Systems biology

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

Nikolaos Tsiantis [1,2], Eva Balsa-Canto [1] and Julio R. Banga [1,*]

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

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \tilde{\mathbf{\Psi}}[\mathbf{x}\{t,\mathbf{p}\},\mathbf{u}\{t\},\mathbf{p},t], \quad \mathbf{x}\{t_0,\mathbf{p}\} = \mathbf{x_0} \tag{S.1}$$

$$\boldsymbol{\eta}[\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}] = 0 \tag{S.2}$$

$$\boldsymbol{\zeta}[\mathbf{x}(t,\mathbf{p}),\mathbf{u}(t),\mathbf{p}] \leq 0 \tag{S.3}$$

$$\boldsymbol{\eta_\iota}[\mathbf{x}(t_\iota,\mathbf{p}),\mathbf{u}(t_\iota),\mathbf{p}] = 0 \tag{S.4}$$

$$\boldsymbol{\zeta_\iota}[\mathbf{x}(t_\iota,\mathbf{p}),\mathbf{u}(t_\iota),\mathbf{p}] \leq 0 \tag{S.5}$$

where $\tilde{\mathbf{\Psi}}$ is the right-hand side of the ordinary differential equations (ODEs) (S.1) describing the dynamics of the states $\mathbf{x}$ given initial conditions $\mathbf{x_0}$. 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 $\boldsymbol{\eta}$ and inequalities $\boldsymbol{\zeta}$ which must be enforced the whole time horizon considered. Similarly, time-point constraints can be equalities $\boldsymbol{\eta_\iota}$ and inequalities $\boldsymbol{\zeta_\iota}$ only to be enforced at a specific point of time t_ι.

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 $\mathbf{p}$ to minimize

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

Subject to:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \tilde{\mathbf{\Psi}}[\mathbf{x}\{t,\mathbf{p}\},\mathbf{u}\{t\},\mathbf{p},t], \quad \mathbf{x}\{t_0,\mathbf{p}\} = \mathbf{x_0} \tag{S.7}$$

