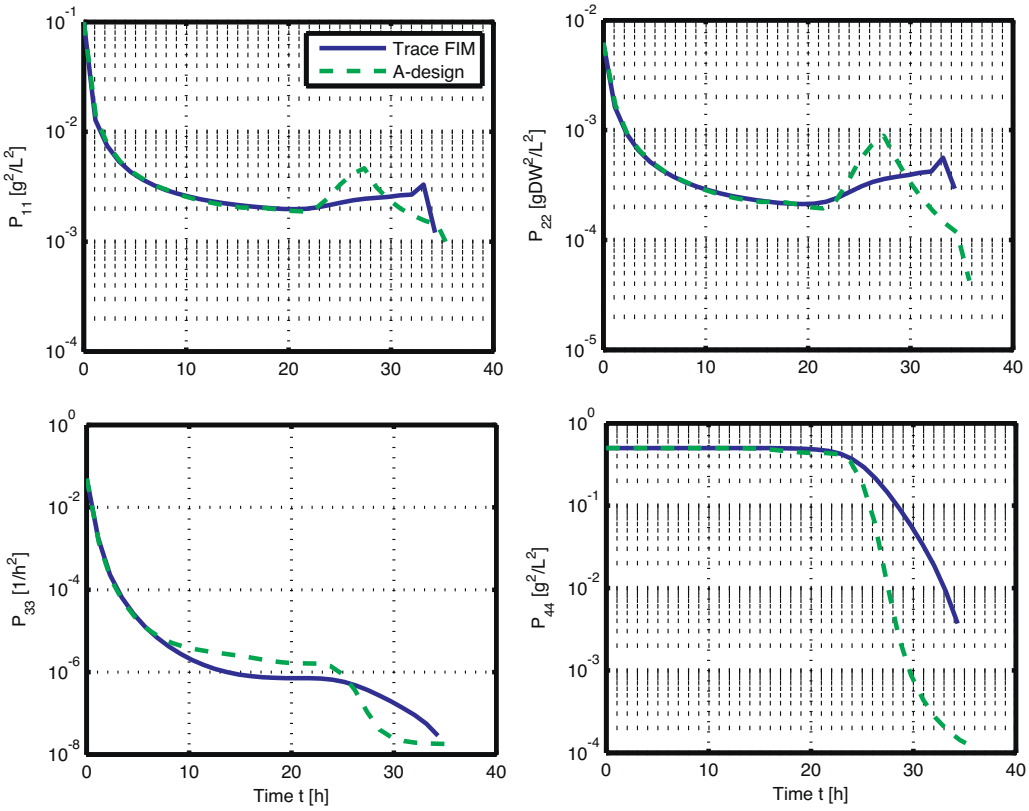

Fig. 2. Evolution of the diagonal elements of the approximate variance–covariance matrix for the experiments obtained by the trace of the Fisher information matrix and the A-criterion approaches.

for a different case study are found in [40]. For all three experiments, a somewhat similar initial feeding pattern is observed, only the experiment obtained by the variance–covariance matrix has an initial short feeding phase. After 20 h the D-design of the Fisher information approach gradually increases until at 30 h its maximum feed rate is reached. The other two experiments each have a peak up to the maximal allowed feeding rate which results in a maximum substrate concentration. The substrate is then consumed for all cases until the end of the experiment.

In Fig. 5, we present the resulting evolution of the diagonal elements of the variance–covariance matrix. From the figure it is clear that K_s is the limiting factor. The uncertainty regarding $\mu_{\max}$ is similar for all three experiments, however for K_s the variance–covariance approach is able to reduce the uncertainty more than the Fisher information matrix approach. Note that the jump to the maximal feeding

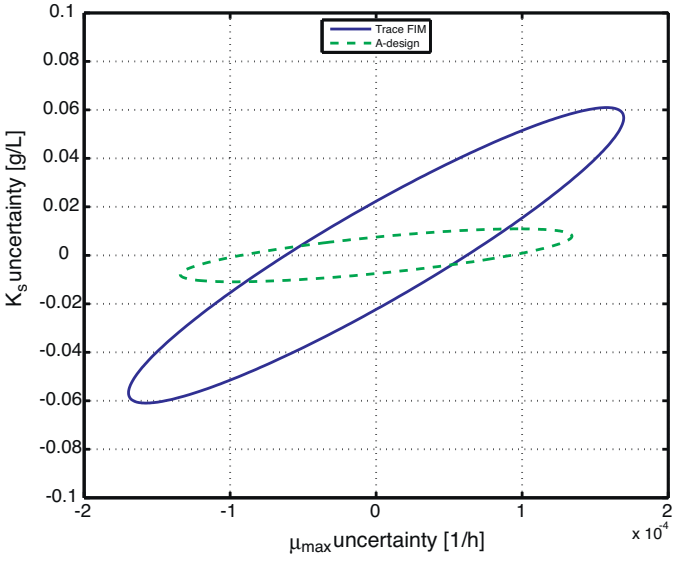

Fig. 3. Uncertainty ellipsoids as measured using the approximate variance–covariance matrix for the designed u^*_{TF} and u^*_A.

