Optimality and identification of dynamic models in systems biology: an inverse optimal control framework

Nikolaos Tsiantis^{1,2}, Eva Balsa-Canto¹ and Julio R. Banga^{1,*}

¹Bioprocess Engineering Group, Spanish National Research Council, IIM-CSIC, C/Eduardo Cabello 6, 36208 Vigo (Spain).

²Department of Chemical Engineering, University of Vigo, Vigo 36310, Spain.

*To whom correspondence should be addressed, <julio@iim.csic.es>.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Optimality principles have been used to explain many biological processes and systems. However, the functions being optimized are in general unknown a priori. Here we present an inverse optimal control (IOC) framework for modeling dynamics in systems biology. The objective is to identify the underlying optimality principle from observed time-series data and simultaneously estimate unmeasured time-dependent inputs and time-invariant model parameters. As a special case, we also consider the problem of optimal simultaneous estimation of inputs and parameters from noisy data. After presenting a general statement of the IOC problem, and discussing special cases of interest, we outline numerical strategies which are scalable and robust.

Results: We discuss the existence, relevance and implications of identifiability issues in the above problems. We present a robust computational approach based on regularized cost functions and the use of suitable direct numerical methods based on the control-vector parameterization approach. To avoid convergence to local solutions, we make use of hybrid global-local methods. We illustrate the performance and capabilities of this approach with several challenging case studies, including simulated and real data. We pay particular attention to the computational scalability of our approach (with the objective of considering large numbers of inputs and states). We provide a software implementation of both the methods and the case studies.

Availability: The code used to obtain the results reported here is available at <https://zenodo.org/record/1009541>

Contact: julio@iim.csic.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Optimality principles in biology have a long history. The optimal design principle hypothesis was formally stated by Rashevsky (1961) and later extended by Rosen (1967). This principle claims that the biological structures necessary to perform a certain function must be of maximum simplicity, and optimal regarding energy and material requirements. Similar optimality principles have been shown to agree with observations about living matter at different levels of organization, from basic molecular and cell biology phenomena, up to the levels of organs, individuals, populations and their evolution (McFarland, 1977; Smith, 1978; Schaffer, 1983; Parker *et al.*, 1990; Alexander, 1996; Popescu, 1998; Sutherland, 2005; Dekel and Alon, 2005; Todorov, 2004; Friston, 2010; de Vos *et al.*, 2013).

The general justification is that evolution by natural selection is the mechanism behind these optimality principles, a concept already present in Darwin's work.

Sutherland (2005) claims that these optimality principles allow biology to move from merely explaining patterns or mechanisms to being able to make predictions from first principles. Bialek (2015) makes the important point that optimality hypotheses should not be adopted because of aesthetic reasons, but as an approach that can be directly tested through quantitative experiments. Mathematical optimization could therefore be regarded as a fundamental research tool in bioinformatics and computational systems biology (Heinrich *et al.*, 1991; Mendes and Kell, 1998; Handl *et al.*, 2007; Banga, 2008; de Vos *et al.*, 2013; Ewald *et al.*, 2017).

Here we consider the role of optimality principles in molecular and cell biology. In particular, we focus on the problem of explaining and predicting

dynamics at those levels based on optimality considerations. In the context of biochemical pathways, pioneering studies were performed in the 1990s considering optimization in metabolic networks (Heinrich et al., 1991; Hatzimanikatis et al., 1996; Heinrich and Schuster, 1996; Heinrich et al., 1997). To our knowledge, the first attempt to explain the dynamics in metabolic networks was presented by Klipp et al. (2002) considering linear networks and the hypothesis of minimum transition time. Interestingly, the predicted pattern of gene expression dynamics was later found experimentally (Zaslaver et al., 2004). Therefore, the pioneering work of Klipp et al. (2002) is a good example of the predictive value of optimality-based methods in systems biology.

Since then, several studies (Oyarzún et al., 2009; Bartl et al., 2010) recognized that the original question considered by Klipp et al. (2002) can be naturally formulated as an optimal control problem. Optimal control (also known as dynamic optimization when the system is open loop) considers the optimization of a dynamic system, that is, one seeks the optimal time-dependent input (open loop control) to minimize a certain performance index (cost function). In recent years, a number of studies have extended these ideas considering more realistic pathways (larger size and more complex topologies), and in some cases making use of increasingly sophisticated optimal control formulations (Oyarzún, 2011; Wessely et al., 2011; Bartl et al., 2013; de Hijas-Liste et al., 2014, 2015; Waldherr et al., 2015; Nimmegeers et al., 2016; Ewald et al., 2017). Similarly, Giordano et al. (2016) have recently used optimal control to obtain novel insights into microbial growth strategies, successfully explaining the dynamical allocation of cellular resources. Interestingly, these authors have also shown that a near-optimal control strategy based on a coarse-grained model exhibited structural similarities with the regulation of ribosomal protein synthesis in *E. coli*.

These studies reflect the potential benefits and the increasing importance of applying control engineering concepts to systems and synthetic biology problems, an area that is still in its infancy (Doyle and Stelling, 2006; Wellstead et al., 2008; He et al., 2016). However, an open question in all these studies is that the cost function (i.e. the performance index being optimized) is unknown a priori. Although multiobjective dynamic optimization can be used to consider several objectives simultaneously (de Hijas-Liste et al., 2014), researchers still have to come up with cost functions describing the hypothetical underlying optimality principle.

It is important to note that these studies should be distinguished from the more standard application of optimal control to biosystems and bioprocesses. This is an area where a lot of research has been carried out, and it includes well known problems as e.g. how to operate a bioreactor to ensure optimal production of a substance of interest (Smets et al., 2004), or how to create or inhibit oscillations or patterns (Lebiedz and Brandt-Pollmann, 2003; Sootla et al., 2016). All these problems have pre-defined performance indexes, which are related to the optimal design and manipulation of biosystems by humans, i.e. the cost function is known a priori.

Here we consider the use of an inverse optimal control perspective to find the optimality principle which can explain (and/or predict) dynamic behavior in biological systems. That is, we face the situation where the cost function is not known a priori. The usual approach so far has been to hypothesize and then test different cost functions for a given biosystem, i.e. computing the optimal control and checking, a posteriori, if there is a match between the predicted and the observed dynamics.

With the aim of finding a more systematic approach, we present here an inverse optimal control framework. Briefly speaking, inverse optimal control aims to find the optimality criteria that can explain a set of given dynamic measurements. In more detail, given a dynamic model of a biosystem with unknown parameters and time-series observations (measurements) of (at least part of) its dynamic states and (some or none) of the inputs, we seek to find the underlying optimality criteria (cost function, or performance index) together with the unmeasured inputs and

Fig. 1. Overall workflow of the solution strategy for Inverse Optimal Control problems

time-invariant model parameters that explain the measurements in the best way. This problem can also be regarded as estimating parameters in optimal control problems, and a number of important variants can be derived depending on what is measured.

2 METHODS

2.1 Approach

We consider two classes of inverse optimal control problems:

- IOCP-1 - input reconstruction as an optimal tracking problem: given a nonlinear dynamic model of a biosystem and a set of measurements (time-series data of observed states), find the time-dependent inputs and time-invariant parameters that explain the available time-series data.
- IOCP-2 - general inverse optimal control problem: given a nonlinear dynamic model of a biosystem operating with an underlying optimality principle, find the cost function, the time-dependent inputs and time-invariant parameters that explain the available data (time-series measurements).

Problem IOCP-1 can be regarded as a special case of problem IOCP-2. It should be noted that several other variants of IOCP-1 are possible, depending for example if the system is partially or fully observed. A few instances of this input reconstruction problem have been recently studied in the domains of systems biology (Kahm et al., 2012; Schelker et al., 2012; Kaschek et al., 2016; Lang and Stelling, 2016) and pharmacokinetics (Trägårdh et al., 2016, 2017). Particularly interesting variants of IOCP-1 have been presented by Engelhardt et al. (2016, 2017) to model biosystems with unknown inputs as well as missing and erroneous interactions. Here we are considering scenarios where all relevant inputs are known and there are no wrong interactions, showing that even under these assumptions the inverse problems can be extremely challenging due to lack of identifiability.

To the best of our knowledge, problem IOCP-2 has not been previously studied in systems biology, but it has recently received attention in the areas of robotics (Panchea and Ramdan, 2015; Clever et al., 2016) and biomechanics (Hatz et al., 2012; Hatz, 2014; Mombaun, 2016).

Next, we present the general problem statements, followed by a section detailing numerical strategies to solve them. We also discuss several important issues related to the structural identifiability of these problems.

2.2 Statement of the inverse optimal control problem

Here we consider dynamic models of biological systems given by sets of deterministic nonlinear ordinary differential equations (ODEs). The systems are assumed to be at least partially observable, i.e. we can obtain

time series measurements of at least some of the states. Here we do not consider the scenarios of errors in the model or unknown inputs.

For the more general problem (IOCP-2), we assume that there is optimality with respect to certain unknown criteria from a finite set of possibilities. Our aim is to find both the criteria and the model parameters that best explain the observed data. This inverse optimal control problem is formulated as a bi-level optimal control structure, consisting of an upper (or outer) and a lower (or inner) optimization problem. The outer problem corresponds to a simultaneous input and parameter estimation problem (as detailed in the Supplementary Information), where the time-invariant model parameters $\mathbf{p}$ and time-dependent inputs (e.g. stimuli) $\mathbf{u}(t)$ are computed so the model predictions $\mathbf{y}$ fit the observed data $\tilde{\mathbf{y}}$. The inner problem corresponds to a multi-objective optimal control problem (see also Supplementary Information), where the optimal solution of the time-dependent inputs $\mathbf{u}(t)$ is computed as Pareto optimal controls for the set of optimality criteria considered. The general mathematical formulation of the inverse optimal control problem IOCP-2 considering free terminal time becomes:

$$\min_{\mathbf{u}(t), \mathbf{p}, t_f} J_{\text{outer}}[\mathbf{x}, \mathbf{u}, \mathbf{p}] \quad (1)$$

Subject to:

$$\min_{\mathbf{u}(t), t_f} \mathbf{J}_{\text{inner}}[\mathbf{x}, \mathbf{u}, \mathbf{p}] \quad (2)$$

Subject to:

$$\frac{d\mathbf{x}}{dt} = \tilde{\Psi}[\mathbf{x}\{t, \mathbf{p}\}, \mathbf{u}\{t, \mathbf{p}\}, t], \quad (3)$$

$$\mathbf{x}\{t_0, \mathbf{p}\} = \mathbf{x}_0$$

$$\mathbf{y}(\mathbf{x}, \mathbf{p}) = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) \quad (4)$$

$$\boldsymbol{\eta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] = 0 \quad (5)$$

$$\boldsymbol{\zeta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] \leq 0 \quad (6)$$

$$\boldsymbol{\eta}_\ell[\mathbf{x}(t_\ell, \mathbf{p}), \mathbf{u}(t_\ell), \mathbf{p}] = 0 \quad (7)$$

$$\boldsymbol{\zeta}_\ell[\mathbf{x}(t_\ell, \mathbf{p}), \mathbf{u}(t_\ell), \mathbf{p}] \leq 0 \quad (8)$$

$$\mathbf{u}^L \leq \mathbf{u}(t) \leq \mathbf{u}^U \quad (9)$$

$$\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U \quad (10)$$

where:

$$J_{\text{outer}} = \sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_s} w_{ijk} (y_{ijk}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{p}) - \tilde{y}_{ijk})^2 \quad (11)$$

$$\mathbf{J}_{\text{inner}}[\mathbf{x}, \mathbf{u}, \mathbf{p}] = [J_1, J_2, \dots, J_N]^T \quad (12)$$

For J_i , ($i \in [1, N]$):

$$J_i[\mathbf{x}, \mathbf{u}, \mathbf{p}] = \Phi_M^i[\mathbf{x}(t_f, \mathbf{p}), \mathbf{p}] + \int_{t_0}^{t_f} \Phi_L^i[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] \quad (13)$$

where $\tilde{\Psi}$ is the right-hand side of the system's dynamics $\frac{d\mathbf{x}}{dt}$ (here we are assuming deterministic nonlinear dynamics given by ordinary differential equations) and equations (5)-(8) are algebraic constraints, including path equalities and inequalities ($\boldsymbol{\eta}$, $\boldsymbol{\zeta}$), and time-point constraints ($\boldsymbol{\eta}_\ell$, $\boldsymbol{\zeta}_\ell$). Equations (9) and (10) correspond to the bounds on the controls (time-dependent inputs, stimuli) ($\mathbf{u}$) and time-invariant parameters ($\mathbf{p}$) respectively. The system observables ($\mathbf{y}$) are defined through an observation functional of states and parameters $\mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$. The upper cost function J_{outer} can

take form of weighted (w_{ijk}) least squares (as above), but alternatively can be described by a maximum likelihood function. The inner problem's cost functional is a multi-objective functional considering the set of N possible optimality criteria to explore. In its general form, each objective function J_i in this set ($i \in [1, N]$) consists of a Mayer (Φ_M^i) and a Lagrange (Φ_L^i) term.

Problem IOCP-1 (input estimation) can be considered as a special case of the formulation above where the lower level optimal control problem is ignored. In other words, it results in a standard optimal control problem where one seeks the optimal input and model parameters that best explain the observed data. This formulation is very general and flexible, allowing us to consider problems with arbitrary inputs, multi-experimental settings and multiple criteria for the optimality hypothesis.

2.3 Solution strategy

The solution of the above bi-level formulation is extremely challenging. Here we transform this hierarchical inverse optimal control problem IOCP-2 into a sequence of two optimal control problems with a subsequent identification of the optimality principle using the set of Pareto optimal controls compatible with the system dynamics and the observed data. In other words, we solve the above bi-level formulation by decomposing it as follows:

- Step 1: we first solve the outer input estimation problem to obtain the time-dependent inputs $\mathbf{u}^*(t)$ and time-invariant parameters $\mathbf{p}^*$ which explain the data in terms of the outer cost function metric. This is a standard single-objective nonlinear optimal control problem. Note that parameters need to be structurally identifiable in order to ensure a unique solution (see further discussion about identifiability below).
- Step 2: we then use parameters $\mathbf{p}^*$ obtained in the previous step to compute the Pareto set of optimal controls (inner problem). Note that this is a multi-objective nonlinear optimal control problem.
- Step 3: finally, we find the solution located in the Pareto set which is consistent with the experimental data.

The corresponding workflow is shown in Figure 1. Note that for the case of IOCP-1, only step 1 is necessary.