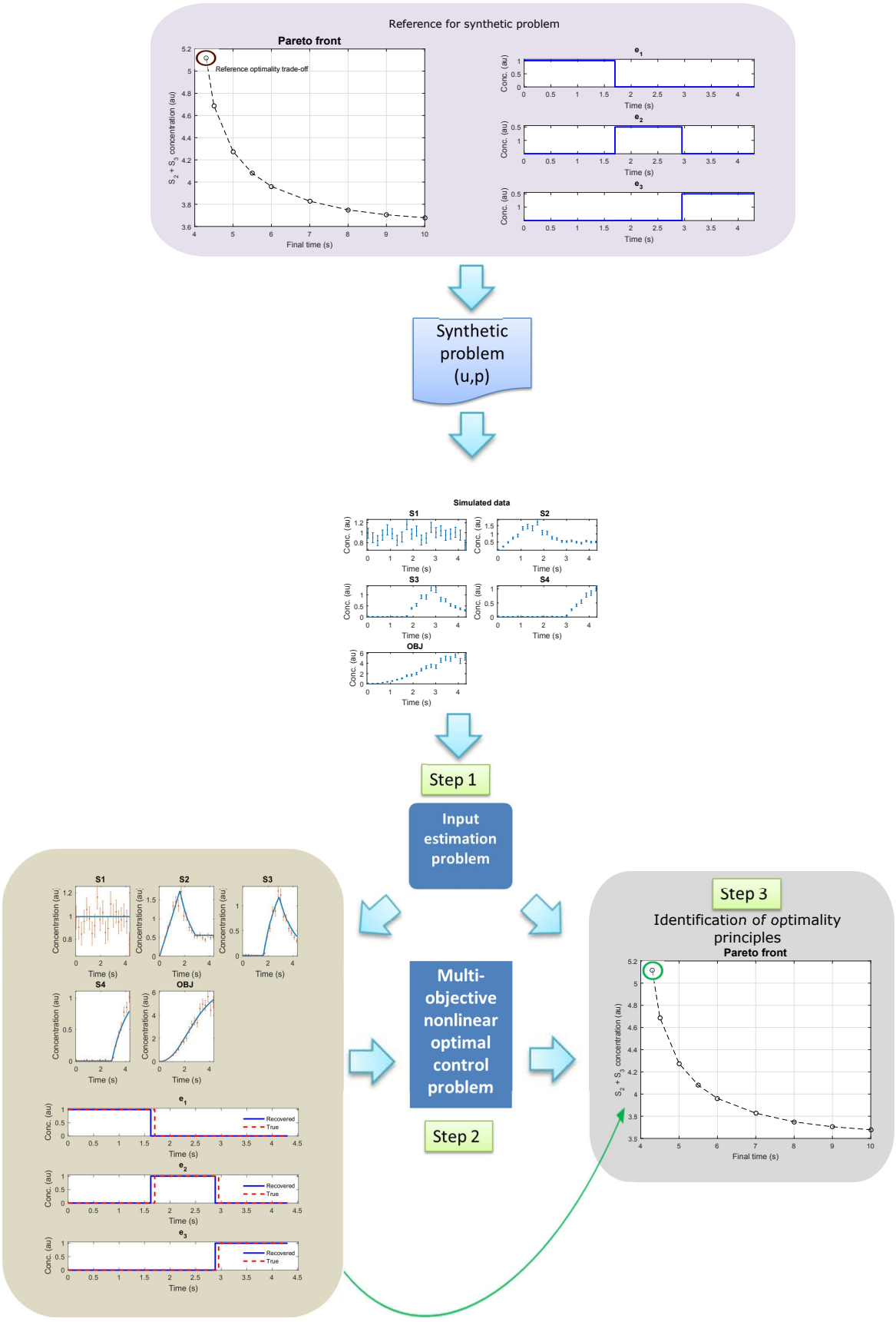

Fig. S1. Diagram of the overall inverse optimal control solution strategy (the illustrations correspond to the particular case of LPN3B synthetic problem).

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

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

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

$$\mathbf{p}^L \leq \mathbf{p} \leq \mathbf{p}^U \tag{S.11}$$

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), $\tilde{y}$ are the experimental measurements of a subset of the (so-called) output state variables, $y(\mathbf{p},t)$ are the model predictions for those outputs, $W(t)$ is a weighting (or scaling) matrix and $\mathbf{x}$ are the differential state variables. Besides, $\tilde{\boldsymbol{\Psi}}$ 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:

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

Additionally we consider bounds for the inputs of the form:

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

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:

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

Subject to:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \tilde{\boldsymbol{\Psi}}[\mathbf{x}\{t,\mathbf{p}\},\mathbf{u}\{t\},\mathbf{p},t], \\ \mathbf{x}\{t_0,\mathbf{p}\} &= \mathbf{x_0} \end{aligned} \tag{S.15}$$

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

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

$$\boldsymbol{\eta_\iota}[\mathbf{x}(t_\iota,\mathbf{p}),\mathbf{u}(t_\iota),\mathbf{p}] = 0 \tag{S.18}$$

$$\boldsymbol{\zeta_\iota}[\mathbf{x}(t_\iota,\mathbf{p}),\mathbf{u}(t_\iota),\mathbf{p}] \leq 0 \tag{S.19}$$

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

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 ($\mathbf{J}[\mathbf{x},\mathbf{u}]$), subject to the system's dynamics as well as the imposed constraints. The objective functional $\mathbf{J}[\mathbf{x},\mathbf{u}]$ corresponds to the optimality criteria hypothesis. In the case of a multicriteria formulation, the cost functional $\mathbf{J}[\mathbf{x},\mathbf{u}]$ is a set of objective functions corresponding to the N different criteria considered:

$$\mathbf{J}[\mathbf{x},\mathbf{u},\mathbf{p}] = \begin{bmatrix} \mathrm{J}_1[\mathbf{x},\mathbf{u},\mathbf{p}] \\ \mathrm{J}_2[\mathbf{x},\mathbf{u},\mathbf{p}] \\ . \\ . \\ . \\ \mathrm{J}_\mathrm{N}[\mathbf{x},\mathbf{u},\mathbf{p}] \end{bmatrix} \tag{S.21}$$

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:

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

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, 1975; 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:

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

being τ the normalized time in each element i:

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

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 (Pohjanpalo, 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:

$$\dot{x_1} = 0 \tag{S.25}$$

$$\dot{x_2} = k_1 u x_1 - k_2 x_2 \tag{S.26}$$

$$y = x_2 \tag{S.27}$$

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:

$$T_0 = 0; \tag{S.28}$$

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

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

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

$$T_4 = k_1 u^3(0) - k_1 k_2 u^2(0) + k_1 k_2^2 u^1(0) - k1 k_2^3 u(0) \tag{S.32}$$