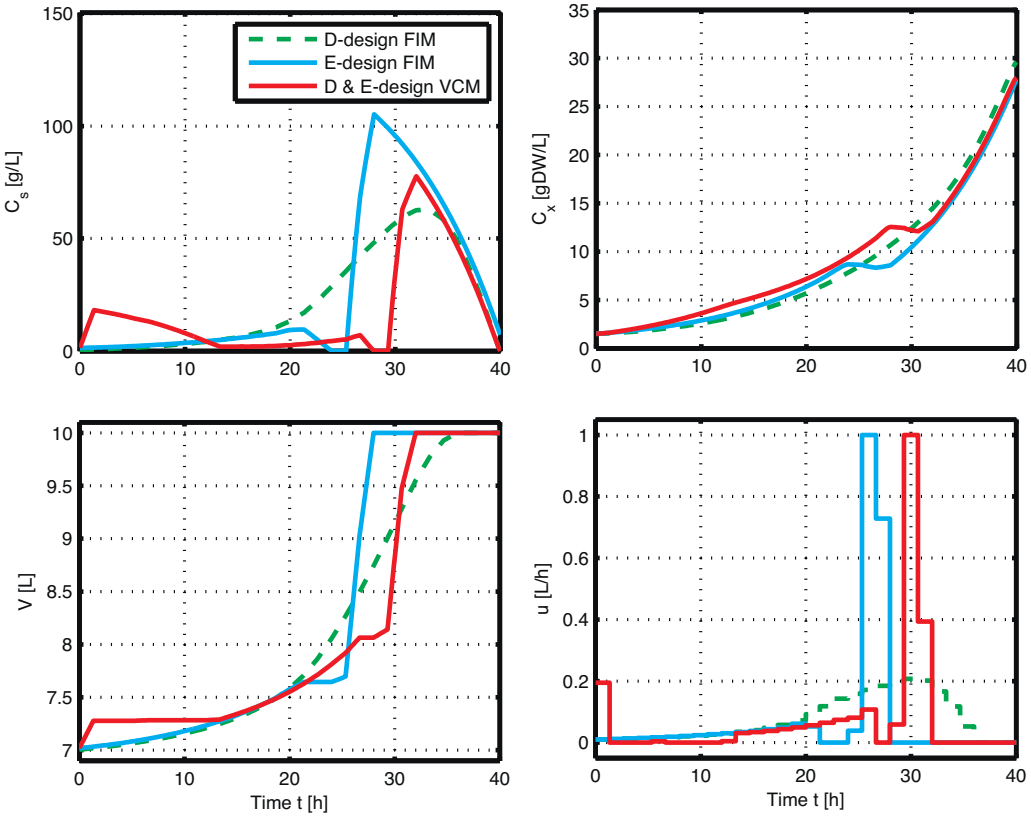

Fig. 4. Evolution of the states and the controls for the D- and E-criterion approach for both the Fisher information approach and the variance–covariance matrix approach.