For the sake of completeness, the following considerations and numerical details were used to implement the above steps in an efficient way:

- the optimal control problem in step 1 is a standard nonlinear dynamic optimization problem. In the case of the multi-objective optimal control problem of step 2, we use an epsilon-constraint decomposition to transform it into a set of single-objective optimal control problems. Thus both steps result in a set of nonlinear optimal control problems.
- these nonlinear optimal control problems are solved using the control vector parameterization approach, CVP (Vassiliadis *et al.*, 1994). The resulting non-linear programming problems are typically multimodal (Banga *et al.*, 2005) so we use a hybrid global-local optimization method combining the enhanced Scatter Search (eSS) and efficient local solvers (Egea *et al.*, 2009), as implemented in AMIGO2 (Balsa-Canto *et al.*, 2016). More details are given in the Supplementary Information.
- due to the ill-conditioning of these problems, regularization terms and/or constraints are used in the cost functions, as detailed below.
- it should be noted that our methodology can handle non-smooth or discontinuous dynamics thanks to the use of the global optimizer. This is illustrated with some of the examples below which have inequality path constraints that behave as state events.

Regularization of cost functions

Regularization methods can be successfully used to reduce ill-conditioning and avoid over-fitting (Gábor and Banga, 2015). Here we make use of regularized cost functions in order to reduce the ill-conditioning of the IOCPs. Thus, we extend the objective functions by adding regularization terms to incorporate problem-specific prior knowledge and/or constraints. In particular, we have implemented (i) a Tikhonov regularization scheme to incorporate possible prior knowledge about inputs and parameters, expressed as reference sets ($\mathbf{u}_{ref}, \mathbf{p}_{ref}$); (ii) a penalty on the second order derivative of the estimated inputs to avoid wild oscillations and ensure convexity in the resulting input profiles. The resulting regularization terms are given by:

$$J_{outer}^{reg} = J_{outer} + R_u + R_p + R_{d2} \quad (14)$$

$$R_u = \alpha \cdot \frac{\sum_{i=1}^{n_u} (\mathbf{u}_i(t) - \mathbf{u}_{i,ref}(t))^2}{n_u \cdot t_f} \quad (15)$$

$$R_p = \beta \cdot \frac{\sum_{j=1}^{n_p} (p_j - p_{j,ref})^2}{n_p} \quad (16)$$

$$R_{d2} = \omega \cdot \sum_{i=1}^{n_u} \int_0^{t_f} \frac{d^2 \mathbf{u}_i(t)}{dt^2} dt \quad (17)$$

where J_{outer} is the cost function, that is considered to be regularized, and α , β and ω are the regularization parameters, i.e. the weights balancing prior knowledge and information in the data. Although a number of regularization tuning methods have been developed to choose these weights, for the general nonlinear case this tuning remains an open question (Gábor and Banga, 2015). Here, we have adopted an iterative re-optimization scheme, similar to that of the L-curve method, where these weights are reduced in each re-optimization. As shown in the case studies below, we have found that these regularization terms can be very helpful for the incorporation of prior knowledge regarding the qualitative behavior of the inputs. For instance, we might want to avoid wild oscillations in the recovered inputs, or to impose a certain behavior compatible with the biochemical mechanisms (e.g. a decreasing profile for any species being degraded).

Implementation

The code implementing the case studies reported here is available at <https://zenodo.org/record/1009541>. This code makes use of the above numerical strategy, which has been implemented as a software (add-on) module for the AMIGO2 toolbox (Balsa-Canto et al., 2016). This new Inverse Optimal Control (IOC) add-on is available at https://sites.google.com/site/amigo2toolbox/home/amigo2_ioc. In all the computations reported here, we have used the eSS (enhanced Scatter Search) optimizer, a hybrid global-local optimization meta-heuristic available in AMIGO2.

2.4 Remarks about identifiability

Identifiability considers the possibility of uniquely inferring model parameters from experimental data (Walter and Pronzato, 1997). A model is structurally identifiable if it is theoretically possible to estimate its parameters under the ideal scenario of continuous noise free data. A model is identifiable in practice if (i) it is structurally identifiable, and (ii) we have sufficiently rich experimental data to achieve high-quality parameter estimates.

This fundamental question becomes even more relevant in the case of IOCPs, where we seek the simultaneous identification of parameters and time-dependent inputs. The latter will in general imply the addition of

Fig. 2. Illustrative example of lack of structural identifiability in inverse optimal control problems (IOCP). The figure shows a forward problem where two different input profiles ($k_u = 1$ and $k_u = 0.5$, with $u(t) = k_u \exp(-0.1t)$) result in the same output y . Thus, if we attempt to solve the inverse problem using observations of this output y , it will not be possible to uniquely recover the input, even for the ideal case of continuous noise-free measurements of the output.

further degrees of freedom, thus inducing lack of structural identifiability. We illustrate this in Figure 2 with a simple example regarding the reaction $X_1 \xrightarrow{u, k_1} X_2 \xrightarrow{k_2}$ which involves an inducer concentration, u . The reactant X_1 is present in excess, therefore we assume that its concentration x_1 is known and constant throughout the process. We can observe x_2 dynamics. The production and degradation rates have kinetic parameters k_1 and k_2 . We considered a simple case in which u follows an exponential decay, $u(t) = k_u e^{-k_t t}$. Figure 2 illustrates how different combinations of k_1 and k_u produce exactly the same x_2 profile for the same value of the product $k_1 k_u$. In other words, for this simple example, it will not be possible to individually identify k_1 and k_u but only their product. Other illustrative examples can be found in the Supplementary Information.

Assessing the structural identifiability of a nonlinear dynamic model can be a rather complicated task. The analysis requires symbolic manipulation, which may become computationally prohibitive (Chis et al., 2011b). Trägårdh et al. (2017) is one of the few studies that has considered identifiability in an input estimation problem, analyzing parametric and stimuli structural identifiability separately. Here we have considered alternative possibilities to assess the identifiability of the IOCP formulation presented above. Since the structural identifiability of the general IOCP is (to the best of our knowledge) an open question, we have considered special cases where we could exploit the possibilities offered by methods developed in the context of parameter estimation (Chis et al., 2011b). For example, a Taylor series approach can be applied by either using time dependent control parameterizations, or by assuming that the successive time derivatives of the stimuli (as evaluated at $t = 0^+$) are unknown parameters to be considered simultaneously alongside the model parameters $\mathbf{p}$. Also, we can apply the generating series approach as implemented in GenSSI (Chis et al., 2011a), provided the stimuli are parametrized using continuous bounded functions of time. We illustrate the use of both approaches in the Supplementary Information.

2.5 Remarks about the reconstruction of the optimality principles

Although optimality principles have been studied at different levels of organization of biological systems, the case of metabolic networks deserves special attention. Heinrich et al. (1991) reviewed this topic in the early 1990s, stating several key ideas. First, the structural design and functioning

of these networks are the outcome of evolution, which can be considered an optimization process. Second, the optimality assumption for these systems is supported by the fact that experimental perturbations (e.g. mutations) typically result in worse function. Third, the cost functions more frequently used were (i) maximization of fluxes, (ii) minimization of the concentration of metabolic intermediates, (iii) minimization of transient times, (iv) maximization of the sensitivity to external signals, and (v) maximization of thermodynamic efficiencies. It should be noted that at that time most studies were restricted to stationary states.

Later, Klipp *et al.* (2002) explicitly considered dynamics and used the concept of minimum transition time to derive optimal temporal gene expression profiles, based on the assumption that cells have developed optimal adaptation strategies as an evolutionary response to changing environmental conditions. Notably, these predictions were experimentally confirmed by Zaslaver *et al.* (2004) for linear pathways, and agreed with earlier experimental observations. A detailed account can be found in Bartl *et al.* (2010), who make the case for a closely related cost function, the minimization of the operation time of substrate to product conversion, with an upper bound on the total amount of enzymes at any given time. These authors show how this alternative cost produces the same temporal pattern in linear pathways (just-in-time expression), while still being consistent with the assumption that high fitness requires high flux.

Although the vast majority of studies in this field have considered single cost functions, it is interesting to realize that Heinrich *et al.* (1991) had already suggested the use of multicriteria optimality, a concept further motivated in their later book (Heinrich and Schuster, 1996). The motivation was based on the idea that appropriate objective functions must be used to evaluate the efficiency (fitness) of a given system, and that it is very unlikely that the structure and dynamic behavior of current metabolic networks can be explained on the basis of a single objective, especially considering their many biological functions and the interactions between them. To the best of our knowledge, de Hijas-Liste *et al.* (2014) were the first to use multicriteria optimal control in the context of metabolic pathways, predicting the full set of best trade-offs (known as the Pareto set) between conflicting objectives.

Here we adopt this multicriteria view for the inverse problem of finding the underlying optimality principles. The method presented above for the general IOCP-2 problem starts from a list of candidate cost functions, and aims to find the point in the Pareto set which agrees with the experimental data and where the reconstructed inputs agree with the observed ones (although the latter are not used in the solution of the inverse problem).

3 RESULTS

We present results for several challenging case studies. We first consider a problem of input estimation (IOCP-1). A second, more challenging problem of the same class is presented in the Supplementary Information to illustrate the scalability of the method regarding the IOCP-1 class. Next we present two instances of the more general inverse optimal control problem (IOCP-2). Simulated data (i.e. pseudo experimental data generated by simulation with different noise levels) are used in order to better assess the methodology. In these scenarios using simulated data, the true input functions (and optimality criteria, in the case of IOCP-2) are known and therefore we can carefully analyze the performance and accuracy of the method. However, it should be noted that in two cases we also consider scenarios with real data. The goodness of fit is evaluated using the normalized root mean square error (NRMSE; more details and other metrics in the Supplementary Information). All the computations were carried out on a PC Intel Xeon E5-2630@2.30GHz using Matlab R2015B under Windows 7.

Fig. 3. Case study JAK-STAT: estimated versus measured input (*pEpoR*) profiles, along with the fit for the two observables (y_1 and y_2). The lower left subfigure shows the goodness of the estimation as measured by the normalized root mean square error (NRMSE) with respect to the observables (y) and the input (u).

3.1 Case study JAK-STAT: input reconstruction for the JAK2-STAT5 signaling pathway

This case study is an example of IOCP-1, based on the problem considered by Schelker *et al.* (2012), where we seek the simultaneous input reconstruction and parameter estimation in a dynamic model of the JAK2-STAT5 signaling pathway. The detailed mathematical statement is given in the Supplementary Information. The problem has one time-dependent stimulus (*pEpoR*, which is treated as unknown here), nine internal state variables and two observables. Real measurements are available for both the observables and the stimulus. Our aim here is to simultaneously estimate the model parameters and the time-dependent trajectory for the input (stimulus), using only the experimental data for the observables in the cost function. The estimated input is then compared with the (unused) measurements of *pEpoR*. In other words, we only use the *pEpoR* measurements for the a posteriori analysis of the goodness of the input estimation.

This example is well suited to illustrate the importance of identifiability analysis in this class of problems. We provide a detailed structural identifiability analysis in the Supplementary Information, revealing that it is not possible to uniquely and simultaneously estimate all model parameters and *pEpoR*(t). In particular, p_1 , p_4 and the scaling parameters are structurally non-identifiable. Based on this, we considered the problem of estimating p_{2-3} , the offsets and *pEpoR*(t). Scaling parameters were fixed to nominal values. A piecewise linear approximation for the input *pEpoR*(t) was considered. Based on the nature of the stimulus, we used regularization to avoid wild oscillatory behavior and over-fitting (i.e. fitting the noise instead of the signal). The results are depicted in Figure 3, showing a good agreement (NRMSE=0.142) between the estimated input and the input measurements (not used in our method), and a very good fit of the model predictions to the observed states measurements (NRMSE=0.084). Computation times were in the range of 1 – 4 minutes.

In the Supplementary Information, we include detailed results for two additional scenarios: (i) a subcase considering the ideal scenario of synthetic noiseless data, for which we achieve almost perfect estimations for both inputs and parameters, and (ii) a subcase using synthetic data generated with the addition of 5% heteroscedastic proportional noise.

Fig. 4. Case study LPN3B10: Pareto front (optimal trade-offs for the cost functions) indicated the reference point selected for the inverse optimal control (IOCP-2) problem. The good agreement between estimated and true inputs are shown in the figure inset. The error of the estimation is $NRMSE_y = 0.13$ and $NRMSE_u = 0.14$.

3.2 Case study LPN3B: linear pathway with three enzymatic reactions

This is a first example of the more general IOCP-2 problem class, where we seek to identify the underlying optimality criteria along with the time-dependent stimuli and the time-invariant parameters considering a metabolic pathway model. This problem is a generalization of the dynamic optimization problem considered by [de Hijas-Liste et al. \(2014\)](#), based on the study by [Bartl et al. \(2010\)](#). In brief, we consider a linear chain of enzymatic reactions where the hypothesis is that the time-dependent enzyme concentrations correspond to an optimal trade-off between minimum conversion time and minimum accumulation of intermediates. The corresponding Pareto front is shown in Figure 4. We take the upper left (circled) point as reference, and generate pseudo-experimental data for those conditions. Then we consider the inverse problem of recovering the inputs and the optimality trade-off from the data. From an optimal control perspective, this case study is computationally challenging due to its multimodality and the presence of algebraic path constraints on the states and the stimuli. We have assumed that the system is fully observed, and we generated pseudo-experimental data for two subcases: LPN3B0 (noiseless data) and LPN3B10 (data with 10% heteroscedastic proportional noise). The detailed mathematical statement is given in the Supplementary Information.

First, we performed a structural identifiability analysis, revealing that it is not possible to simultaneously and uniquely estimate the kinetic constants and the time-dependent enzyme profiles. We also confirmed this property numerically, i.e. we were able to obtain different combinations of enzyme activation profiles and kinetic parameters which result in the same good fit to the data (full details of these identifiability issues are given in the Supplementary Information).

Next, in order to handle this lack of structural identifiability, we used prior knowledge so as to enforce a unique numerical solution. This prior information was embedded into the regularization scheme and the selected bounds. We also considered several types of control parameterizations for the inputs (enzyme activation profiles). We achieved a very good reconstruction of the original stimuli trajectories and the parameter values in both subcases. The goodness of the input estimation for the noisy subcase is given in Figure 4, showing a recovered just-in-time activation profile in

good agreement with the true one (not used in the reconstruction). Note that the corresponding NRMSE of 14% is due to the small shifts in the switching times. Computation times were in the range of 1 – 20 minutes. Further details for both subcases can be found in the Supplementary Information.