$$\dots \tag{S.33}$$

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

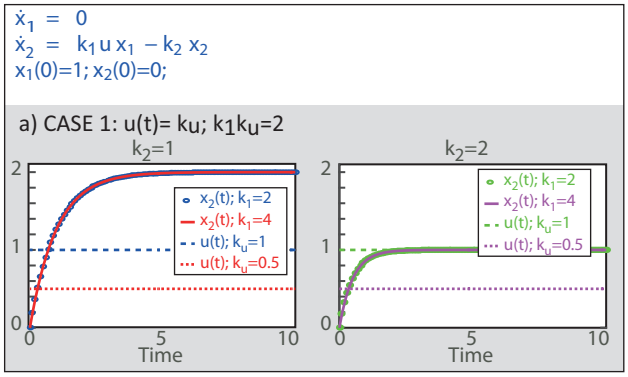

Fig. S2. Illustrative examples of lack of structural identifiability under sustained stimulation. Figures correspond to the controls and the corresponding observables in two different scenarios: i) $k_u = 0.5; k_1 = 4; k_2 = 1$ (red) and $k_u = 1; k_1 = 2; k_2 = 1$ (blue); and ii) $k_u = 0.5; k_1 = 4; k_2 = 2$ (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.

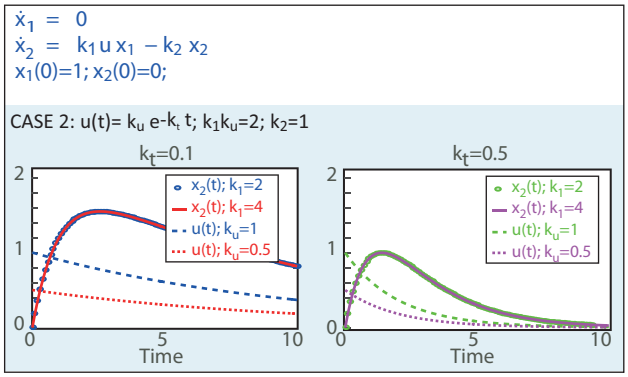

Fig. S3. Illustrative examples of lack of structural identifiability under exponential stimulation. Figures correspond to the controls and the corresponding observables in two different scenarios: i) $k_u = 0.5; k_1 = 4; k_t = 0.1$ (red) and $k_u = 1; k_1 = 2; k_t = 0.1$ (blue); and ii) $k_u = 0.5; k_1 = 4; k_t = 0.5$ (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 probed 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.

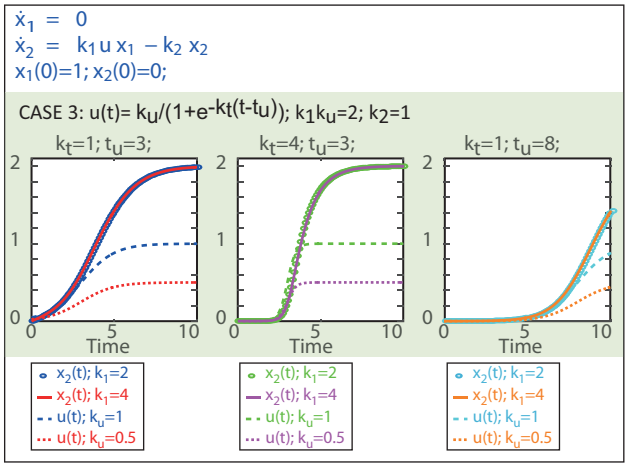

Fig. S4. Illustrative examples of lack of structural identifiability under logistic stimulation. Figures correspond to the controls and the corresponding observables in three different scenarios each: i) $k_u = 0.5; k_1 = 4; k_t = 1; t_u = 3$ (red) and $k_u = 1; k_1 = 2; k_t = 1; t_u = 3$ (blue); ii) $k_u = 0.5; k_1 = 4; k_t = 4; t_u = 3$ (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:

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

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:

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

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_{udata} 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:

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

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:

$$\tilde{y}_{o,s} = y_{o,s} + \epsilon_{o,s} \tag{S.37}$$

with :

$$\epsilon_{o,s} = \sigma \times randC_{o,s} \times y_{e,o} \tag{S.38}$$

where $\epsilon_{o,s}$ are normally distributed independent random variables with standard deviation $\sigma \times y_{e,o}$ and $randC_{o,s}$ 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.

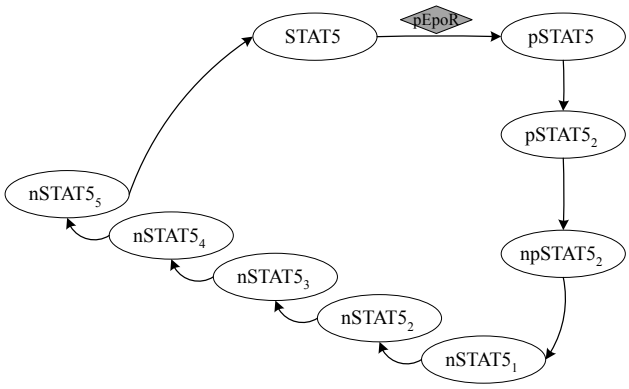

Fig. S5. Case JAKSTAT: Network representation for the JAK-STAT case study.