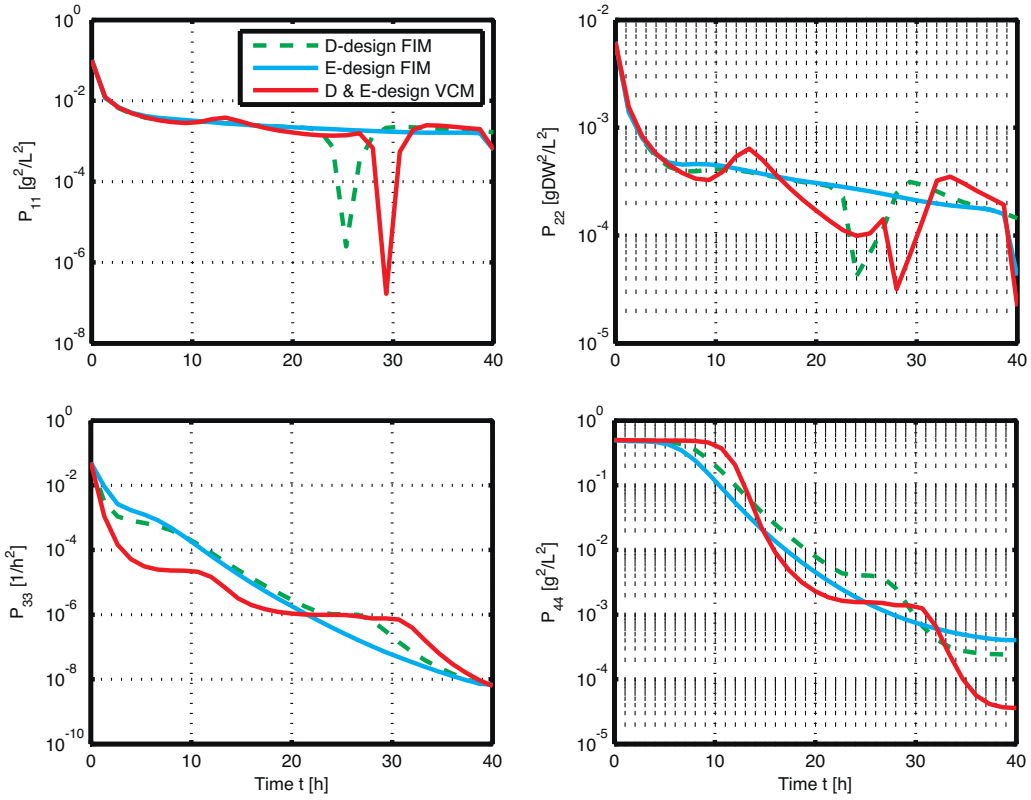

Fig. 5. Evolution of the diagonal elements of the approximate variance–covariance matrix for the experiments obtained by the D- and E-criterion approach for both the Fisher information approach and the variance–covariance matrix approach.

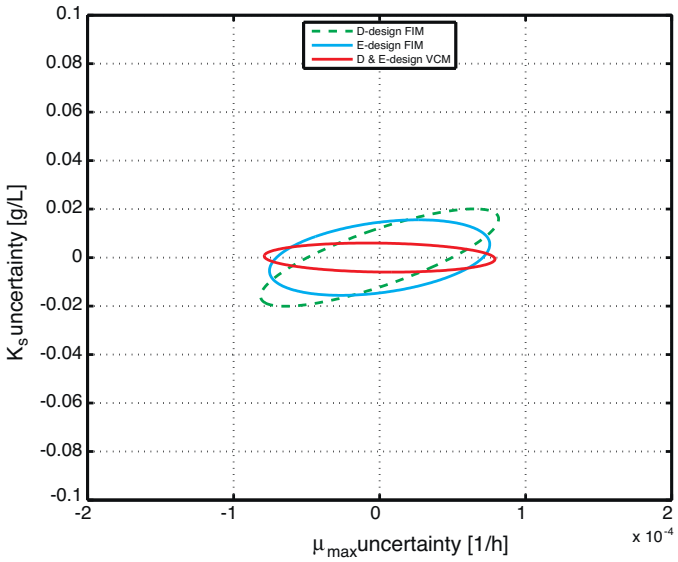

Fig. 6. Uncertainty ellipsoids as measured using the approximate variance–covariance matrix for the designed $u^*_{D_{FIM}}$, $u^*_{E_{FIM}}$ and $u^*_{D,E_{AVC}}$.

rate is clearly visible in the uncertainty of the states. Also the lowest variance of the states is not at the end of the experiment as we do not aim at reducing the state uncertainty in our objective.

The expected confidence ellipsoids are illustrated in Fig. 6. The smallest region is obtained by the design of the variance–covariance matrix approach. When we compare the D-design of the Fisher information approach with the E-design, we observe a slightly smaller joint confidence region for the D-design. These observations are similar as in the A-criterion versus trace of the Fisher information case, but the difference between the two approaches is less pronounced.

4.3. Optimal experiment design with process noise

In this section we study the influence of process noise on the optimal experiment design. One of the advantages of the proposed method for computing the variance–covariance matrix is that this process noise can be taken into account in the optimal design process. The influence of the assumed process noise is indicated by w_1 and w_2 in the model description. We distinguish two process noise levels. In the first case, the corresponding variance–covariance matrix of the noise is given by $W_1(t) = \mathrm{diag}\left(0.005\,\mathrm{g}^2/\mathrm{L}^2\,\mathrm{s}^2, 0.005\,\mathrm{g}^2/\mathrm{L}^2\,\mathrm{s}^2\right)^\top$. In the second case we set $W_2(t) = \mathrm{diag}\left(0.05\,\mathrm{g}^2/\mathrm{L}^2\,\mathrm{s}^2, 0.05\,\mathrm{g}^2/\mathrm{L}^2\,\mathrm{s}^2\right)^\top$ assuming a higher level of process noise.

We solve the optimization problems again with a multiple-shooting approach. The number of discretization intervals is also set to 30. The integrator tolerances and the SQP routine tolerance are set to the same values as in the previous section.