3.3 Case study SC: central metabolism of *Saccharomyces cerevisiae* during diauxic shift

As a second example of problem IOCP-2, we considered the dynamics of the central metabolism of yeast during diauxic shift in a nutrient depletion scenario. Our formulation is based on the problems studied by [Klipp et al. \(2002\)](#) and [de Hijas-Liste et al. \(2014\)](#), where the model considers six reactions: upper and lower glycolysis, ethanol formation and ethanol consumption, the TCA cycle and the respiratory chain. This description results in 9 dynamic states, 8 parameters and 6 time-dependent enzyme concentrations. In the previous study by [de Hijas-Liste et al. \(2014\)](#), the objective was to find the enzyme activation profiles so as to compute the Pareto set of optimal control considering two objective functions (survival time and enzyme synthesis cost). A path inequality constraint is used to ensure that concentrations of NADH and ATP remain above critical values at all times. In other words, Pareto-optimal enzyme profiles were computed which ensure long-term homeostasis of key metabolites under conditions of a diauxic shift.

Here, we formulate a synthetic problem by selecting a reference point in the Pareto front and considering the full inverse optimal control problem: given synthetic time-series data for the observed states at such a reference point, estimate the model parameters, the inputs (enzyme time-dependent activation profiles) and the optimality criteria which best explain such data.

This case study is more complex than LPN3B, including more time-dependent stimuli and several challenging path constraints on the concentrations of NADH and ATP. The detailed mathematical formulation and a diagram of the network are given in the Supplementary Information. We considered two different subcases based on the generated synthetic data: SC0 (noiseless) and SC10 (noisy, with 10% heteroscedastic proportional noise). Note that this case study is particularly interesting because real experimental data are also available, allowing qualitative comparisons ([Klipp et al. \(2002\)](#)).

As expected, the problem is very ill-conditioned and exhibits identifiability issues of a similar nature to those in [3.2](#). The controls e_i appear only multiplying the parameters, making both of them structurally unidentifiable. Therefore, to reduce the identifiability issues, we re-formulated the model, estimating the products of parameters and stimuli (details in the Supplementary Information). This re-formulated problem has 9 dynamic states, 2 parameters and 6 time-dependent products of enzyme concentrations and kinetic parameters which we consider as our new inputs.

The results for the noiseless subcase SC0 are given in the Supplementary Information, showing a very good reconstruction of the inputs ($NRMSE_u = 0.2$) and an almost perfect fit to the data ($NRMSE_y = 0.001$). The more realistic SC10 subcase was also successfully solved, showing that despite the 10% noise, our methodology was able to find a very reasonable reconstruction of the inputs ($NRMSE_u = 0.29$), and therefore a good identification of the optimality trade-off explaining the dynamics. In Figure 5 we show the matching between the estimated and the true input values, the corresponding fit of the estimated states to the data and the NRMSE values. Computation times were in the range of 15 – 90 minutes.

It is worth discussing the biological meaning of the estimated states and inputs. At early times, enzyme e_1 shows a peak of activity which allows the rapid consumption of glucose. Enzyme e_2 and e_3 also present significant activity at early times, resulting in the production of ethanol

Fig. 5. Case study SC: estimated versus true input values (left; NRMSE=0.29), and the corresponding fit to observed data (right; NRMSE=0.13), for the SC10 (10% noise) subcase.

while glucose is still present. Later on, e_4 increases its activity, so ethanol is consumed to maintain the required levels of NADH and ATP. These results are consistent with the observed behavior of diauxic shift of yeast under glucose starvation conditions. An additional qualitative comparison with experimental data is given in the Supplementary Information.

4 CONCLUSIONS

Here we consider the question of systematically inferring optimality principles in cellular pathways from noisy dynamic data. We present a general inverse optimal control (IOC) framework to address this question, where the objective is to simultaneously identify the underlying optimality principle, as well as the time-dependent inputs and time-invariant model parameters. We present a solution strategy where the IOCP is decomposed into an optimal tracking problem and a multi-criteria optimal control problem. As a special case, we also consider the subproblem of input and parameter estimation from noisy data.

Importantly, we discuss the existence and implications (in terms of fundamental limitations) of structural identifiability issues in the above problems. Further, we present a robust computational solution strategy based on regularized cost functions and the use of suitable direct numerical methods based on the control-vector parameterization approach. To avoid convergence to local solutions, we make use of hybrid global-local optimization methods.

We illustrate the performance of this framework by successfully solving several challenging case studies dealing with biochemical pathways of increasing size and complexity, including scenarios with simulated and real data. We highlight the importance of analyzing the structural identifiability issues present in these problems. We show how the limitations arising from these issues can be surmounted by model reformulation and/or the exploitation of prior knowledge. Similarly, practical identifiability issues (due to lack of information in the data) could be surmounted by incorporating more informative experiments and/or by performing model-reduction. Finally, we find that the methodology exhibits a satisfactory scalability in terms of computational cost as a function of problem size (details are provided in the Supplementary Information). It should be noted that large speedups could be obtained by exploiting the inherent parallelism of the methodology.

Funding

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 675585 (MSCA ITN "SyMBioSys") and from the Spanish MINECO/FEDER projects SYNBIOfACTORY (grant number DPI2014-55276-C5-2-R), SYN-BIOCONTROL (DPI2017-82896-C2-2-R) and IMPROWINE (grant number AGL2015-67504-C3-2-R). Author NT is a MSCA ESR at IIM-CSIC (Spain).

References

- Alexander, R. M. (1996). *Optima for animals*. Princeton University Press.
- Balsa-Canto, E. *et al.* (2016). Amigo2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*, **32**(21), 3357–3359.
- Banga, J. R. (2008). Optimization in computational systems biology. *BMC Syst. Biol.*, **2**(1), 47.
- Banga, J. R. *et al.* (2005). Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. *J. Biotechnol.*, **117**(4), 407–419.
- Bartl, M. *et al.* (2010). Modelling the optimal timing in metabolic pathway activation—use of pontryagin's maximum principle and role of the golden section. *Biosystems*, **101**(1), 67–77.
- Bartl, M. *et al.* (2013). Dynamic optimization identifies optimal programmes for pathway regulation in prokaryotes. *Nat. Commun.*, **4**.
- Bialek, W. (2015). Perspectives on theory at the interface of physics and biology. *arXiv preprint arXiv:1512.08954*.
- Chis, O. *et al.* (2011a). GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics*, **27**(18), 2610–2611.
- Chis, O. *et al.* (2011b). Structural identifiability of systems biology models: A critical comparison of methods. *PloS One*, **6**(11), e27755.
- Clever, D. *et al.* (2016). Inverse optimal control based identification of optimality criteria in whole-body human walking on level ground. In *IEEE EMBS Conf. Biomed. Robot. Biomech.*, pages 1192–1199.
- de Hijas-Liste, G. M. *et al.* (2014). Global dynamic optimization approach to predict activation in metabolic pathways. *BMC Syst. Biol.*, **8**(1), 1.
- de Hijas-Liste, G. M. *et al.* (2015). Optimal programs of pathway control: dissecting the influence of pathway topology and feedback inhibition on pathway regulation. *BMC Bioinf.*, **16**(1), 163.

- de Vos, M. G. *et al.* (2013). Optimality in evolution: new insights from synthetic biology. *Curr. Opin. Biotechnol.*, **24**(4), 797–802.
- Dekel, E. and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**(7050), 588.
- Doyle, F. J. and Stelling, J. (2006). Systems interface biology. *J. R. Soc. Interface*, **3**(10), 603–616.
- Egea, J. A. *et al.* (2009). Dynamic optimization of nonlinear processes with an enhanced scatter search method. *Ind. Eng. Chem. Res.*, **48**(9), 4388–4401.
- Engelhardt, B. *et al.* (2016). Learning (from) the errors of a systems biology model. *Sci. Rep.*, **6**.
- Engelhardt, B. *et al.* (2017). A bayesian approach to estimating hidden variables as well as missing and wrong molecular interactions in ordinary differential equation-based mathematical models. *J. R. Soc. Interface*, **14**(131), 20170332.
- Ewald, J. *et al.* (2017). Deciphering the regulation of metabolism with dynamic optimization: an overview of recent advances. *Biochem. Soc. Trans.*, **45**(4), 1035–1043.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.*, **11**(2), 127–138.
- Gábor, A. and Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst. Biol.*, **9**(1), 74.
- Giordano, N. *et al.* (2016). Dynamical allocation of cellular resources as an optimal control problem: novel insights into microbial growth strategies. *PLoS Comput. Biol.*, **12**(3), e1004802.
- Handl, J. *et al.* (2007). Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **4**(2).
- Hatz, K. (2014). *Efficient numerical methods for hierarchical dynamic optimization with application to cerebral palsy gait modeling*. Ph.D. thesis, University of Heidelberg.
- Hatz, K. *et al.* (2012). Estimating parameters in optimal control problems. *SIAM J. Sci. Comput.*, **34**(3), A1707–A1728.
- Hatzimanikatis, V. *et al.* (1996). Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.*, **42**(5), 1277–1292.
- He, F. *et al.* (2016). Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering. *J. R. Soc. Interface*, **13**(117), 20151046.
- Heinrich, R. and Schuster, S. (1996). *The regulation of cellular systems*. Springer Science & Business Media.
- Heinrich, R. *et al.* (1991). Mathematical analysis of enzymic reaction systems using optimization principles. *FEBS J.*, **201**(1), 1–21.
- Heinrich, R. *et al.* (1997). Theoretical approaches to the evolutionary optimization of glycolysis. *Eur. J. Biochem.*, **243**(1–2), 191–201.
- Kahm, M. *et al.* (2012). Potassium starvation in yeast: Mechanisms of homeostasis revealed by mathematical modeling. *PLoS Comput. Biol.*, **8**(6).
- Kaschek, D. *et al.* (2016). Testing the pattern of akt activation by variational parameter estimation. *IEEE Life Sci. Lett.*, **2**(2), 13–16.
- Klipp, E. *et al.* (2002). Prediction of temporal gene expression. *Eur. J. Biochem.*, **269**(22), 5406–5413.
- Lang, M. and Stelling, J. (2016). Modular parameter identification of biomolecular networks. *SIAM J. Sci. Comput.*, **38**(6), B988–B1008.
- Lebiedz, D. and Brandt-Pollmann, U. (2003). Manipulation of self-aggregation patterns and waves in a reaction-diffusion system by optimal boundary control strategies. *Phys. Rev. Lett.*, **91**(20), 208301/1–208301/4.
- McFarland, D. (1977). Decision making in animals. *Nature*, **269**(5623), 15–21.
- Mendes, P. and Kell, D. (1998). Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**(10), 869–883.
- Mombaur, K. (2016). *Optimal Control for Applications in Medical and Rehabilitation Technology: Challenges and Solutions*, pages 103–145. Springer International Publishing, Cham.
- Nimmegeers, P. *et al.* (2016). Dynamic optimization of biological networks under parametric uncertainty. *BMC Syst. Biol.*, **10**(1), 86.
- Oyarzún, D. (2011). Optimal control of metabolic networks with saturable enzyme kinetics. *IET Syst. Biol.*, **5**(2), 110–119.
- Oyarzún, D. A. *et al.* (2009). Sequential activation of metabolic pathways: a dynamic optimization approach. *Bull. Math. Biol.*, **71**(8), 1851–1872.
- Panchea, A. M. and Ramdani, N. (2015). Towards solving inverse optimal control in a bounded-error framework. In *2015 American Control Conference (ACC)*, pages 4910–4915.
- Parker, G. A. *et al.* (1990). Optimality theory in evolutionary biology. *Nature*, **348**(6296), 27–33.
- Popescu, A. I. (1998). Bionics, biological systems and the principle of optimal design. *Acta Biotheor.*, **46**(4), 299–310.
- Rashevsky, N. (1961). *Mathematical principles in biology and their applications*. Thomas.
- Rosen, R. (1967). *Optimality principles in biology*. Springer.
- Schaffer, W. M. (1983). The application of optimal control theory to the general life history problem. *Am. Nat.*, **121**(3), 418–431.
- Schelker, M. *et al.* (2012). Comprehensive estimation of input signals and dynamics in biochemical reaction networks. *Bioinformatics*, **28**(18), i529–i534.
- Smets, I. Y. *et al.* (2004). Optimal adaptive control of (bio) chemical reactors: past, present and future. *J. Process Control*, **14**(7), 795–805.
- Smith, J. M. (1978). Optimization theory in evolution. *Annu Rev Ecol Syst*, **9**(1), 31–56.
- Sootla, A. *et al.* (2016). Shaping pulses to control bistable systems: Analysis, computation and counterexamples. *Automatica*, **63**, 254–264.
- Sutherland, W. J. (2005). The best solution. *Nature*, **435**(7042), 569–569.
- Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci.*, **7**(9), 907.
- Trägårdh, M. *et al.* (2016). Input estimation for drug discovery using optimal control and markov chain monte carlo approaches. *J. Pharmacokinet. Phar.*, **43**(2), 207–221.
- Trägårdh, M. *et al.* (2017). Input estimation for extended-release formulations exemplified with exenatide. *Front. Bioeng. Biotechnol.*, **5**(24).
- Vassiliadis, V. *et al.* (1994). Solution of a class of multistage dynamic optimization problems. 2. problems with path constraints. *Ind. Eng. Chem. Res.*, **33**(9), 2123–2133.
- Waldherr, S. *et al.* (2015). Dynamic optimization of metabolic networks coupled with gene expression. *J. Theor. Biol.*, **365**, 469–485.
- Walter, E. and Pronzato, L. (1997). *Identification of parametric models from experimental data*. Springer Verlag.
- Wellstead, P. *et al.* (2008). The role of control and system theory in systems biology. *Annu. Rev. Control*, **32**(1), 33–47.
- Wessely, F. *et al.* (2011). Optimal regulatory strategies for metabolic pathways in escherichia coli depending on protein costs. *Mol. Syst. Biol.*, **7**(1), 515.
- Zaslaver, A. *et al.* (2004). Just-in-time transcription program in metabolic pathways. *Nat. Genet.*, **36**(5), 486–491.

SUPPLEMENTARY INFORMATION - Optimality and identification of dynamic models in systems biology: an inverse optimal control framework

Nikolaos Tsiantis^{1,2}, Eva Balsa-Canto¹ and Julio R. Banga^{1,*}

¹Bioprocess Engineering Group, Spanish National Research Council, IIM-CSIC, C/Eduardo Cabello 6, 36208 Vigo (Spain).

²Department of Chemical Engineering, University of Vigo, Vigo 36310, Spain.

*To whom correspondence should be addressed, <julio@iim.csic.es>.

Abstract

Here we provide detailed additional information regarding (i) the statements of the mathematical optimization problems, (ii) the numerical solution strategy, and (iii) the three case studies considered in the main article, plus an additional example of optimal input reconstruction in a multi-experiment setting. The code to reproduce the results reported here is available at <https://zenodo.org/record/1009541>

Contact: julio@iim.csic.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 SUPPLEMENTARY INFORMATION

In the main article, we provide the statement of the general inverse optimal control problem (IOCP-2), and then we describe a solution strategy involving the solution of two sub-problems, i.e. (i) a simultaneous input and parameter estimation problem, and (ii) a multiobjective optimal control problem. In Figure S1 we present a schematic representation of the overall workflow of our strategy (plots correspond to one of the case studies; more details are given below).