1.7.1 Problem formulation

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

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

Where:

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

Subject to:

$$\begin{aligned}
\frac{dSTAT5}{dt} &= p_4 \cdot nSTAT5_5 \cdot \frac{0.45}{1.4} - p_1 \cdot pEpoR \cdot STAT5 \\
\frac{dpSTAT5}{dt} &= p_1 \cdot pEpoR \cdot STAT5 - 2 \cdot p_2 \cdot pSTAT5^2 \\
\frac{dpSTAT5_2}{dt} &= 2 \cdot p_2 \cdot pSTAT5^2 - p_3 \cdot pSTAT5_2 \\
\frac{dnpSTAT5_2}{dt} &= p_3 \cdot pSTAT5_2 \cdot \frac{1.4}{0.45} - p_4 \cdot npSTAT5_2 \\
\frac{dnSTAT5_1}{dt} &= 2 \cdot p_4 \cdot npSTAT5_2 - p_4 \cdot nSTAT5_1 \\
\frac{dnSTAT5_2}{dt} &= p_4 \cdot nSTAT5_1 - p_4 \cdot nSTAT5_2 \\
\frac{dnSTAT5_3}{dt} &= p_4 \cdot nSTAT5_2 - p_4 \cdot nSTAT5_3 \\
\frac{dnSTAT5_4}{dt} &= p_4 \cdot nSTAT5_3 - p_4 \cdot nSTAT5_4 \\
\frac{dnSTAT5_5}{dt} &= p_4 \cdot nSTAT5_4 - p_4 \cdot nSTAT5_5
\end{aligned} \tag{S.41}$$

$$y_1(\mathbf{x}, \mathbf{p}) = offset\_tSTAT5 + scale\_tSTAT5 \cdot (STAT5 + pSTAT5 + 2 \cdot pSTAT5_2) \tag{S.42}$$

$$y_2(\mathbf{x}, \mathbf{p}) = offset\_pSTAT5 + scale\_pSTAT5 \cdot (pSTAT5 + 2 \cdot pSTAT5_2) \tag{S.43}$$

Where the states vector is:

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

And the initial conditions are:

$$\mathbf{x}(t_0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{S.45}$$

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 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, $EpoA$, $Epot$, $EpoS$].

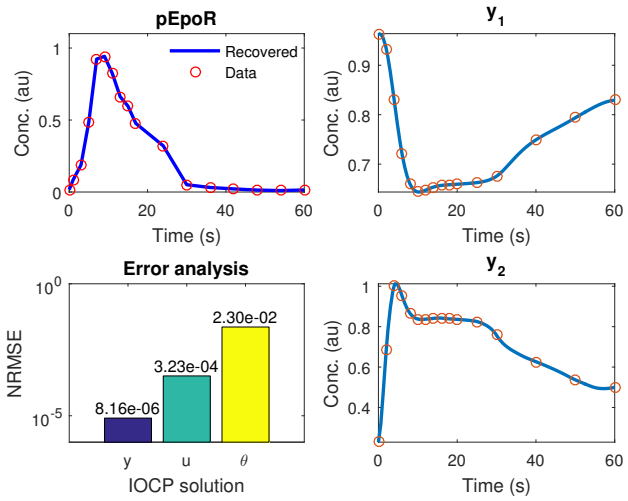

Fig. S6. Case JAKSTAT0: estimated versus experimental stimuli measurements, along with the fit for the observed states. The overall quality of the estimation is given as normalized root mean square error (NRMSE) for the observables (y), the stimuli (u) and the parameters (θ). This solution corresponds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