We display the resulting states of the optimization in Fig. 7. Both cases taking process noise into account are compared with the nominal case of the design without process noise. All cases start with a similar amount of initial substrate concentration. The nominal case is in a batch phase and is consuming its substrate concentration until 8 h of the experiment. The cases including the process noise have a short additional feed early on in the experiment, bringing the substrate concentration to around 60 g/L. The feed rates of the process noise cases both exhibit a longer feeding phase than the A-design. It runs up to 18 h for the first process noise case (u^*_{AW1}) and up to 20 h for the second process noise case (u^*_{AW2}). The substrate concentrations is held more constant with these feed rates, u^*_{AW1} and u^*_{AW2} during the first 20 h. All three experiments have after this initial feeding phase a batch phase where the substrate concentration is consumed until no more substrate is left in the reactor. After the depletion of the substrate, a second feed phase starts which is similar for the three cases. The experiment is terminated when the maximal allowed volume is reached and all substrate is consumed. We also would like to point out that with the addition of process noise, the amount of biomass formed is decreasing for increasing levels of process noise. Similar observations were made in [27,28]. Note that the experimentation time increases for increasing levels of process noise.

In order to illustrate the evolution of the uncertainty, we depict the diagonal elements of the variance–covariance matrix for the A-design and the two process noise levels in Fig. 8. We observe that the initial uncertainty for both states, C_s and C_x, decreases over time. However, with increasing levels of process noise this decrease becomes smaller. In the case of the biomass concentration, C_x, we observe even a more or less steady state level after an initial decrease for the two process noise cases. After an initial decrease in uncertainty, there is for around 23 h in the experiment an increase in the uncertainty, which is maximal around 28–29 h. This is related to the steady consumption of substrate. For the uncertainty with respect to the parameters, we observe a steady decrease for $\mu_{\max}$. With the addition of process noise, this decrease is slower over time and does not reach the lowest uncertainty level of the A-design. The accuracy of K_s is also for the process noise cases the limiting factor in the optimal experiment design. We notice that the decrease in uncertainty for K_s only takes place in the last part of the experiment and is thus mainly influenced by the last feeding phase. We would also like to emphasize that the objective function focuses on the uncertainty in Q_{33} and Q_{44}. This explains why the uncertainty regarding the states C_s and C_x remains high throughout the experiment.

In order to illustrate the expected uncertainty, we depict the uncertainty ellipsoids of the parameters, $\mu_{\max}$ and K_s at the end of the experiment in Fig. 9 for the A-design and the two process noise cases. We observe that the A-design that does not take the process noise into account results in an uncertainty ellipsoid with the smallest area. By taking $W_1(t)$ into account, the uncertainty region is larger than in the nominal design. When increasing the process noise to the levels of $W_2(t)$, the uncertainty is even further increased. In order to measure

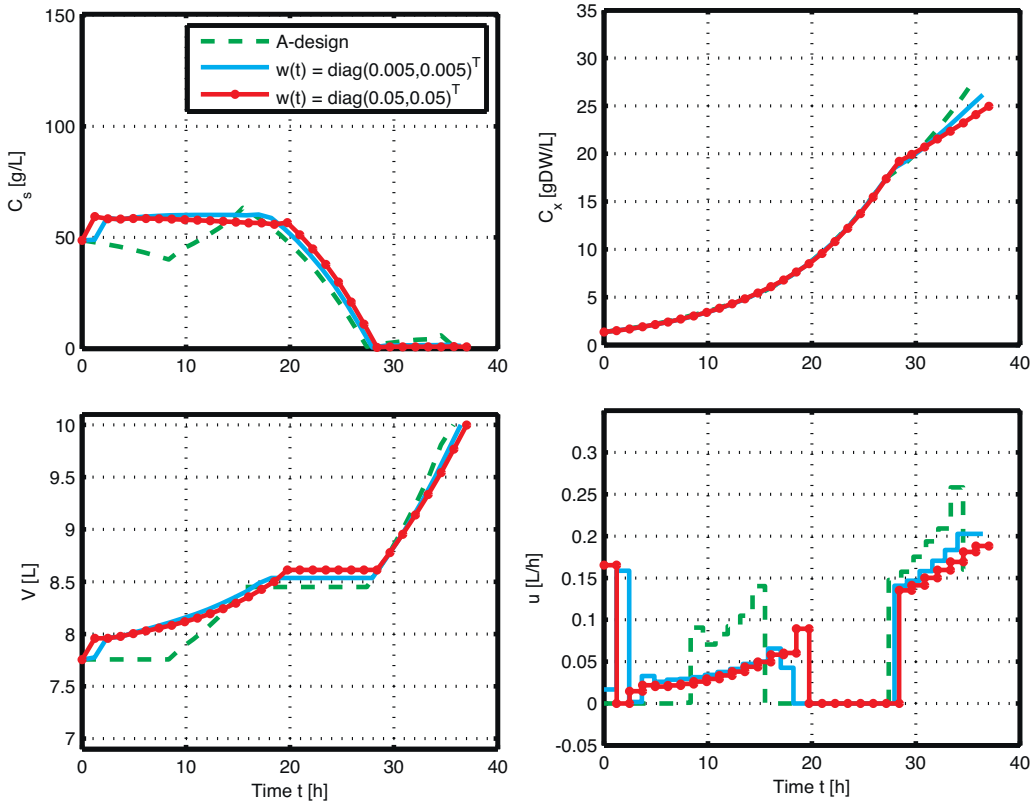

Fig. 7. Evolution of the states and the controls for the A-criterion (AVC based) approach with and without process noise added.