In this supplementary information document we give, for the sake of completeness, further details regarding the mathematical statements of these subproblems. We also provide further information regarding the numerical strategy to solve these nonlinear optimal control problems. Finally, we also give additional details and results regarding the four case studies considered.

1.1 Dynamic models of biological systems

Here we consider dynamic models of biological systems given by sets of deterministic nonlinear ordinary differential equations (ODEs), with possible additional constraints. We use the classical state space representation:

$$\frac{dx}{dt} = \tilde{\Psi}[x\{t, p\}, u\{t\}, p, t], \quad (S.1)$$

$$x\{t_0, p\} = x_0$$

$$\eta[x(t, p), u(t), p] = 0 \quad (S.2)$$

$$\zeta[x(t, p), u(t), p] \leq 0 \quad (S.3)$$

$$\eta_L[x(t_L, p), u(t_L), p] = 0 \quad (S.4)$$

$$\zeta_L[x(t_L, p), u(t_L), p] \leq 0 \quad (S.5)$$

where $\tilde{\Psi}$ is the right-hand side of the ordinary differential equations (ODEs) (S.1) describing the dynamics of the states x given initial conditions x_0 . Constraints can also be present in these dynamic models, either as path constraints (S.2, S.3) or as time-point constraints (S.4, S.5). Path constraints can be equalities η and inequalities ζ which must be enforced the whole time horizon considered. Similarly, time-point constraints can be equalities η_L and inequalities ζ_L only to be enforced at a specific point of time t_L .

Although the above formulation corresponds to the widely used ODE formalism, it should be noted that the inverse optimal control approach presented in our paper is also applicable to other dynamic modeling formalisms.

1.2 Statement of the general estimation problem

The problem of *parameter estimation* (PE) in nonlinear dynamic systems is usually formulated as a dynamic optimization problem. Given some data and a dynamic model, the optimal values for the model parameters are computed minimizing an objective (cost) function quantifying the differences between the model's predictions and the data. The generalized mathematical formulation of such problems corresponds to a nonlinear programming problem (NLP) with differential and algebraic constraints:

Find p to minimize

$$J = \sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_s} w_{ijk} (y_{ijk}(x(t_i, p), p) - \tilde{y}_{ijk})^2 \quad (S.6)$$

Subject to:

$$\frac{dx}{dt} = \tilde{\Psi}[x\{t, p\}, u\{t\}, p, t], \quad (S.7)$$

$$x\{t_0, p\} = x_0$$

Fig. S1. Diagram of the overall inverse optimal control solution strategy (the illustrations correspond to the particular case of LPN3B synthetic problem).

$$\mathbf{y}(\mathbf{x}, \mathbf{p}) = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) \quad (\text{S.8})$$

$$\boldsymbol{\eta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] = 0 \quad (\text{S.9})$$

$$\boldsymbol{\zeta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] \leq 0 \quad (\text{S.10})$$

$$\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U \quad (\text{S.11})$$

where J is the cost function to be minimized, $\mathbf{p}$ is the vector of decision variables of the optimization problem (the set of parameters to be estimated), $\tilde{\mathbf{y}}$ are the experimental measurements of a subset of the (so-called) output state variables, $\mathbf{y}(\mathbf{p}, t)$ are the model predictions for those outputs, $W(t)$ is a weighting (or scaling) matrix and $\mathbf{x}$ are the differential state variables. Besides, $\tilde{\Psi}$ is the set of differential and algebraic equality constraints (S.7) describing the system dynamics (i.e. the nonlinear process model), and $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ are the possible equality (S.9) and inequality (S.10) path and point constraints. Finally, $\mathbf{p}$ is subject to upper and lower bounds acting as inequality constraints (S.11).

The simultaneous input and parameter estimation problem (described in the main text as IOCP-1) is a generalization of the above problem where one seeks to estimate both time-invariant $\mathbf{p}$ and time-dependent $\mathbf{u}(t)$. This more general estimation problem is actually a particular case of nonlinear optimal control, which we describe next.

In the usual PE problem time-dependent inputs are treated as known (measured) quantities, and model parameters as unknown variables to be estimated from experimental data. However, in many other cases, this is not true, either due to large uncertainty in the input measurements or due to lack of such measurements. The above general estimation formulation allows us to address those situations where one needs to calibrate the dynamic model to the time-series data and simultaneously estimate unmeasured time-dependent inputs.

Extending the PE mathematical formulation, the cost functional for the IOCP-1 case needs to account for the dependency of the model's predictions on the inputs, so instead of Eqn. (S.6) we should use:

$$\min_{\mathbf{u}(t), \mathbf{p}, t_f} \sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_s} w_{ijk} (y_{ijk}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{p}, \mathbf{u}) - \tilde{y}_{ijk})^2 \quad (\text{S.12})$$

Additionally we consider bounds for the inputs of the form:

$$\mathbf{u}^L \leq \mathbf{u}(t) \leq \mathbf{u}^U \quad (\text{S.13})$$

Note that in IOCP-1, no inference of the underlying optimality principles is considered. In other words, the problem is restricted to estimating the unknown inputs and parameters of the model that best explain (fit) the data. To make the connection with the underlying optimality principles, we need to combine IOCP-1 with a multiobjective optimal control problem (OCP) in what is referred and described in the main text as IOCP-2. In the next section we provide a brief description of the optimal control problem, including its multi-objective formulation.

1.3 Statement of the general optimal control problem

Considering nonlinear dynamic systems, the problem of *optimal control* (OCP) consists of computing the optimal decision variables (time-varying inputs, or controls, and time-invariant parameters) that minimize (or maximize) a given cost functional (or performance index), subject to a set of differential equations and possibly algebraic constraints. Mathematically, the OCP is usually stated as follows:

$$\min_{\mathbf{u}(t), t_f} J[\mathbf{x}, \mathbf{u}] \quad (\text{S.14})$$

Subject to:

$$\frac{d\mathbf{x}}{dt} = \tilde{\Psi}[\mathbf{x}\{t, \mathbf{p}\}, \mathbf{u}\{t\}, \mathbf{p}, t], \quad (\text{S.15})$$

$$\mathbf{x}\{t_0, \mathbf{p}\} = \mathbf{x}_0$$

$$\boldsymbol{\eta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] = 0 \quad (\text{S.16})$$

$$\boldsymbol{\zeta}[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] \leq 0 \quad (\text{S.17})$$

$$\boldsymbol{\eta}_\iota[\mathbf{x}(t_\iota, \mathbf{p}), \mathbf{u}(t_\iota), \mathbf{p}] = 0 \quad (\text{S.18})$$

$$\boldsymbol{\zeta}_\iota[\mathbf{x}(t_\iota, \mathbf{p}), \mathbf{u}(t_\iota), \mathbf{p}] \leq 0 \quad (\text{S.19})$$

$$\mathbf{u}^L \leq \mathbf{u}(t) \leq \mathbf{u}^U \quad (\text{S.20})$$

where the time-dependent control variables ($\mathbf{u}(t)$) (along with the final time (t_f) in the case of the free terminal time problems), are computed in order to minimize (or maximize) an objective (cost) functional ($J[\mathbf{x}, \mathbf{u}]$), subject to the system's dynamics as well as the imposed constraints. The objective functional $J[\mathbf{x}, \mathbf{u}]$ corresponds to the optimality criteria hypothesis. In the case of a multicriteria formulation, the cost functional $J[\mathbf{x}, \mathbf{u}]$ is a set of objective functions corresponding to the N different criteria considered:

$$J[\mathbf{x}, \mathbf{u}, \mathbf{p}] = \begin{bmatrix} J_1[\mathbf{x}, \mathbf{u}, \mathbf{p}] \\ J_2[\mathbf{x}, \mathbf{u}, \mathbf{p}] \\ \vdots \\ J_N[\mathbf{x}, \mathbf{u}, \mathbf{p}] \end{bmatrix} \quad (\text{S.21})$$

where, in its general form, each objective function J_i in this set ($i \in [1, N]$) consists of a Mayer and a Lagrange term:

$$J_i[\mathbf{x}, \mathbf{u}, \mathbf{p}] = \Phi_M^i[\mathbf{x}(t_f, \mathbf{p}), \mathbf{p}] + \int_{t_0}^{t_f} \Phi_L^i[\mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}] \quad (\text{S.22})$$

The system's dynamics are described by Eqns. (S.15), i.e. the set of ordinary differential equations and the corresponding initial values ($\mathbf{x}\{t_0\}$), forming the so-called initial value problem (IVP). Equality ($\boldsymbol{\eta}$) and inequality ($\boldsymbol{\zeta}$) path constraints are represented in the sets of equations (S.16) and (S.17), respectively, as constraints to be enforced (e.g. total enzyme capacity, critical thresholds for specific concentrations, etc.). Also, the sets of equality ($\boldsymbol{\eta}_\iota$) and inequality ($\boldsymbol{\zeta}_\iota$) time-point constraints are given in (S.18) and (S.19), respectively, corresponding to constraints valid for a specific point of the time horizon. Finally, the upper and lower bounds ($\mathbf{u}^U, \mathbf{u}^L$) of the control variables' vector throughout the process are given in (S.20).

1.4 Numerical solution of nonlinear optimal control problems

Methods for the numerical solution of nonlinear optimal control problems can be classified under three categories: dynamic programming, indirect and direct approaches. Dynamic programming (Bellman, 1956; Bertsekas et al., 1995) suffers from the so-called *curse of dimensionality*, so the latter two are the most promising strategies for realistic problems. Indirect approaches were historically the first developed and are based on the transformation of the original optimal control problem into a multi-point boundary value problem using Pontryagin's necessary conditions (Bryson, 1973; Liberzon, 2012). Direct methods are based on the discretization of the control, known as the sequential strategy (Vassiliadis et al., 1994a), or both the control and the states, known as the simultaneous strategy (Biegler et al., 2002). Here we have chosen the control parameterization approach.

1.4.1 Control vector parameterization

The control vector parameterization (CVP) approach proceeds by dividing the time horizon into a number of elements (ρ). Each of the control variables ($j = 1 \dots n_u$) are then approximated within each interval ($i = 1 \dots \rho$) by means of some basis functions (usually, low order Lagrange polynomials (Vassiliadis et al., 1994b)) as follows:

$$u_j^{(i)}(t) = \sum_{k=1}^{M_j} u_{ijk} \ell_k^{(M_j)}(\tau^{(i)}) \quad t \in [t_{i-1}, t_i] \quad (\text{S.23})$$

being τ the normalized time in each element i :

$$\tau^{(i)} = \frac{t - t_{i-1}}{t_i - t_{i-1}} \quad (\text{S.24})$$

and M_j the order of the Lagrange polynomial (ℓ). In this work we will consider $M_j = 1$ or $M_j = 2$, i.e. step-wise or linear-wise control approximations.

In the CVP approach, the controls are expressed as functions of a new set of time invariant parameters corresponding to the polynomial coefficients ($\mathbf{w}$). Therefore the original infinite dimensional problem is transformed into a set of non-linear programming problems, with dynamic (the model) and algebraic constraints, in which the decision variables correspond to the original unknown parameters in θ and $\mathbf{w}$, which will be part of the overall set of parameters to determine.

1.5 Structural identifiability

Here we illustrate how the Taylor approach could be used to analyze the structural identifiability of a given model. The approach is based on the fact that observations are unique analytic functions of time and so all their derivatives with respect to time should also be unique (Pohjanpalto, 1978). It is thus possible to represent the observables by the corresponding Maclaurin series expansion and it is the uniqueness of this representation that will guarantee the structural identifiability of the system. The idea is to establish a system of non-linear algebraic equations on the unknowns, based on the calculation of the Taylor series coefficients, and to check whether the system has a unique solution.

Importantly for the IOCP problem, the unknowns include the vector of parameters and the vector of inputs (plus their derivatives with respect to time as evaluated at a given time, typically $t = 0^+$).

Let's consider the simple illustrative example:

$$\dot{x}_1 = 0 \quad (\text{S.25})$$

$$\dot{x}_2 = k_1 u x_1 - k_2 x_2 \quad (\text{S.26})$$

$$y = x_2 \quad (\text{S.27})$$

with $x_1(0) = 1$ and $x_2(0) = 0$. For illustrative purposes, we now use the Taylor approach to perform the structural identifiability analysis of the associated IOC problem. For the purpose of the analysis we will assume that $u(t)$ is a bounded continuously differentiable function of time. The first Taylor coefficients would read as follows:

$$T_0 = 0; \quad (\text{S.28})$$

$$T_1 = k_1 u(0); \quad (\text{S.29})$$

$$T_2 = k_1 u^1(0) - k_1 k_2 u(0); \quad (\text{S.30})$$

$$T_3 = k_1 u^2(0) - k_1 k_2 u^1(0) + k_1 k_2^2 u(0) \quad (\text{S.31})$$

$$T_4 = k_1 u^3(0) - k_1 k_2 u^2(0) + k_1 k_2^2 u^1(0) - k_1 k_2^3 u(0) \quad (\text{S.32})$$

$$\dots \quad (\text{S.33})$$

where u^i with $i = 1, 2, 3, \dots$ regards the first, second, third, etc. derivative of the control with respect to time.

Fig. S2. Illustrative examples of lack of structural identifiability under sustained stimulation. Figures correspond to the controls and the corresponding observables in two different scenarios: i) $k_u = 0.5$; $k_1 = 4$; $k_2 = 1$ (red) and $k_u = 1$; $k_1 = 2$; $k_2 = 1$ (blue); and ii) $k_u = 0.5$; $k_1 = 4$; $k_2 = 2$ (purple) and $k_u = 1$; $k_1 = 2$; $k_2 = 2$ (green). Figures illustrate how the inputs in each figure are different while the observations coincide but also that results vary in both figures. The fact that the observation does not change despite the value of u has to do with the fact that $k_u \cdot k_1 = 2$ in both cases; these results mean that k_u and k_1 are not uniquely identifiable (only their product is). Remarkably modifications in k_2 result in a different system behaviour, so k_2 becomes structurally identifiable.

Fig. S3. Illustrative examples of lack of structural identifiability under exponential stimulation. Figures correspond to the controls and the corresponding observables in two different scenarios: i) $k_u = 0.5$; $k_1 = 4$; $k_t = 0.1$ (red) and $k_u = 1$; $k_1 = 2$; $k_t = 0.1$ (blue); and ii) $k_u = 0.5$; $k_1 = 4$; $k_t = 0.5$ (purple) and $k_u = 1$; $k_1 = 2$; $k_t = 0.5$ (green). Figures illustrate how the inputs in each figure are different while the observations coincide but also that results vary in both figures. The fact that the observation does not change despite the value of u has to do with the fact that $k_u \cdot k_1 = 2$ in both cases; these results mean that k_u and k_1 are not uniquely identifiable (only their product is). Remarkably modifications in k_t result in a different system behavior, so k_t becomes structurally identifiable.