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

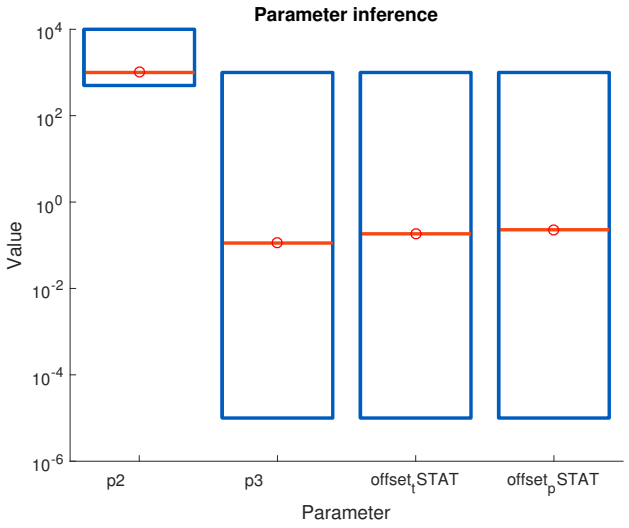

Fig. S7. Case JAKSTAT0: parameter inference in the inverse optimal control problem. The box corresponds to the bounds used in the estimation, the orange dash to the true value used to generate the synthetic data and the orange dot to the estimate. This solution corresponds to the synthetic noiseless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

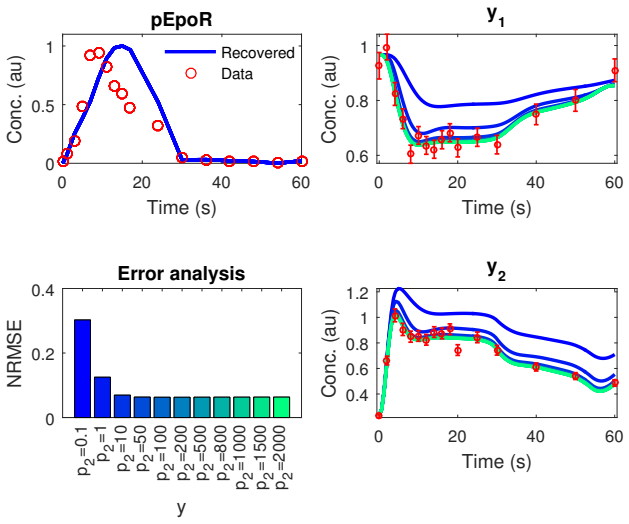

Fig. S8. Case JAKSTAT5: illustration of practical non-identifiability in p_2 simulating a local solution ($pEpoR$ and $\mathbf{p}$) for different values of p_2 and showing how it affects the observables y_1 and y_2. This corresponds to the synthetic noisy case.

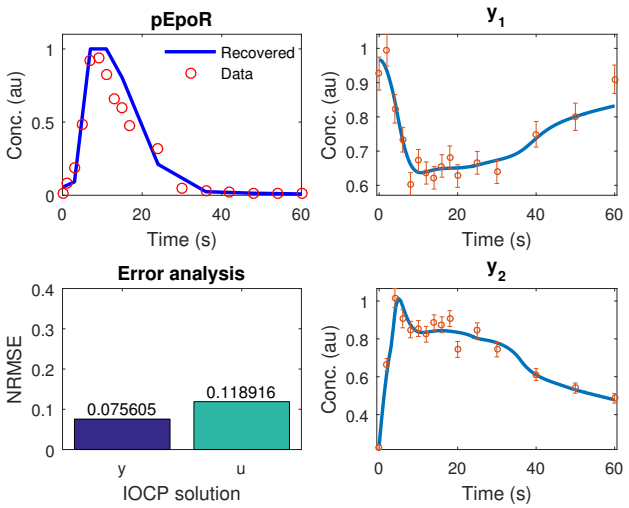

Fig. S9. Case JAKSTAT5: estimated versus experimental stimuli measurements, along with the fit for the observed states. The overall quality of the estimation is given as normalized root mean square error (NRMSE) for the observables (y) and the stimuli (u). This solution corresponds to the synthetic noisy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

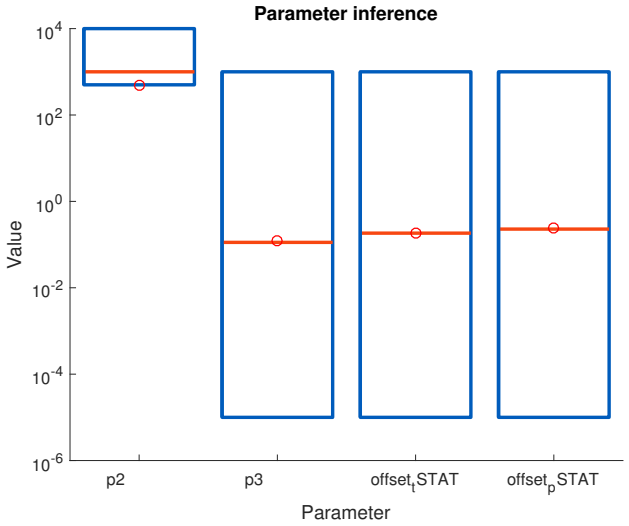

Fig. S10. Case JAKSTAT5: parameter inference in the inverse optimal control problem. The box corresponds to the bounds used in the estimation, the orange dash illustrates the true value used to generate the synthetic data and the orange dot shows to the estimate. This solution corresponds to the synthetic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