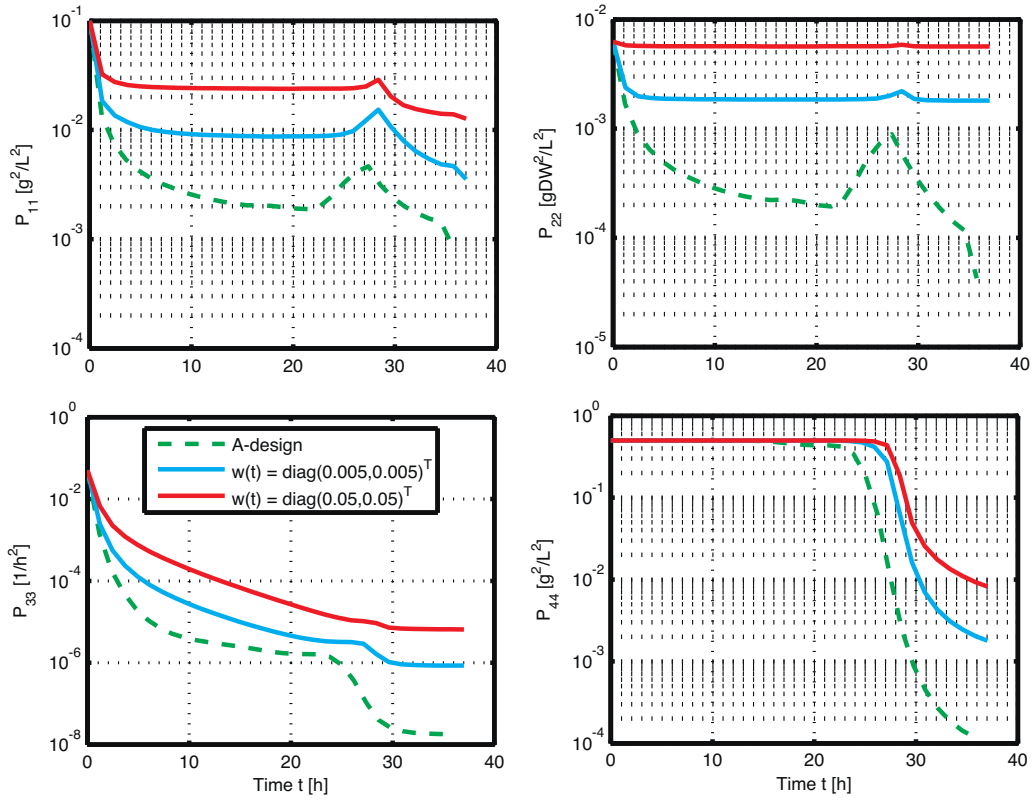

Fig. 8. Evolution of the diagonal elements of the approximate variance–covariance matrix for the experiments obtained by the A-criterion approach with and without process noise.

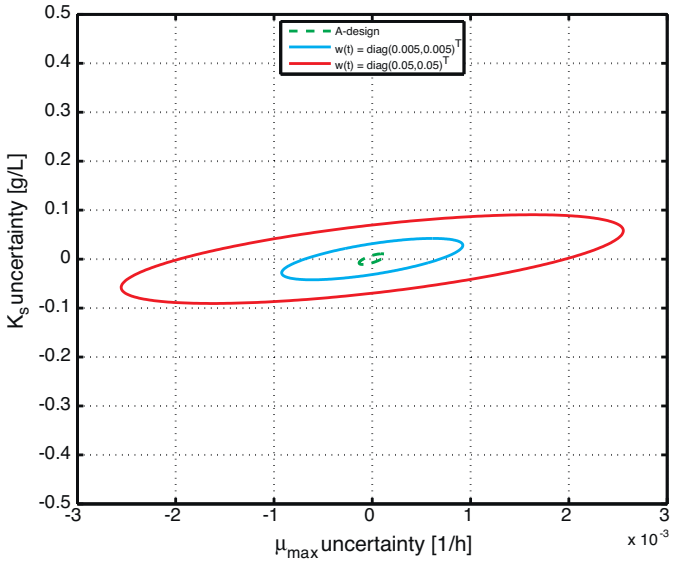

Fig. 9. Uncertainty ellipsoids as measured using the approximate variance–covariance matrix for the designed u^*_{A}, u^*_{AW1} and u^*_{AW2}.