Since $k_1, k_2, u(0), u^1(0)$ and u^i with $i = 1, 2, 3, \dots$ are unknown in the general IOCP problem, we need to assess whether the solution of the system of equations S.29-S.33 on the unknowns is unique.

It is straightforward to see that the Jacobian of the Taylor coefficients with respect to the unknowns is rank deficient, indicating that parameters and u cannot be simultaneously uniquely estimated.

To get further information we may define specific control parameterizations. For example, if $u(t) = k_u$, it can be clearly seen that only $k_1 k_u$ and k_2 can be simultaneously estimated, but it will be impossible to give unique values for k_1 and k_u . Similarly, if $u(t) = k_u \exp(-k_t t)$, it can be again proved that $k_1 k_u, k_2$ and k_t can be uniquely identified in $\mathbf{R}^+$.

In Figures S2, S3 and S4 several simple illustrative examples of structural non-identifiability are depicted.

Fig. S4. Illustrative examples of lack of structural identifiability under logistic stimulation. Figures correspond to the controls and the corresponding observables in three different scenarios each: i) $k_u = 0.5; k_1 = 4; k_t = 1; t_u = 3$ (red) and $k_u = 1; k_1 = 2; k_t = 1; t_u = 3$ (blue); ii) $k_u = 0.5; k_1 = 4; k_t = 4; t_u = 3$ (purple) and $k_u = 1; k_1 = 2; k_t = 4; t_u = 3$ (green) and iii) $k_u = 0.5; k_1 = 4; k_t = 1; t_u = 8$ (orange) and $k_u = 1; k_1 = 2; k_t = 1; t_u = 8$ (cyan). Figures illustrate how the inputs in each figure are different while the observations coincide but also that results vary in all three figures. The fact that the observation does not change despite the value of u has to do with the fact that $k_u \cdot k_1 = 2$ in all three cases; these results mean that k_u and k_1 are not uniquely identifiable (only their product is). Remarkably modifications in k_t and t_u result in different system behaviors, so both parameters become structurally identifiable.

1.6 Error analysis

1.6.1 Goodness of fit

For each case study, the performance of our method is evaluated using the normalized root mean square error (NRMSE). This metric is a way to quantify the relative deviation of the model's predictions with respect to the experimental measurements or the known problem solution. For the case of the error with respect to the experimental measurements (observables), the NRMSE is computed as:

$$NRMSE_y = \sqrt{\frac{\sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_s} \left(\frac{y_{ijk}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{u}(t_i), \mathbf{p}) - \tilde{y}_{ijk}}{\max(\tilde{y}_{ijk}) - \min(\tilde{y}_{ijk})} \right)^2}{N_{data}}} \quad (\text{S.34})$$

Where n_{exp}, n_{obs}, n_s stand for the numbers of experiments, the number of observables and the number of time-points respectively, as N_{data} is the total number of data points available (all experiments, all observables and all time-points).

This is a typical metric used in parameter estimation (PE) problems. In the IOCP, we are not only estimating the unknown parameter set (θ), trying to fit the model's prediction to the observables, but also reconstructing the model's inputs. Therefore, in synthetic problems (where the input used to create the data and should be inferred is known), considering only the error related to the observables would not be an adequate quantification of the method's error.

An equivalent expression is used to compute the NRMSE with respect to the time-dependent inputs:

$$NRMSE_u = \sqrt{\frac{\sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{sti}} \sum_{i=1}^{n_{su}} \left(\frac{u_{ijk} - \tilde{u}_{ijk}}{\max(\tilde{u}_{ijk}) - \min(\tilde{u}_{ijk})} \right)^2}{N_{data}}} \quad (\text{S.35})$$

Where n_{exp}, n_{sti}, n_{su} stand for the numbers of experiments, the number of inputs and the number of elements of the inputs' time-vector respectively, as N_{data} is the total number of data points for the inputs available (all experiments, all inputs and all time-points). In this case of problems using simulated data, the true value of the parameters is known, so this metric can also be computed with respect to the estimated parameters themselves:

$$NRMSE_{\theta} = \sqrt{\frac{\sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{\theta}} \left(\frac{\theta_{jk} - \tilde{\theta}_{jk}}{\max(\tilde{\theta}_{jk}) - \min(\tilde{\theta}_{jk})} \right)^2}{N_{\theta data}}} \quad (\text{S.36})$$

Where n_{exp}, n_{θ} stand for the numbers of experiments (in case some or all parameters are local unknowns in a multi-experimental scheme) and the number of unknown parameters respectively, as $N_{\theta data}$ is the total number of data points for all unknown parameters available (all experiments, all unknown parameters).

1.6.2 Data generation

Several case studies included subcases with noisy synthetic data. We have generated realizations of this pseudo-experimental data by adding noise to the output of the models considering a Gaussian distribution with heteroscedastic variance as follows:

$$\tilde{y}_{o,s} = y_{o,s} + \epsilon_{o,s} \quad (\text{S.37})$$

with :

$$\epsilon_{o,s} = \sigma \times \text{rand}C_{o,s} \times y_{e,o} \quad (\text{S.38})$$

where $\epsilon_{o,s}$ are normally distributed independent random variables with standard deviation $\sigma \times y_{e,o}$ and $\text{rand}C_{o,s}$ is a random number drawn from the standard normal distribution $N(0, \infty)$ defined for every observable o and sampling time s . In the rest of the manuscript, we are referring to this error model as heteroscedastic proportional.

1.7 Case study JAK-STAT

This case study is an example of IOCP-1, based on the problem considered by Schelker *et al.* (2012), where we seek the simultaneous input reconstruction and parameter estimation in a dynamic model of the JAK2-STAT5 signaling pathway. The detailed mathematical statement was taken from <http://data2dynamics.github.io/d2d/> and is detailed below. In Figure S5 we show the corresponding network representation.

Fig. S5. Case JAKSTAT: Network representation for the JAK-STAT case study.

1.7.1 Problem formulation

The mathematical formulation of the simultaneous input and parameter estimation problem is:

$$\min_{pEpoR(t), \mathbf{p}} J[\mathbf{x}, pEpoR, \mathbf{p}] \quad (\text{S.39})$$

Where:

$$J[\mathbf{x}, pEpoR, \mathbf{p}] = \sum_{q=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_{data}} w_{ijq} (y_{ijq}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{p}) - \tilde{y}_{ijq})^2 \quad (\text{S.40})$$

Subject to:

$$\begin{aligned} \frac{dSTAT5}{dt} &= p_4 \cdot nSTAT5 \cdot \frac{0.45}{1.4} - p_1 \cdot pEpoR \cdot STAT5 \\ \frac{dpSTAT5}{dt} &= p_1 \cdot pEpoR \cdot STAT5 - 2 \cdot p_2 \cdot pSTAT5^2 \\ \frac{dpSTAT5_2}{dt} &= 2 \cdot p_2 \cdot pSTAT5^2 - p_3 \cdot pSTAT5_2 \\ \frac{dnSTAT5_2}{dt} &= p_3 \cdot pSTAT5_2 \cdot \frac{1.4}{0.45} - p_4 \cdot npSTAT5_2 \\ \frac{dnSTAT5_1}{dt} &= 2 \cdot p_4 \cdot npSTAT5_2 - p_4 \cdot nSTAT5_1 \\ \frac{dnSTAT5_2}{dt} &= p_4 \cdot nSTAT5_1 - p_4 \cdot nSTAT5_2 \\ \frac{dnSTAT5_3}{dt} &= p_4 \cdot nSTAT5_2 - p_4 \cdot nSTAT5_3 \\ \frac{dnSTAT5_4}{dt} &= p_4 \cdot nSTAT5_3 - p_4 \cdot nSTAT5_4 \\ \frac{dnSTAT5_5}{dt} &= p_4 \cdot nSTAT5_4 - p_4 \cdot nSTAT5_5 \end{aligned} \quad (\text{S.41})$$

$$y_1(\mathbf{x}, \mathbf{p}) = offset_tSTAT5 + scale_tSTAT5 \cdot (STAT5 + pSTAT5 + 2 \cdot pSTAT5_2) \quad (\text{S.42})$$

$$y_2(\mathbf{x}, \mathbf{p}) = offset_pSTAT5 + scale_pSTAT5 \cdot (pSTAT5 + 2 \cdot pSTAT5_2) \quad (\text{S.43})$$

Where the states vector is:

$$\mathbf{x} = [STAT5, pSTAT5, pSTAT5_2, npSTAT5_2, nSTAT5_1, nSTAT5_2, nSTAT5_3, nSTAT5_4, nSTAT5_5] \quad (\text{S.44})$$

And the initial conditions are:

$$\mathbf{x}(t_0) = [1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (\text{S.45})$$

1.7.2 Structural identifiability analysis

The structural identifiability analysis was performed using the generating series approach as implemented in GenSSI software toolbox (Chis et al., 2011a,b), available at <https://github.com/genSSI-developer/GenSSI>. For the analysis we assumed a Gaussian parameterization for $pEpoR(t) = EpoAe^{(-0.5(t-Epot)^2/EpoS^2)}$. This parameterization captures the main features of the measured data and it only requires the addition of three parameters: the maximum value ($EpoA$), the width of the Gaussian ($EpoS$) and the location of the center ($Epot$) of the Gaussian.

The model has been reported as non-identifiable regarding standard parameter estimation (Raue et al., 2009). Here we analyze the structural identifiability of the more general IOCP to assess what the added

Fig. S6. Case JAKSTAT0: estimated versus experimental stimuli measurements, along with the fit for the observed states. The overall quality of the estimation is given as normalized root mean square error (NRMSE) for the observables (y), the stimuli (u) and the parameters (θ). This solution corresponds to the synthetic noiseless case.

difficulties are. Our analysis reveals that the model is structurally non identifiable, i.e. it is impossible to uniquely determine all parameters plus the parameterization of the stimulus. The complete analysis reveals that $offset_tSTAT5$ and $offset_pSTAT5$ are globally structurally identifiable, p_4 is non identifiable and for the remaining parameters it is not possible to conclude identifiability.

Only if we fix the observables scaling parameters, p_1 and p_4 it is possible to obtain a full rank Jacobian in the generating series approach to guarantee local structural identifiability for $[offset_tSTAT5, offset_pSTAT5, p_2, p_3, EpoA, Epot, EpoS]$.

1.7.3 Numerical results

In this case study, we considered solving the IOCP-1 problem type for three different scenarios, one considering real experimental data (JAKSTATreal) and two additional scenarios using a synthetic data set. The real experimental data and the nominal values for the model parameters were taken from <http://data2dynamics.github.io/d2d/>. The synthetic data for the additional scenarios were generated by simulation of a chosen input profile either without the addition of numerical noise (JAKSTAT0) or with 5% heteroscedastic proportional noise (JAKSTAT5).

Given the structural identifiability analysis presented above, we first considered the synthetic subcases (JAKSTAT0 and JAKSTAT5) to study the impact on identifiability of fixing parameters to their nominal values. For the noiseless case JAKSTAT0, the inference of the input profile and the fit on the observables were almost perfectly recovered, as presented in Figure S6. The quality of inference is given in terms of NRMSE with respect to both the observables and the inputs. Additionally, the almost perfect identification of the known parameters is shown in Figure S7 where the bounds considered for the estimation of the parameters are shown as blue boxes, their true nominal values as orange lines, and the estimated values as dots.

For the noisy synthetic subcase (JAKSTAT5), we used regularization to achieve a satisfactory solution, avoiding possible artefacts as described in the main paper. We made use of two regularization terms, the first acting as a constraint on the second order derivative in order to avoid obtaining noisy input solutions that result in over-fitting the observables. At the same time, we avoid oscillatory behavior in the input profile and incorporate the prior knowledge on the behavior we expect from a biological inducer. We

Fig. S7. Case JAKSTAT0: parameter inference in the inverse optimal control problem. The box corresponds to the bounds used in the estimation, the orange dash to the true value used to generate the synthetic data and the orange dot to the estimate. This solution corresponds to the synthetic noiseless case.

Fig. S8. Case JAKSTAT5: illustration of practical non-identifiability in p_2 simulating a local solution ($pEpoR$ and p) for different values of p_2 and showing how it affects the observables y_1 and y_2 . This corresponds to the synthetic noisy case.

also used regularization on the input, providing a very simplistic initial guess, simulating in the simplest way the early peak that we expect to observe qualitatively. We were able to infer with high quality the input profile (Figure S9), obtaining a simultaneous good fit to the observables. We provide the NRMSE metrics in the same figure. The identification of the parameters is illustrated in Figures S10 and S11. These results indicate lack of practical identifiability for parameter p_2 . This issue is further illustrated in Figure S8 for p_2 values greater than 10, the outputs (and therefore the quality of the fit, as indicated by the NRMSE) are essentially the same.

Finally, for the real data subcase (JAKSTATreal) we took advantage of the analysis we performed in the synthetic subcases. We used the same discretization, a simple initial guess for the input and identical regularization strategy and settings. The input inference and the fit to the real

Fig. S9. Case JAKSTAT5: estimated versus experimental stimuli measurements, along with the fit for the observed states. The overall quality of the estimation is given as normalized root mean square error (NRMSE) for the observables (y) and the stimuli (u). This solution corresponds to the synthetic noisy case.

Fig. S10. Case JAKSTAT5: parameter inference in the inverse optimal control problem. The box corresponds to the bounds used in the estimation, the orange dash illustrates the true value used to generate the synthetic data and the orange dot shows to the estimate. This solution corresponds to the synthetic noisy case.

data are presented in the main paper, showing an excellent recovery. The corresponding convergence curve is given in Figure S54.

1.7.4 Uncertainty analysis

Analyzing the uncertainty in the estimates is non-trivial for this class of problems. Here, we will show how to quantify the uncertainty of the IOCP for one of the case studies presented (JAKSTAT5).

Our starting point is to consider IOCP as a generalization of the parameter estimation problem involving dynamic systems. Roughly speaking, for the parameter estimation problem there are four ways to compute errors in the parameter estimates: (i) using metrics based on the Fisher Information Matrix (FIM) and the Cramer-Rao inequality, (ii) bootstrapping approaches, (iii) Bayesian methods, and (iv) the profile likelihood

Fig. S11. Case JAKSTAT5: parameter inference in the inverse optimal control problem. Relative difference between the true values used to generate the synthetic data and the estimates. This solution corresponds to the synthetic noisy case.

method. A detailed review of these and related approaches can be found in [Vanlier et al. \(2013\)](#).