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 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 generelization 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

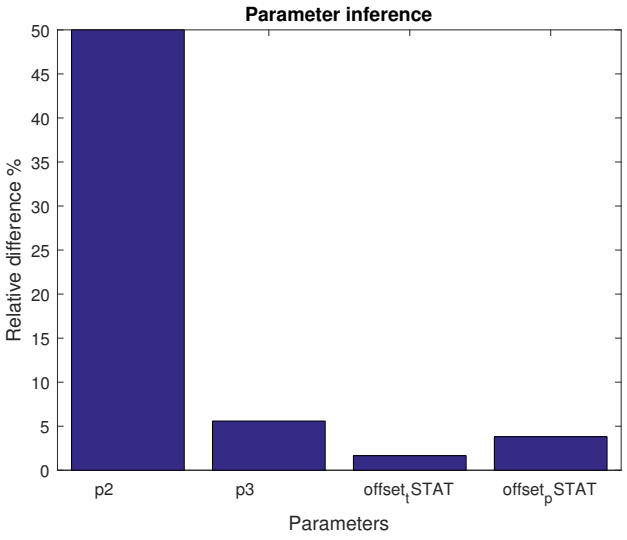

Fig. S11. Case JAKSTAT5: parameter inference in the inverse optimal control problem. Relative difference between the true values used to generate the synthetic data and the estimates. This solution corresponds to the synthetic noisy case.

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.

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 soolving he 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.

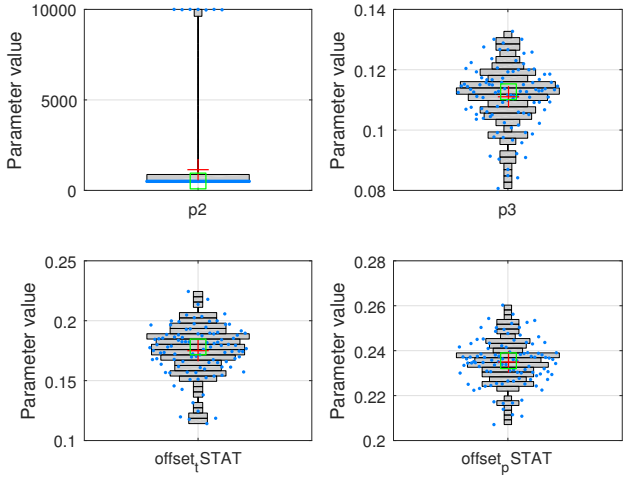

Fig. S12. Uncertainty analysis for JAKSTAT5: violin plots depicting the distribution of the parameters found using the bootstrap approach.

1.8.1 Problem formulation

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

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

Where:

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

Subject to:

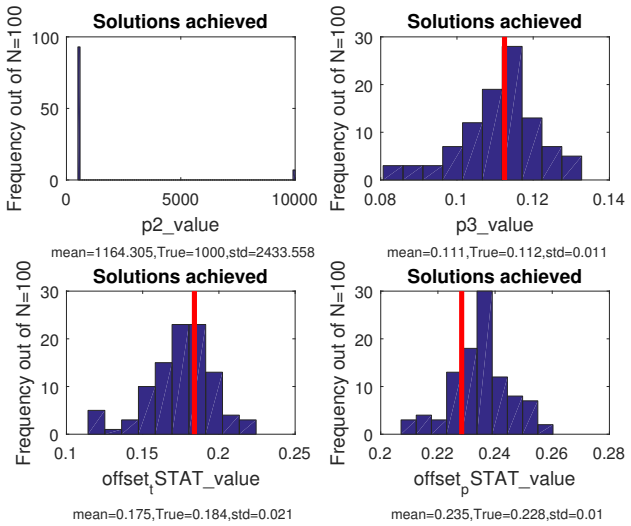

Fig. S13. Uncertainty analysis for JAKSTAT5: histograms depicting the distribution of the parameters resulting from the bootstrap approach. The true value of the parameters used to generate all data realizations is marked with the red vertical line. Note the bi-modality of p_2 due to its lack of identifiability.