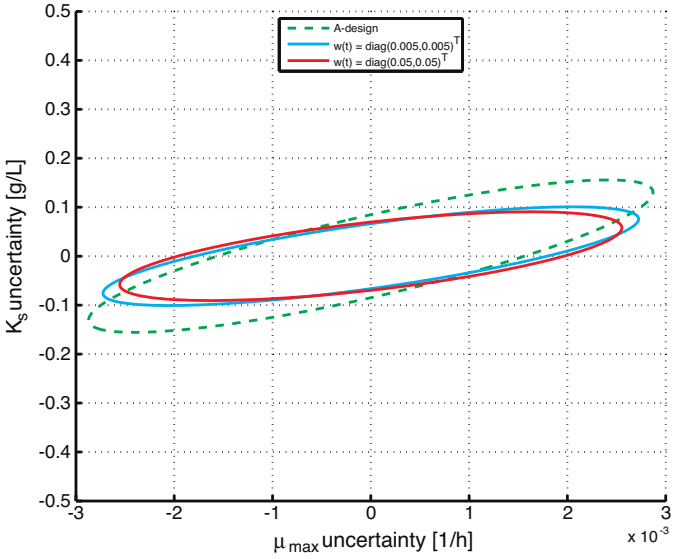

Fig. 10. Uncertainty ellipsoids as measured using the approximate variance–covariance matrix for with process noise added equal to u^*_{AW2} for the nominal A-design, u^*_{AW1} and u^*_{AW2}.

this increase of uncertainty, we calculate the total area of the expected uncertainty ellipsoid. In the case of the nominal design, this is 3.17×10^{-6}. For the first process noise level, $W_1(t)$, the total area is increased to 9.13×10^{-5} in the case of the second noise level, $W_2(t)$, this even becomes 5.59×10^{-4}. As we would expect, the uncertainty region increases as the process noise level increases, however the proposed method is able to take process noise into account. In order to investigate the effect of process noise on the experiments designed with less or no process noise, we depict Fig. 10. We add the same level of process noise to all experiments. This illustrates that the process noise has an effect on the parameter accuracy and that it is important to take this process noise into account. From Fig. 10, it is clear that the A-design leads to the largest uncertainty region while the experiment, u^*_{AW1}, has a slightly smaller uncertainty region than the A-design but still much larger than u^*_{AW2} case.

To conclude the discussion of the numerical results, we present in Table 2 the run-times. We compare the Fisher information approach with the proposed Riccati based method. Based on this specific case study we observe a decrease of 40% in time needed for each SQP step. In

Table 2
Run-time comparison for the Fisher information matrix approach and the proposed Riccati based variance–covariance matrix calculation.

	Fisher approach	Riccati approach	Riccati under process noise
Time per step [s]	2.2	1.3	1.5
QP solution [s]	2.37×10^{-1}	5.75×10^{-2}	5.69×10^{-2}
Simulation and sensitivity generation [s]	1.93	1.24	1.45

the case we add process noise, a decrease per SQP step of 30% is observed for this specific case study. If we additionally exploit the structure of the Riccati differential equations in the multiple-shooting code as explained in Theorem 3.1, we can gain another 33% of speed-up in terms of overall computation time for this case study.

4.4. Final remarks and open problems

In this paper, we have introduced a computational framework for solving optimal experiment design problems based on Riccati differential equations. Here, our problem formulation follows the classical framework of linear approximation. When applying this framework in practice, we have to be aware of some general difficulties of optimal experiment design. The aim of this section is to collect these issues and discuss possible remedies. It turns out that some of the difficulties can be addressed by using existing approaches from the literature while other problems are still open.

- Firstly, the approximation of the variance–covariance matrix which we propose to optimize, is – analogous to the Fisher information matrix – only locally correct. This aspect is especially important when computing confidence levels. For example, if we choose a very large confidence level the associated confidence ellipsoid might be larger than the region in which the linearization is correct. One way to overcome this limitation, is to regard robust optimal experiment design formulations which are for example proposed in [9,34]. These robust formulations can additionally reduce the effect of the initial guess for the states and parameters influencing the linearization point. Note that the robust optimal experiment design formulations in [9] could in principle be combined with our approaches aiming at a robustification with respect to both: the linearization point and the process noise. Other approaches which can deal with the nonlinearity effects are the bootstrap method which has been proposed in [2]. This method is based on a Monte-Carlo approach which generates a set of noisy simulations each yielding a different paraemeter estimate such that the variance-covariance matrix can be approximated after simulating a suitably large number of experiments. However, this approach is extremely expensive and hardly applicable if process noise is present as we would have to simulate a huge number of process noise scenarios in our case. Finally, we note that the sigma-point or unscented Kalman filtering methods as proposed in [36] or [17] are a heuristic but computationally tractable way to reduce the approximation error in the computation of the variance–covariance matrix.
- Secondly, another important difficulty in optimal experiment design are structural model-plant mismatches. To a certain extent these mismatches can be taken into account in our approach. For example, if we have two model right-hand sides f_1 and f_2 we may combine them by introducing an auxiliary parameter "$\theta \in [0, 1]$" using $f := \theta f_1 + (1-\theta) f_2$ as our new right-hand side while θ has to be estimated from measurements, too. However, more general model-plant mismatches – especially, those which cannot be parameterized – cannot be taken into account in the presented framework. Finally, we refer to the model discrimination methods developed in [11,38] which can be used for addressing the problem of model plant mismatch.
- And thirdly, our formulation relies on uncertainty models for both: the process noise as well as the measurement errors. More precisely, we have assumed that these errors are white-noise processes with given variance–covariance matrices $V(t)$ and $W(t)$. However, in practice finding a good uncertainty model can be as difficult as modeling the dynamic process itself. Our formulation only helps us to design an experiment for measuring the system parameters and initial states but it does not help us to identify a stochastic model of the uncertainties. While many sensors come along with a good estimate of the function V, finding a good estimate for W can be challenging.