However, when we consider problems IOCP-I and IOCP-II and a frequentist framework, bootstrapping seems to be the only approach that can be extended easily. Essentially, in a bootstrap approach we perturb the data and solve the problem for each realization of the data. This is a straight-forward approach that generally has the disadvantage of being rather demanding computationally.

Considering the JAKSTAT5 case, we used the bootstrap approach generating different realizations of synthetic data and solving the problem for each one of them. Then the resulting distributions were analyzed statistically. In [Figure S13](#) the distribution obtained for the unknown parameters is presented. Note that the distributions in this figure, as well as the ones corresponding to the time elements of the input (pEpoR), are non-normal. Therefore we represent them using violin plots in [Figure S12](#) (for the case of model parameters).

Due to their non-normal nature, the concepts of standard deviation or confidence intervals are not applicable. Instead, we can use the concept of credible intervals, introduced in Bayesian statistics ([Kruschke, 2014](#)), ([Kruschke, 2013](#)), as an equivalent for our purposes. In [Figure S14](#) the input (pEpoR) bootstrap result is presented. The 90% high density interval (credible interval) was computed ([Table S.T1](#)) and is illustrated with the green envelope.

1.8 Case study TSP

This case study is another, more complex, instance of the IOCP-1 class. It is based on the biochemical pathway studied by [Moles et al. \(2003\)](#) for parameter estimation. Here we consider the simultaneous input and parameter estimation in a dynamic model of a fully observed 3 step pathway, with 8 states, 36 parameters and 2 inputs (which take different values for a set of 16 different experimental conditions). It should be noted that practical identifiability issues for the parameter estimation problem studied by [Moles et al. \(2003\)](#) were reported by [Rodriguez-Fernandez et al. \(2006\)](#). Thus, we expect the extended IOCP version presented here to be very challenging, and therefore a good opportunity to evaluate the advantages of the regularization methodology discussed above. The associated network representation is given in [Figure S15](#).

Table S.T1. Uncertainty analysis: credible intervals computed for the JAKSTAT5 case with 90% probability. The pEpoR time vector elements (ioc1-9) correspond to the value of pEpoR for $t=[0 \ 3 \ 7 \ 11 \ 15 \ 24 \ 36 \ 48 \ 60]$. The high density interval (HDI) computed is then used to plot [Figure S14](#).

IOC_upar	mode90%	90%HDI lb	90%HDI ub
p2	-	500	500.0235
p3	0.113052	0.095799	0.1301
offset_tSTAT	0.179433	0.138071	0.20572
offset_pSTAT	0.235583	0.222378	0.256238
ioc1	0.067923	0.044708	0.116407
ioc2	0.074264	0.023111	0.109239
ioc3	0.981378	0.775684	1
ioc4	0.946703	0.116469	1
ioc5	0.196829	0.072747	0.889468
ioc6	0.12167	0.056991	0.307032
ioc7	0.025029	0.011514	0.044097
ioc8	0.011836	0.00499	0.020522
ioc9	0.010573	0.003899	0.016652

HDI:High Density Intervals, lb: lower bound, ub:upper bound.

Fig. S12. Uncertainty analysis for JAKSTAT5: violin plots depicting the distribution of the parameters found using the bootstrap approach.

1.8.1 Problem formulation

The formulation of the simultaneous input and parameter estimation problem is:

$$\min_{\mathbf{u}(t), \mathbf{p}} J[\mathbf{X}, \mathbf{u}, \mathbf{p}] \quad (\text{S.46})$$

Where:

$$J[\mathbf{X}, \mathbf{u}, \mathbf{p}] = \sum_{q=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_{data}} w_{ijq} (y_{ijq}(\mathbf{x}(t_i, \mathbf{p}), \mathbf{p}) - \tilde{y}_{ijq})^2 \quad (\text{S.47})$$

Subject to:

Fig. S13. Uncertainty analysis for JAKSTAT5: histograms depicting the distribution of the parameters resulting from the bootstrap approach. The true value of the parameters used to generate all data realizations is marked with the red vertical line. Note the bi-modality of p_2 due to its lack of identifiability.

Fig. S14. Uncertainty analysis for JAKSTAT5: comparison between the mean and the true trajectories of the time-dependent input resulting from the bootstrap approach. The high density interval (90% HDI) is represented as the gray envelope.

Fig. S15. Case TSP: Network representation for the TSP case study.

Fig. S16. Case TSP: goodness of the estimation for the two subcases, TSP0 (noiseless) and TSP5 (5% noise) in terms of normalized root mean square error (NRMSE) of the observables (y), the stimuli (u) and the parameters (θ). The observables, stimuli and parameters are described in (S.49), (S.51), (S.52).

$$\begin{aligned}
 \frac{dG_1}{dt} &= \frac{V_1}{1 + \left(\frac{P}{K_{i1}}\right)^{ni_1} + \left(\frac{Ka_1}{S}\right)^{na_1}} - k_1 \cdot G_1 \\
 \frac{dG_2}{dt} &= \frac{V_2}{1 + \left(\frac{P}{K_{i2}}\right)^{ni_2} + \left(\frac{Ka_2}{M_1}\right)^{na_2}} - k_2 \cdot G_2 \\
 \frac{dG_3}{dt} &= \frac{V_3}{1 + \left(\frac{P}{K_{i3}}\right)^{ni_3} + \left(\frac{Ka_3}{M_2}\right)^{na_3}} - k_3 \cdot G_3 \\
 \frac{dE_1}{dt} &= \frac{V_4 \cdot G_1}{K_4 + G_1} - k_4 \cdot E_1 \\
 \frac{dE_2}{dt} &= \frac{V_5 \cdot G_2}{K_5 + G_2} - k_5 \cdot E_2 \\
 \frac{dE_3}{dt} &= \frac{V_6 \cdot G_3}{K_6 + G_3} - k_6 \cdot E_3 \\
 \frac{dM_1}{dt} &= \frac{kcat_1 \cdot E_1 \cdot \left(\frac{1}{Km_1}\right) \cdot (S - M_1)}{1 + \frac{S}{Km_1} + \frac{M_1}{Km_2}} \\
 &\quad - \frac{kcat_2 \cdot E_2 \cdot \left(\frac{1}{Km_3}\right) \cdot (M_1 - M_2)}{1 + \frac{M_1}{Km_3} + \frac{M_2}{Km_4}} \\
 \frac{dM_2}{dt} &= \frac{kcat_2 \cdot E_2 \cdot \left(\frac{1}{Km_3}\right) \cdot (M_1 - M_2)}{1 + \frac{M_1}{Km_3} + \frac{M_2}{Km_4}} \\
 &\quad - \frac{kcat_3 \cdot E_3 \cdot \left(\frac{1}{Km_5}\right) \cdot (M_2 - P)}{1 + \frac{M_2}{Km_5} + \frac{P}{Km_6}}
 \end{aligned} \tag{S.48}$$

$$y(\mathbf{X}, \mathbf{p}) = \mathbf{X} \tag{S.49}$$

Where the states vector is:

$$\mathbf{X} = [G_1, G_2, G_3, E_1, E_2, E_3, M_1, M_2] \tag{S.50}$$

And the inputs and parameters vectors are, respectively:

$$\mathbf{u} = [S, P] \tag{S.51}$$

Fig. S17. Case TSP: normalized root mean square error for noisy and noiseless case, with respect to the observables (y), the inputs (u) and the true parameters (θ). TSP5 corresponds to the left y-axis while TSP0 to the right one. The units of the latter are presented scaled in 10^{-5} . The observables, stimuli and parameters are described in (S.49), (S.51), (S.52).

$$\mathbf{p} = [V_1, Ki_1, ni_1, Ka_1, na_1, k_1, V_2, Ki_2, ni_2, Ka_2, na_2, k_2, V_3, Ki_3, ni_3, Ka_3, na_3, k_3, V_4, K_4, k_4, V_5, K_5, k_5, V_6, K_6, k_6, kcat_1, Km_1, Km_2, kcat_2, Km_3, Km_4, kcat_3, Km_5, Km_6] \quad (\text{S.52})$$

And the initial conditions are:

$$\mathbf{x}^T(t_0) = \begin{bmatrix} 0.66667 \\ 0.57254 \\ 0.41758 \\ 0.4 \\ 0.36409 \\ 0.29457 \\ 1.419 \\ 0.93464 \end{bmatrix} \quad (\text{S.53})$$

where $M_1, M_2, E_1, E_2, E_3, G_1, G_2$ and G_3 represent concentrations of the species involved in the different biochemical reactions and S and P keep fixed initial values for each experiment (i.e. parameters under our control). The parameters are divided in two different classes: Hill coefficients, allowed to vary within the range (0.1, 10), and all the others, allowed to vary within the range (10^{-6} , 500).

1.8.2 Numerical results

We considered two scenarios, noisy (TSP5) and noiseless (TSP0) data. Thus, we generated data for 16 experiments with different inputs, both with and without 5% heteroscedastic proportional noise. Our results are summarized in Figure S16 where we give the NRMSE values (for inputs, states and parameters) for the two different subcases. Note, this is a synthetic scenario and knowing the true parameter values allows us to compute their NRMSE.

For the noiseless subcase (TSP0), there was no need to use regularization, and we were able to recover almost exactly all the inputs and parameters (note the very low NRMSE values). The input reconstruction is summarized in Figure S18 showing an analysis based on the relative difference of the values inferred and the true inputs for each of the 16 experimental conditions and each of the two time-invariant inputs. The parameter identification results are presented in Figure S19 as the relative difference of the parameters computed to their true values (since the

Table S.T2. Case TSP: evolution of NRMSE for noisy and noiseless cases while iteratively decreasing the regularization parameters.

Noise	Iterations	$NRMSE_y$	$NRMSE_u$	$NRMSE_\theta$
0%	1st	0.000030	0.000021	0.000033
10%	1st	0.375198	0.644647	0.842095
10%	2nd	0.308303	0.677463	0.774919
10%	3rd	0.128832	0.148638	0.168275

y:observables, u:stimuli, θ :parameters.

Fig. S18. Case TSP0: stimuli inference per experiment. In this case study, synthetic multi-experimental datasets were used for the inverse optimal control problem. Here the relative difference between the true input values used to generate the noiseless data for each experiment and the input estimates is analyzed, per experiment, for both inputs.

problem is synthetic and therefore we know the true values of parameters). The fit on the observables for experiment I is given in Figure S22 (the fit for the remaining 15 experiments are similar, and not shown here for the sake of brevity).

In the case of TSP5, we achieved a good inference of the true solution despite the practical identifiability issues of this problem. In contrast to TSP0, where no regularization was used, in TSP5 we used two of the implemented regularization terms in our cost function. In particular, we used a regularization scheme with re-optimizations, starting with relatively large values for the regularization parameters for both stimuli and parameters. Then we iterate, re-optimizing using the solution obtained as an initialization guess for the new optimization, while decreasing the values of the regularization parameters. This procedure resulted in very good estimations of the stimuli for the multiple experiments. We also achieved good parameter estimates, although we also found the identifiability issues previously reported (Rodriguez-Fernandez et al., 2006), as expected. The impact of the iterative decrease of regularization on the performance of our method can be visualized in Figure S17 and is also given in table format in Table S.T2. These results illustrate how the NRMSE improves at each iteration (TSP5 on the left y-axis). We also give as a reference the much lower NRMSE of the noiseless case (TSP0, on the right y-axis). The inference of the inputs and parameters is presented in Figures S20 and S21. The fit on the observables for one of the experiments is given in Figure S23 and the convergence curve in Figure S55.

Fig. S24. Case LPN3B: Network representation for the LPN3B case study.

Where:

$$J_{\text{outer}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] = \sum_{q=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_{data}} w_{ijq} (y_{ijq}(\mathbf{S}(t_i, \mathbf{k}), \mathbf{e}, \mathbf{k}) - \tilde{y}_{ijq})^2 \quad (\text{S.55})$$

Subject to:

$$\min_{\mathbf{e}(t), t_f} J_{\text{inner}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] \quad (\text{S.56})$$

Where:

$$J_{\text{inner}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] = \left[t_f, \int_{t_0}^{t_f} (S_2 + S_3) dt \right]^T \quad (\text{S.57})$$

Subject to the system dynamics:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v} \quad (\text{S.58})$$

$$\mathbf{y}(\mathbf{S}, \mathbf{k}) = \mathbf{S} \quad (\text{S.59})$$

Where the states' vector is:

$$\mathbf{S} = [S_1, S_2, S_3, S_4] \quad (\text{S.60})$$

While the stoichiometric matrix $\mathbf{N}$ is:

$$\mathbf{N} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

And the kinetics are described by:

$$v_i = k_i \cdot S_i \cdot e_i \quad (\text{S.61})$$

With the following end-point constraint:

$$S_4(t_f) = P(t_f) \quad (\text{S.62})$$

and path constraint:

$$\sum_{i=1}^3 e_i \leq E_T \quad (\text{S.63})$$

with: $E_T = 1$ M, $S_1(t_0) = 1$ M, $S_i(t_0) = 0$ for $i = 2, 3, 4$ and $P(t_f) = 0.9$ M.

1.9.2 Structural identifiability analysis

The structural identifiability analysis was performed using the generating series approach as implemented in the GenSSI software toolbox (Chis et al. (2013a)), available at github.com/genssideveloper/GenSSI.

We first needed to parameterize the solution for the activation profiles for the different enzymes $e_i(t)$, $i = 1 \dots 3$. With this aim, we exploited the fact that an underlying optimality principle drives the system, i.e. the system evolves to minimize the final time to achieve a certain amount of product. Under this assumption, the enzyme activation profiles can be approximated by a boxcar function as follows:

$$e_i = k_{ei1}H(t - t_{i1}) - k_{ei2}H(t - t_{i2}) \quad (\text{S.64})$$

where $H(t)$ is the Heaviside step function; k_{ei1} and k_{ei2} regard the amplitudes of the enzyme activation, and t_{i1} and t_{i2} determine the switching times from 0 to k_{ei} or the other way around.

Table S.T3. Case LPN3B: brief description of sub-cases considered for the LPN3B case study. LPN3B0 and LPN3B10 are the main sub-cases solved. The rest of the sub-cases are variations presented to illustrate different issues of the inverse optimal control problem.