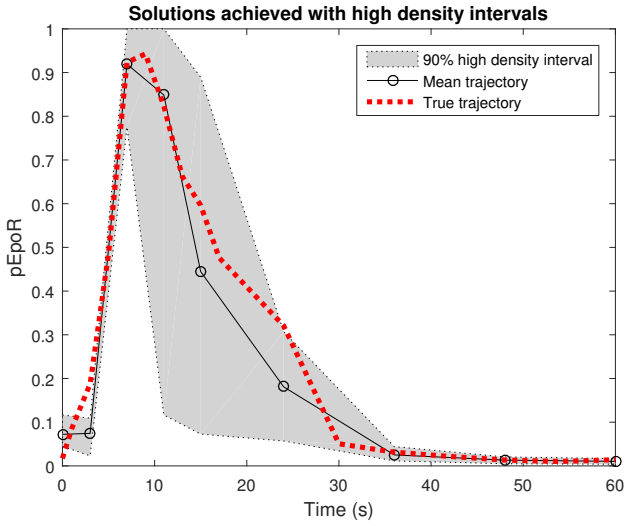

Fig. S14. Uncertainty analysis for JAKSTAT5: comparison between the mean and the true trajectories of the time-dependent input resulting from the bootstrap approach. The high density interval (90% HDI) is represented as the gray envelope.

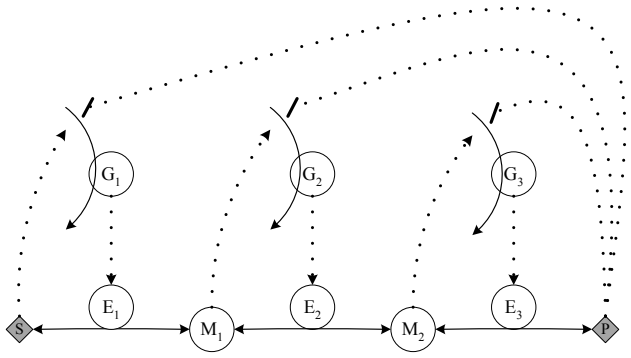

Fig. S15. Case TSP: Network representation for the TSP case study.

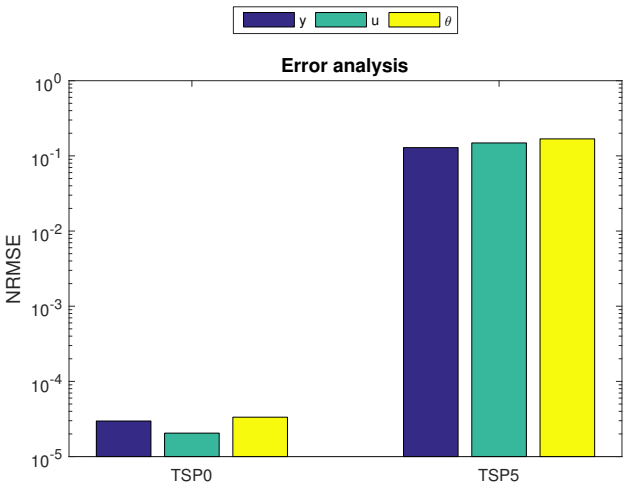

Fig. S16. Case TSP: goodness of the estimation for the two subcases, TSP0 (noiseless) and TSP5 (5% noise) in terms of normalized root mean square error (NRMSE) of the observables (y), the stimuli (u) and the parameters (θ). The observables, stimuli and parameters are described in (S.49),(S.51),(S.52).

$$
\begin{aligned}
\frac{dG_1}{dt} &= \frac{V_1}{1+(\frac{P}{Ki_1})^{ni_1}+(\frac{Ka_1}{S})^{na_1}} - k_1 \cdot G_1 \\
\frac{dG_2}{dt} &= \frac{V_2}{1+(\frac{P}{Ki_2})^{ni_2}+(\frac{Ka_2}{M1})^{na_2}} - k_2 \cdot G_2 \\
\frac{dG_3}{dt} &= \frac{V_3}{1+(\frac{P}{Ki_3})^{ni_3}+(\frac{Ka_3}{M_2})^{na_3}} - k_3 \cdot G_3 \\
\frac{dE_1}{dt} &= \frac{V_4 \cdot G_1}{K_4+G_1} - k_4 \cdot E_1 \\
\frac{dE_2}{dt} &= \frac{V_5 \cdot G_2}{K_5+G_2} - k_5 \cdot E_2 \\
\frac{dE_3}{dt} &= \frac{V_6 \cdot G_2}{K_6+G_3} - k_6 \cdot E_3 \\
\frac{dM_1}{dt} &= \frac{kcat_1 \cdot E_1 \cdot (\frac{1}{Km_1}) \cdot (S-M_1)}{1+\frac{S}{Km_1}+\frac{M1}{Km_2}} \\
&\quad - \frac{kcat_2 \cdot E_2 \cdot (\frac{1}{Km_3}) \cdot (M_1-M_2)}{1+\frac{M1}{Km_3}+\frac{M2}{Km_4}} \\
\frac{dM_2}{dt} &= \frac{kcat_2 \cdot E_2 \cdot (\frac{1}{Km3}) \cdot (M_1-M_2)}{1+\frac{M_1}{Km_3}+\frac{M_2}{Km_4}} \\
&\quad - \frac{kcat_3 \cdot E_3 \cdot (\frac{1}{Km_5}) \cdot (M_2-P)}{1+\frac{M_2}{Km_5}+\frac{P}{Km_6}}
\end{aligned}
\tag{S.48}
$$