5. Conclusions

In this work, we have discussed a Riccati differential equation based method for optimal experiment design under process noise. A main contribution of the paper is the modified reformulation of optimal experiment design problems into a standard optimal control problem which allows explicitly accounting for process noise, i.e., unmodeled inputs which affect the dynamic system. The resulting standard optimal control problem consists of fewer states than the one based on the Fisher information matrix and can efficiently be solved with existing multiple-shooting or collocation based algorithms. Additionally, we suggested to exploit the particular structure of the Riccati differential equations, as summarized in Theorem 3.1. As the computation of the variance–covariance matrix is cheaper, we have proposed to use design criteria based on this variance–covariance matrix instead of those based on the Fisher information matrix.

We have illustrated the proposed method on a nonlinear fed-batch bioreactor benchmark case-study. We observed that the method yields an informative experiment. Furthermore, we also illustrated that process noise can easily be accounted for, and that this incorporation of noise resulted -as expected- in larger uncertainty regions. When looking at the obtained optimal time-varying feed profiles, the experiments obtained under process noise exhibited a more gradual feeding compared with the A-design.

Acknowledgments

We thank the three anonymous referees for their precise and substantial remarks which significantly improved this manuscript.

D. Telen has a Ph.D. grant of the Agency for Innovation through Science and Technology in Flanders (IWT). J. Van Impe holds the chair Safety Engineering sponsored by the Belgian Chemistry and Life Sciences Federation essenscia. The research was supported by KUL, OT/10/035, via GOA/11/05 Ambiorics, GOA/10/09 MaNet, PFV/10/002 (OPTEC), IOF-SCORES4CHEM, the Flemish Government via FWO-projects and via IWT-projects, the Belgian Federal Science Policy Office: IAP VII/19 (DYSCO) and the EU (ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC HIGHWIND (259 166)).

References

[1] M. Baltes, R. Schneider, C. Sturm, M. Reuss, Optimal experimental design for parameter design in unstructured growth models, Biotechnology Progress 10 (1994) 480–488.
[2] E. Balsa-Canto, M. Rodriguez-Fernandez, J.R. Banga, Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation, Journal of Food Engineering 82 (2007) 178–188.
[3] J. Banga, K. Versyck, J. Van Impe, Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization approach, Industrial and Engineering Chemistry Research 41 (2002) 2425–2430.