Label	Noise level	Description
LPN3B0	0%	Noiseless case
LPN3B10	10%	Noisy case
LPN3B0re	0%	LPN3B0 without initial assumption of input discretization and using re-optimizations
LPN3B10re	0%	LPN3B10 without initial assumption of input discretization and using re-optimizations
LPN3B0noreg	0%	LPN3B0 without regularization
LPN3B0p2l2	0%	LPN3B0 with different parameter bounds
LPN3B0noreg2	0%	LPN3B0 without regularization and with different parameter and input bounds

A continuous approximation of the boxcar function would be:

$$e_i \approx \frac{k_{ei1}}{1 + e^{-2k_t(t-t_{i1})}} - \frac{k_{ei2}}{1 + e^{-2k_t(t-t_{i2})}} \quad (\text{S.65})$$

with k_t large enough (for example, $kt > 20$).

The number of parameters under this approximation is 15, three kinetic constants plus the boxcar approximation related parameters. The generating series approach results in a rank deficient Jacobian. The addition of further derivatives to the method does not increase the Jacobian's rank. Therefore it is concluded that the model in its more general form is non-identifiable.

To further reduce the number of unknowns, we exploit again the underlying optimality principle: 1) we assume that the enzymes are either active or inactive, i.e. $k_{ei} = k_{ei1} = k_{ei2}$, $i = 1 \dots 3$; 2) we apply the constraint $\sum_{i=1}^3 k_{ei} = E_T$, i.e. the sum of the enzymes corresponds to the maximum E_T over time; and 3) we use the just-in-time activation property (Zaslaver et al. (2004)) of linear pathways to fix $t_{11} = 0$, i.e. first enzyme is active from initial time and $t_{22} = t_{31}$. The number of unknowns is 9.

After several derivatives the rank of the Jacobian is 7, which means that the model is still non-identifiable. The model becomes, at least, locally structurally identifiable if we fix either the kinetic constants (k_i) or the amplitudes of the enzyme activation (k_{ei}).

In summary, exploiting the biological knowledge about the optimality principle underlying the dynamics of these linear pathways enabled the reduction of the unknowns in the IOCP problem. However it was not possible to guarantee structural identifiability. We discuss the consequences of this lack of identifiability in the next section.

1.9.3 Numerical results

We solved the IOCP-2 formulation considering synthetic data (generated by simulation) for two scenarios: data with 10% heteroscedastic proportional noise (LPN3B10), and noiseless data (LPN3B0). The quality of the solutions computed for these two subcases are shown in Table S.T4 in terms of the normalized root mean square error (NRMSE) for the observables, inputs and (since we are dealing with a synthetic problem, where the true values of parameters are known) the model parameters.

In the noiseless case, we achieved almost perfect reconstruction of inputs and parameters. In the noisy case we were able to achieve a very good solution using a piecewise constant approximation for the inputs. In order to tackle the identifiability issues discussed above we used regularization in an iterative scheme for both the inputs and parameters. Later we

Fig. S25. Case LPN3B0: stimuli inference in 3rd re-optimization of the inverse optimal control problem and the relative difference between the true parameter values used to generate the synthetic noiseless data and the estimates.

Fig. S26. Case LPN3B0: fit for the observables in 3rd re-optimization of the inverse optimal control problem for the noiseless case.

analyzed the impact of this strategy on the inference of the true solution. The results for the LPN3B0 case are summarized in Figures S25 and S26. The respective results for the noisy case are presented in the Figures S27, S28 and S56.

It is worth using this relatively simple (but challenging) case study to illustrate several additional key points of our approach. In Figure S1 we present a schematic representation of the workflow we consider for synthetic problems. Assuming a set of parameters for our model, as well as the underlying optimality principles, we solve the multi-objective OCP. We then obtain a set of all optimal solutions (Pareto front) as different trade-offs of the criteria considered. We generate synthetic data considering the inputs resulted from the OCP (and the same assumed set of parameters) and that correspond to one of the Pareto optimal points. To parallelize with the general IOCP work-flow figure presented in the main paper, we then refer to the first step, that is solving the input estimation problem from the synthetic data set. As a result, we have the reconstructed input and the inferred parameters that can be now used in step 2. In this step, we solve

Fig. S27. Case LPN3B10: goodness of the estimation for the noisy and noiseless subcase in comparison, showing the normalized root mean square error and the reconstructed stimuli in the noisy subcase (versus the true value used for the pseudo-experimental noisy data generation).

Fig. S28. Case LPN3B10: fit to the observables in the inverse optimal control problem for the noisy case.

the multicriteria OCP (using the parameter set that we got from step 1). We, thus, obtain a Pareto front where we proceed in step 3 matching the reconstructed solution of step 1 to one of the Pareto optimal points and thus infer the specific trade-off of the criteria considered that is consistent with the data used. In Figure 4 (in the main paper) such a comparison is made for the solution taken in the noisy case (LPN3B10) and presented along with the multi-objective OCP Pareto front, in which it is located.

It is also worth noting the challenges present in this problem. For instance, in the main subcases of the problem (LPN3B10 and LPN3B0) we used a 3-step approximation for the inputs profiles, where each step has varying duration of time. This is something that can be easily justified in this case by the optimality principle underlying in linear metabolic networks of minimal transition time, described as the just-in-time activation profile. If we ignore this assumption, the suitable number of steps to approximate the inputs can be determined by solving the problem using

Fig. S29. Case LPN3B0re: stimuli inference in re-optimization 4_3 and the relative difference between the true parameter values used to generate the synthetic noiseless data and the estimates. This subcase corresponds to the solution of the noiseless case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

Fig. S30. Case LPN3B0re: fit on the observables in re-optimization 4_3 in the inverse optimal control problem. This subcase corresponds to the solution of the noiseless case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

Table S.T4. Case LPN3B: NRMSE for noiseless (LPN3B0) and noisy (LPN3B10) case.

Noise	Iterations	$NRMSE_y$	$NRMSE_u$	$NRMSE_\theta$
0%	1st	0.412983	0.577643	0.999990
0%	2nd	0.382238	0.482203	0.931450
0%	3rd	0.153955	0.141419	0.396655
0%	4th	0.023829	0.115464	0.098381
10%	1st	0.405368	0.577643	0.999990
10%	2nd	0.487641	0.486206	0.935376
10%	3rd	0.208916	0.141912	0.447917
10%	4th	0.131578	0.141407	0.112145

y:observables, u:stimuli, θ :parameters.

Fig. S31. Case LPN3B0re: Quality of fit in the solution of the inverse optimal control problem presented in Fig. S29 corresponding to the noiseless case, in terms of normalized root mean square (with respect to the observables (y), inputs (u) and parameters (θ)) and how it changes by the iterative use of regularized re-optimization.

Fig. S32. Case LPN3B10re: stimuli inference in re-optimization 4_3 and the relative difference between the true parameter values used to generate the synthetic noisy data and the estimates. This subcase corresponds to the solution of the noisy case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

increasing numbers of discretization elements. In order to test the capabilities of our methods by using zero prior knowledge (worst case scenario for a problem of unknown solution), we considered obtaining the solution of the same two main scenarios using input profiles that have been refined through sequential re-optimizations increasing the number of discretization elements. In that way we can approximate any profile of inputs starting from very few piecewise constant elements of fixed time duration. These illustrative subcases are referred to as LPN3B10re and LPN3B0re, corresponding to the noisy and noiseless case respectively. All the subcases considered for illustrative purposes are summarized in Table S.T3.

The results for the noisy case are presented in Figures S32, S33 and S34. Here each iteration is labeled as the number of times the input discretization has been doubled (starting from the initial number of 5 elements) with the number of iterative decreases in the regularization parameters α and β as a

Fig. S33. Case LPN3B10re: fit on the observables in re-optimization 4_3 in the inverse optimal control problem. This subcase corresponds to the solution of the noisy case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

Fig. S34. Case LPN3B10re: Quality of fit in the solution of the inverse optimal control problem presented in Fig. S32 corresponding to the noisy case, in terms of normalized root mean square (with respect to the observables (y), inputs (u) and parameters (θ)) and how it changes by the iterative use of regularized re-optimization.

subscript. In Figure S34 we can see that after a certain point, by increasing more the discretization level and relaxing the regularization parameters the solution was insignificantly improving in terms of NRMSE, e.g. for the parameters, but was actually becoming worse for the inputs. The later can also be visualized by the obvious numerical noise in the input profiles in Figure S32. In LPN3B0re, due to the absence of noise, the case is not the same and the solution is being improved until the last iteration of the scheme that we implemented (Figure S31). In addition to that, the input profiles are smoother and of course the metrics (NRMSE) better, as it can be expected. The results are summarized in Figures S29, S30 and S31.

Furthermore, it is worth illustrating numerically the structural identifiability issues arising in these formulations, i.e. we can obtain the same fit to the model's observables with completely different input profiles. Following three more subcases are presented to support our claim (LPN3B0noreg,

Fig. S35. Case LPN3B0noreg: two different optimal solutions with the same quality of fit to the observables as a result of lack of structural identifiability. The red dashed line is the true solution used to generate the synthetic noiseless data and the blue line is the estimate computed without the use of regularization for this ill-posed case. Their relative difference in terms of parameters estimated is also presented. The underestimation in the inputs is compensated by the massive difference in parameter values.

Fig. S36. Case LPN3B0noreg: fit to the observables for the optimal solutions depicted in Fig. S35.

LPN3B0p2l2 and LPN3B0noreg2), all of them considering noiseless synthetic data to emphasize the issues even in the ideal scenario.

In particular, in subcase LPN3B0noreg, we can obtain a solution of the problem without any use of regularization (as expected for a noiseless case). We summarize the results in Figures S35 and S36. The inputs profiles are quite different with respect to the original ones, even though the model's output fit the observables extremely well. As we also discussed in the main paper, this is a consequence of the problem being ill-posed (due to the lack of structural identifiability). We can reduce these difficulties using regularization, i.e. we surmount ill-posedness by incorporating prior information via regularization.

In addition, in subcases LPN3B0p2l2 and LPN3B0noreg2 we show how introducing prior information (in the bounds and using regularization)

Fig. S37. Case LPN3B0p2l2: two different optimal enzyme profiles which result in the same quality of fit. Since the optimal p_2 value is twice its true value, the computed enzyme profiles compensate for this modification. This result illustrates the lack of structural identifiability and corresponds to the solution of the noiseless case with different bounds on the parameter set.

Fig. S39. Case LPN3B0norege2: two different optimal enzyme profiles which result in the same quality of fit. The computed enzyme profiles compensate for the massive differences in the parameter estimates. This result illustrates the lack of structural identifiability and corresponds to the solution of the noiseless case with different bounds on parameter set and inputs.

Fig. S38. Case LPN3B0p2l2: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S37

Fig. S40. Case LPN3B0norege2: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S39

can improve the inferred solution. In LPN3B0p2l2, while using regularization in the exact same way as to obtain the LPN3B0 solution, we constrain the bounds of parameter p_2 , to force a solution far for the true one ($= 1$) and imposing a lower bound equal to 3 in the IOCP formulation. As a result, the input profiles we obtain show that the increase in the resulting p_2 is totally compensated by a decrease in the e_2 activation, yet the output of the model remains unaltered. The results are given in Figures S37 and S38. In subcase LPN3B0norege2, we do not make use of any regularization term in our cost function. We constrain p_2 between 0 and 0.1 and we increase the upper bound of e_2 from 1 to 100. The results are shown in Figures S39 and S40. As one can expect, the same (with the LPN3B0) output is obtained but with much different input profiles. The big decrease in p_2 is compensated by a massive increase in e_2 while having no regularization in use e_1 and e_3 activations are decreased and

compensated by increased p_1 and p_3 values resulting again in an excellent fit.

1.10 Case study SC

This case study is presented as the second example of the general IOCP-2 problem. The model corresponds to the central carbon metabolism of yeast during diauxic shift in a nutrient depletion scenario. The model formulation is taken from Klipp *et al.* (2002) and de Hijas-Liste *et al.* (2014). It consists of six metabolic reactions: the upper and lower glycolysis, the ethanol formation and consumption, the TCA cycle and the respiratory chain. This description results in 9 dynamic states, 8 parameters and 6 time-dependent enzyme concentrations. Additionally, we consider the critical values of ATP and NADH as constraints on the states of the model. In the multicriteria OCP formulation considered by de Hijas-Liste *et al.* (2014), the objective was to find the enzyme activation profiles which are Pareto

optimal, considering two objective functions (maximization of survival time and minimization of enzyme synthesis cost). Here the synthetic data used to fit the model are generated considering one of the points of the Pareto front resulted by this OCP in the same way previously described for LPN3B. In Figure S41 we give the network representation that was considered.

Fig. S41. Case SC: Network representation for the SC case study. Glucose and triose phosphates are represented in this model by X_1 and X_2 , while pyruvate and ethanol, respectively, by X_3 and X_4 .

1.10.1 Problem formulation

The mathematical formulation of the inverse optimal control problem is given as:

$$\min_{\mathbf{e}(t), \mathbf{k}, t_f} J_{\text{outer}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] \quad (\text{S.66})$$

Where:

$$J_{\text{outer}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] = \sum_{q=1}^{n_{\text{exp}}} \sum_{j=1}^{n_{\text{obs}}} \sum_{i=1}^{n_{\text{data}}} w_{ijq} (y_{ijq}(\mathbf{S}(t_i, \mathbf{k}), \mathbf{e}, \mathbf{k}) - \tilde{y}_{ijq})^2 \quad (\text{S.67})$$

Subject to:

$$\max_{\mathbf{e}(t), t_f} \mathbf{J}_{\text{inner}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] \quad (\text{S.68})$$

Where:

$$\mathbf{J}_{\text{inner}}[\mathbf{S}, \mathbf{e}, \mathbf{k}] = \left[t_f, -\int_{t_0}^{t_f} (\sum_{i=1}^6 e_i) dt \right]^T \quad (\text{S.69})$$

Subject to:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N} \cdot \mathbf{v}, \mathbf{S}\{t_0\} = \mathbf{S}_0 \quad (\text{S.70})$$

$$\mathbf{y}(\mathbf{S}, \mathbf{k}) = \mathbf{S} \quad (\text{S.71})$$

$$\sum_{i=1}^6 e_i \leq E_T \quad (\text{S.72})$$

$$NADH \geq NADH_c \quad (\text{S.73})$$

$$ATP \geq ATP_c \quad (\text{S.74})$$

Where the states' vector is:

$$\mathbf{S} = [X_1, X_2, X_3, X_4, NADH, ATP, NAD, ADP] \quad (\text{S.75})$$

And the stoichiometric matrix along with the reactions' kinetics corresponds to (S.76) and (S.77) respectively and are given below:

$$\mathbf{N} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 4 & -1 & 0 & -1 \\ -2 & 2 & 0 & 0 & 0 & 3 & -1 & 0 \\ 0 & -1 & 1 & -1 & -4 & 1 & 0 & 1 \\ 2 & -2 & 0 & 0 & 0 & -3 & 1 & 0 \end{bmatrix} \quad (\text{S.76})$$

$$v_1 = k_1 \cdot e_1 \cdot X_1 \cdot ATP$$

$$v_2 = k_2 \cdot e_2 \cdot X_2 \cdot NAD \cdot ADP$$

$$v_3 = k_3 \cdot e_3 \cdot X_3 \cdot NADH$$

$$v_4 = k_4 \cdot e_4 \cdot X_4 \cdot NAD$$

$$v_5 = k_5 \cdot e_5 \cdot X_3 \cdot NAD$$

$$v_6 = k_6 \cdot e_6 \cdot NADH \cdot ADP$$

$$v_7 = k_7 \cdot ATP$$

$$v_8 = k_8 \cdot NADH$$

(S.77)

With the given initial values of:

$$\mathbf{S}_0 = [1 \ 1 \ 1 \ 10 \ 0.7 \ 0.8 \ 0.3 \ 0.2] \quad (\text{S.78})$$

A path constraint on the total amount of enzymes is implemented in Eqn. (S.72) where E_T is 11.5. In (S.73) and (S.74) the critical values $NADH_c = 0.5$ and $ATP_c = 0.7$, of NADH and ATP respectively, above which the cell is considered to survive are implemented also as path constraints.

We reformulated the model as explained below. After considering the computation of the products $\tilde{e}_i = k_i \cdot e_i : i \in [1, 6]$, (S.66), (S.67), (S.68), (S.69), respectively, become:

$$\min_{\tilde{\mathbf{e}}(t), \mathbf{k}, t_f} J_{\text{outer}}[\mathbf{S}, \tilde{\mathbf{e}}, \mathbf{k}] \quad (\text{S.79})$$

Where:

$$J_{\text{outer}}[\mathbf{S}, \tilde{\mathbf{e}}, \mathbf{k}] = \sum_{q=1}^{n_{\text{exp}}} \sum_{j=1}^{n_{\text{obs}}} \sum_{i=1}^{n_{\text{data}}} w_{ijq} (y_{ijq}(\mathbf{S}(t_i, \mathbf{k}), \tilde{\mathbf{e}}, \mathbf{k}) - \tilde{y}_{ijq})^2 \quad (\text{S.80})$$

Subject to:

$$\max_{\tilde{\mathbf{e}}(t), t_f} \mathbf{J}_{\text{inner}}[\mathbf{S}, \tilde{\mathbf{e}}, \mathbf{k}] \quad (\text{S.81})$$

Where:

$$\mathbf{J}_{\text{inner}}[\mathbf{S}, \tilde{\mathbf{e}}, \mathbf{k}] = \left[t_f, -\int_{t_0}^{t_f} (\sum_{i=1}^6 \tilde{e}_i) dt \right]^T \quad (\text{S.82})$$

Additionally (S.72), (S.77) become, respectively:

$$\sum_{i=1}^6 \tilde{e}_i \leq \tilde{E}_T \quad (\text{S.83})$$

$$v_1 = \tilde{e}_1 \cdot X_1 \cdot ATP$$

$$v_2 = \tilde{e}_2 \cdot X_2 \cdot NAD \cdot ADP$$

$$v_3 = \tilde{e}_3 \cdot X_3 \cdot NADH$$

$$v_4 = \tilde{e}_4 \cdot X_4 \cdot NAD$$

$$v_5 = \tilde{e}_5 \cdot X_3 \cdot NAD$$

$$v_6 = \tilde{e}_6 \cdot NADH \cdot ADP$$

$$v_7 = k_7 \cdot ATP$$

$$v_8 = k_8 \cdot NADH$$

(S.84)

This model's reactions correspond to a simple representation of the central carbon metabolism of yeast. The states X_{1-4} represent the key metabolites of the pathways included. In detail, X_1 stands for glucose and X_4 for ethanol, the two carbon sources between which the diauxic shift is occurring in the dynamic scenario taken into account. The triose phosphates in the glycolysis are represented by X_2 , as pyruvate is connecting glycolysis with the TCA cycle and ethanol formation/consumption

Fig. S42. Case SC0A: stimuli inference in the inverse optimal control problem. This solution corresponds to the noiseless case and scenario A, where we assume that we can measure the total enzyme concentration to be measured.

Fig. S43. Case SC0A: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S42

as state X_3 . Therefore in Figures S43 and S50 these states are shown by the names corresponding to the metabolites they represent. However, the inputs of the model do not correspond to a specific enzyme due to the fact that the reactions they are catalyzing are simplifications of whole pathways. Therefore in Figure S52 where a qualitative comparison between the model's predictions and real experimental measurements is shown, the model's inputs are compared with multiple real enzymes included in their corresponding pathway. Note that for e_6 there is no such correspondence and therefore no comparison is made.

1.10.2 Numerical results

This model presents a similar issues as in case LPN3B. In the dynamics, the inputs (enzyme concentrations e_i) are always found multiplied by the kinetic parameters (k_{1-6}). In the case of LPN3B, where the structural identifiability analysis was illustrated both analytically and numerically, we found serious problems due to this structure. It should be noted that in LPN3B the unknowns were three inputs, approximated with 3 piecewise constant elements, and 3 parameters, while the data were 21 time-points for five observables. Here, in SC, where we have 8 parameters and 6 inputs, approximated with 6 piecewise linear steps (aka 7 elements) and computed for 15 time-points of 8 observables, we expect such problems to

Fig. S44. Case SC0A: estimated inputs from the inverse optimal control problem being matched to the corresponding point in the Pareto front. This corresponds to the noiseless case (scenario A) optimal solution presented in Fig. S42

Table S.T5. Case SC: NRMSE for the noisy and noiseless cases.

Noise	Scenario	$NRMSE_y$	$NRMSE_u$	$NRMSE_\theta$
0%	A	0.001231	0.197015	0.740545
0%	B	0.007045	0.644559	1.563603
10%	-	0.129583	0.287440	0.642375

y:observables, u:stimuli, θ :parameters.

be even more serious. Therefore, in order to quickly reduce the problem ill-posedness we reformulated the model as stated above, i.e. by estimating the products of $\tilde{e}_i = k_i \cdot e_i$, for $i \in [1, 6]$. By doing this, we reduced the NLP size by 6 structurally unidentifiable unknowns.

For this case study, we considered data: (i) generated with the addition of 10% heteroscedastic proportional noise (SC10), and (ii)w noiseless (SC0). In Table S.T5 we summarize the normalized root mean square error for the solutions corresponding to these scenarios (NRMSE values were computed for observables and inputs). Since we are dealing with a synthetic problem of known solution, we can also use the NRMSE metric for the parameter inference.

As expected, in the noiseless case (SC0), there was no need to use regularization terms in the cost function. For illustrative purposes we divide this subcase into two scenarios. In scenario A we assume that we can measure the total enzyme concentration, while in scenario B we do not. Regarding scenario A, the controls (enzymes) inference is shown in Figures S42 while the fit to the observables is shown in Figure S43. The corresponding NRMSE values are given for observables, inputs and parameters in Figure S45 for both iterations of the solution (1st solution and re-optimization). The respective results for scenario B are presented in Figures S46 S47 and S48. Finally, in the same way as presented for LPN3B, in Figure S44 we are illustrating the final step of the IOCP-2, matching the solution taken from the input reconstruction with one of the points of the Pareto front resulted in the OCP and, i.e., correctly identifying the underlying optimality principle.

We note that for the noisy case (SC10) no regularization was used in the objective function (described by a log likelihood cost function). In Figure S49 we present the IOCP result for the inputs, while in Figures S50 and S57 we show the model fit to the observables and the convergence

Fig. S45. Case SC0A: normalized root mean square error with respect to the observables (y), the inputs (u) and the true parameters (θ), corresponding to the noiseless case optimal solution presented in Fig. S42

Fig. S47. Case SC0B: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S46

Fig. S46. Case SC0B: stimuli inference in the inverse optimal control problem. This solution corresponds to the noiseless case and scenario B, where we assume we can not measure the total enzyme concentration.

Fig. S48. Case SC0B: normalized root mean square error with respect to the observables (y), the inputs (u) and the true parameters (θ), corresponding to the noiseless case optimal solution presented in Fig. S46

curve, respectively. The corresponding NRMSE values with respect to all observables, inputs and parameters are given in Figure S51 and Table S.T5

As mentioned in the main paper, experimental time-series data are available that allow qualitative comparisons, as performed in previous works considering this case study as an optima control problem (Klipp *et al.*, 2002), (de Hijas-Liste *et al.*, 2014). Even though the scaling is very different, we can still perform a qualitative comparison between the enzymes of the model and the corresponding enzymes that have been measured, as was done in these previous papers. In figure S52 we can see that such a qualitative comparison exhibits a good agreement.

1.11 Computational requirements and scalability

The purpose of this paper is to introduce and illustrate the IOCP approach, the fundamental issues in this class of problems and how to handle them. We illustrate all the above using four different case studies, two for each type of IOCP as described previously in the manuscript. These case studies are of various size and complexity, so it is interesting to analyze their different requirements in terms of computational effort. We present the computation times required to obtain the solutions presented in this work in Table S.T6 while in Figures S53-S57 we present typical convergence

Fig. S49. Case SC10: stimuli inference in the inverse optimal control problem. This solution corresponds to the noisy case.

Fig. S50. Case SC10: fit to the observables corresponding to the noisy case optimal solution presented in Fig. S49

Fig. S51. Case SC10: normalized root mean square error with respect to the observables (y), the inputs (u) and the true parameters (θ) corresponding to the noisy case optimal solution presented in Fig. S49

curves for each case study. The rest of the convergence curve figures, corresponding to all subcases considered in this work, can be found in Zenodo. All the computations were carried out on a PC Intel Xeon E5-2630@2.30GHz using Matlab R2015b under Windows 7.

Table S.16 shows that the computation times required to solve case studies JAKSTAT and LPN3B were rather modest (order of minutes). Case study SC required computation times of 1 or more hours, as expected due to the larger network involved. Case study TSP was even more demanding because it considers 16 different experiments. Note however that multi-experiment settings can be easily parallelized. Overall, we believe that these results indicate a rather satisfactory scalability of our approach with problem size (Figure S53), especially considering that we have used a Matlab implementation of the framework (typically 1-2 orders of magnitude slower than compiled languages like C) and how challenging these problems are. Further, significant speed-ups can be obtained by adapting high performance computing strategies, including coarse and fine-grained parallelization strategies, as we have recently shown for the parameter estimation problem [Penas et al. \(2017\)](#).

Fig. S52. Case SC0A: qualitative comparison of the estimated inputs with relevant experimental enzyme data. Our estimated inputs (blue line), presented in Fig. S42 correspond to the left axis while the experimental data (black lines) correspond to the right axis.

Fig. S53. Computation (CPU) time ranges for the case studies considered in this paper as a function of degrees of freedom of the optimization problems. The different colors correspond to different subcases.

References

- Bellman, R. (1956). Dynamic programming and lagrange multipliers. *Proceedings of the National Academy of Sciences*, **42**(10), 767–769.
- Bertsekas, D. P. et al. (1995). *Dynamic programming and optimal control*. Athena Scientific Belmont, MA.
- Biegler, L. T. et al. (2002). Advances in simultaneous strategies for dynamic process optimization. *Chemical Engineering Science*, **57**(4), 575–593.

Table S.T6. Computation (CPU) times per subcase. Note that the reported CPU time corresponds to the time needed to converge. Notation: Dof : degrees of freedom, N_{st} : number of states, N_u : number of inputs, N_θ : number of unknown parameters to be estimated, N_{exp} : number of experiments, CPU_{range} : computation time range. The computations were carried out on a PC Intel Xeon E5-2630@2.30GHz using Matlab R2015b under Windows 7

Subcase	Dof	N_{st}	N_u	N_θ	N_{exp}	$CPU_{range}(min)$
JAKSTAT0	21	9	1	4	1	10-30
JAKSTAT5	13	9	1	4	1	1-4
JAKSTATreal	13	9	1	4	1	1-4
LPN3B0	14	6	3	3	1	1-20
LPN3B10	14	6	3	3	1	1-20
SC0A	44	11	6	2	1	60-360
SC0B	44	11	6	2	1	40-141
SC10	44	11	6	2	1	15-90
TSP0	68	8	2	36	16	45-250
TSP5	68	8	2	36	16	494-908

Fig. S54. Case JAKSTATreal: Convergence curve for the reported solution corresponding to the real experimental data.

Fig. S55. Case TSP5: Convergence curve for the reported noisy solution.

Fig. S56. Case LPN3B10: Convergence curve for the reported noisy solution.

Bryson, A. E. (1975). *Applied optimal control: optimization, estimation and control*. CRC Press.

Chis, O. et al. (2011a). GenSSI: a software toolbox for structural identifiability analysis of biological models. *Bioinformatics*, **27**(18), 2610–2611.

Chis, O. et al. (2011b). Structural identifiability of systems biology models: A critical comparison of methods. *Plos One*, **6**(11), e27755.

de Hijas-Liste, G. M. et al. (2014). Global dynamic optimization approach to predict activation in metabolic pathways. *BMC systems biology*, **8**(1), 1.

Klipp, E. et al. (2002). Prediction of temporal gene expression. *European journal of biochemistry*, **269**(22), 5406–5413.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, **142**(2), 573–603.

Liberzon, D. (2012). *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press.

Moles, C. G. C. et al. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research*, **13**(11), 2467–74.

Penas, D. R. et al. (2017). Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy. *BMC Bioinformatics*, **18**(1), 52.

Pohjanpalo, H. (1978). System identifiability based on power-series expansion of solution. *Math. Biosci.*, **41**(1-2), 21–33.

Raue, A. et al. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**(15), 1923–1929.

Rodriguez-Fernandez, M. et al. (2006). A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, **83**(2), 248 – 265. 5th International Conference on Systems Biology.

Schelker, M. et al. (2012). Comprehensive estimation of input signals and dynamics in biochemical reaction networks. *Bioinformatics*, **28**(18), i529–i534.

Vanlier, J. et al. (2013). Parameter uncertainty in biochemical models described by ordinary differential equations. *Mathematical biosciences*, **246**(2), 305–314.

Vassiliadis, V. et al. (1994a). Solution of a class of multistage dynamic optimization problems. 2. problems with path constraints. *Industrial & Engineering Chemistry Research*, **33**(9), 2123–2133.

Vassiliadis, V. S. et al. (1994b). Solution of a class of multistage dynamic optimization problems. 1. problems without path constraints. *Industrial & Engineering Chemistry Research*, **33**(9), 2111–2122.

Fig. S57. Case SC10: Convergence curve for the reported noisy solution.

Zaslaver, A. *et al.* (2004). Just-in-time transcription program in metabolic pathways. *Nature genetics*, **36**(5), 486–491.