[4] I. Bauer, G. Bock, S. Körkel, J. Schloder, Numerical methods for optimum experimental design in DAE systems, Journal of Computational and Applied Mathematics 120 (2000) 1–25.
[5] L. Biegler, Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation, Computers and Chemical Engineering 8 (1984) 243–248.
[6] L. Biegler, An overview of simultaneous strategies for dynamic optimization, Chemical Engineering and Processing: Process Intensification 46 (2007) 1043–1053.
[7] H. Bock, Recent advances in parameter identification techniques for ODE, in: P. Deuflhard, E. Hairer (Eds.), Numerical Treatment of Inverse Problems in Differential and Integral Equations, Birkhäuser, Boston, 1983, pp. 95–121.
[8] H. Bock, K. Plitt, A multiple shooting algorithm for direct solution of optimal control problems, in: Proceedings of the 9th IFAC World Congress, Budapest, Pergamon Press, 1984, pp. 243–247.
[9] H.G. Bock, S. Körkel, E. Kostina, J.P. Schlöder, Robustness Aspects in Parameter Estimation, Optimal Design of Experiments and Optimal Control, Springer, Berlin, 2007, pp. 117–146.
[10] G. Box, H. Lucas, Design of experiments in non-linear situations, Biometrika 46 (1959) 77–90.
[11] G. Buzzi-Ferraris, P. Forzatti, A new sequential experimental design procedure for discriminating among rival models, Chemical Engineering Science 38 (1983) 225–232.
[12] A. Cappuyns, K. Bernaerts, I. Smets, O. Ona, E. Prinsen, J. Vanderleyden, J. Van Impe, Optimal fed batch experiment design for estimation of Monod kinetics of *Azospirillum brasilense*: from theory to practice, Biotechnology Progress 23 (2007) 1074–1081.
[13] D. Espie, S. Machietto, The optimal design of dynamic experiments, A.I.Ch.E. Journal 35 (1989) 223–229.
[14] R. Fisher, The Design of Experiments, Oliver & Boyd, Edinburgh, 1935.
[15] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter precision: state of the art, Chemical Engineering Science 63 (2008) 4846–4872.
[16] M. Gevers, L. Ljung, Optimal experiment designs with respect to the intended model application, Automatica 22 (1986) 543–554.
[17] T. Heine, M. Kawohl, R. King, Derivative-free optimal experimental design, Chemical Engineering Science 63 (2008) 4873–4880.
[18] H. Hjalmarsson, From experiment design to closed-loop control, Automatica 41 (2005) 393–438.
[19] B. Houska, H. Ferreau, M. Diehl, ACADO Toolkit – an open-source framework for automatic control and dynamic optimization, Optimal Control Applications and Methods 32 (2011) 298–312.
[20] R. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME: Journal of Basic Engineering 82 (1960) 35–45.
[21] R.E. Kalman, R.S. Bucy, New results in linear filtering and prediction theory, Journal of Basic Engineering 83 (1961) 95–108.
[22] J. Kiefer, J. Wolfowitz, Optimum designs in regression problems, Annals of Mathematical Statistics 30 (1959) 271–294.
[23] S. Körkel, E. Kostina, H. Bock, J. Schlöder, Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes, Optimization Methods and Software Journal 19 (3–4) (2004) 327–338.
[24] S. Körkel, A. Potschka, H. Bock, S. Sager, A multiple shooting formulation for optimum experimental design, Mathematical Programming, submitted for publication.
[25] D. Leineweber, I. Bauer, H. Bock, J. Schlöder, An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part I: theoretical aspects, Computers and Chemical Engineering 27 (2003) 157–166.
[26] L. Ljung, System Identification: Theory for the User, Prentice Hall, Englewood Cliffs, 1999.
[27] F. Logist, B. Houska, M. Diehl, J. Van Impe, Fast pareto set generation for nonlinear optimal control problems with multiple objectives, Structural and Multidisciplinary Optimization 42 (2010) 591–603.
[28] F. Logist, B. Houska, M. Diehl, J. Van Impe, Robust multi-objective optimal control of uncertain (bio)chemical processes, Chemical Engineering Science 66 (20) (2011) 4670–4682.
[29] R. Mehra, Optimal inputs for linear system identification, IEEE Transactions on Automatic Control 19 (1974) 192–200.
[30] A. Munack, Optimal Feeding Strategy for Identification of Mono-Type Models by Fed-Batch Experiments, Elsevier, Amsterdam, 1988.
[31] A. Munack, C. Posten, Design of optimal dynamic experiments for parameter estimation., in: Proceedings of the 1989 American Control Conference, 1989, pp. 2010–2016.
[32] L. Pontryagin, V. Boltyanskiy, R. Gamkrelidze, Y. Mishchenko, The Mathematical Theory of Optimal Processes, Wiley-Interscience, New York, 1962.
[33] F. Pukelsheim, Optimal Design of Experiments, John Wiley & Sons, Inc., New York, 1993.
[34] C. Rojas, J. Welsh, G. Goodwin, A. Feuer, Robust optimal experiment design for system identification, Automatica 43 (2007) 993–1008.
[35] R. Sargent, G. Sullivan, The development of an efficient optimal control package, in: J. Stoer (Ed.), Proceedings of the 8th IFIP Conference on Optimization Techniques, Springer-Verlag, Heidelberg, 1978, pp. 158–168.
[36] R. Schenkendorf, A. Kremling, M. Mangold, Optimal experiment design with the sigma point method, Systems Biology 3 (2009) 10–23.
[37] K. Schittkowski, Experimental design tools for ordinary and differential equations, Industrial and Engineering Chemistry Research 46 (2007) 9137–9147.
[38] M. Schwaab, J.L. Monteiro, J.C. Pinto, Sequential experimental design for model discrimination. Taking into account the posterior covariance matrix of parameter uncertainties, Chemical Engineering Science 63 (2008) 2408–2419.
[39] D. Telen, F. Logist, E. Van Derlinden, I. Tack, J. Van Impe, Optimal experiment design for dynamic bioprocesses: a multi-objective approach, Chemical Engineering Science 78 (2012) 82–97.
[40] E. Van Derlinden, Quantifying microbial dynamics as a function of temperature: towards an optimal trade-off between biological and model complexity, Ph.D. Thesis, 2009.
[41] K. Versyck, J. Van Impe, Feed rate optimization for bed-batch bioreactors: from optimal process performance to optimal parameter estimation, Chemical Engineering Communications 172 (1999) 107–124.
[42] E. Walter, L. Pronzato, Qualitative and quantitative experiment design for phenomenological models – a survey, Automatica 26 (1990) 195–213.
[43] E. Walter, L. Pronzato, Identification of Parametric Models from Experimental Data, Springer, Paris, 1997.