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

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:

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

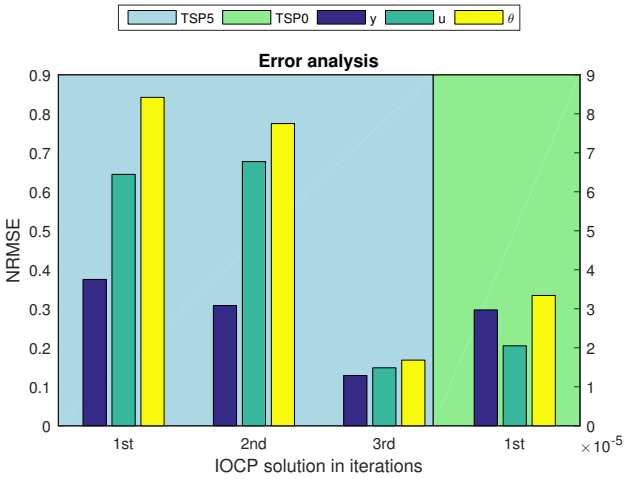

Fig. S17. Case TSP: normalized root mean square error for noisy and noiseless case, with respect to the observables (y), the inputs (u) and the true parameters (θ). TSP5 corresponds to the left y-axis while TSP0 to the right one. The units of the latter are presented scaled in 10^{-5}. The observables,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

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

And the initial conditions are:

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

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 (TPS5) and noiseless (TPS0) 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 NRME 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 the their true values (since the 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).

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.

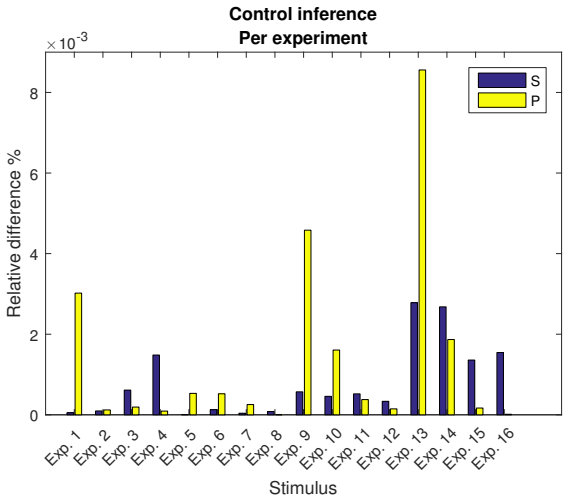

Fig. S18. Case TSP0: stimuli inference per experiment. In this case study, synthetic multi-experimental datasets were used for the inverse optimal control problem. Here the relative difference between the true input values used to generate the noiseless data for each experiment and the input estimates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

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.

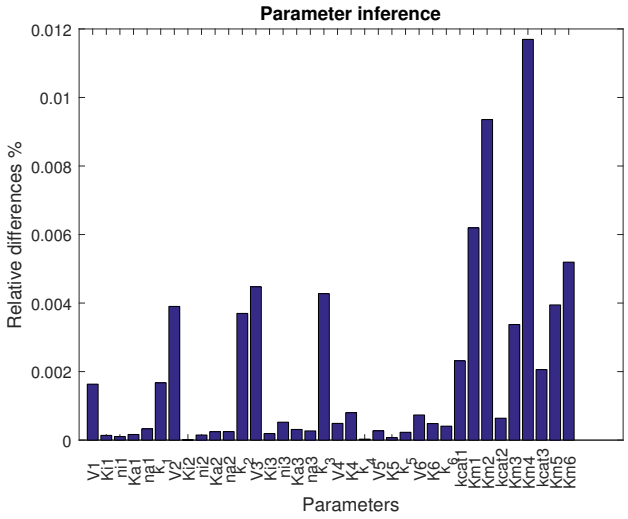

Fig. S19. Case TSP0: parameter inference in the inverse optimal control problem as relative difference between the true values used to generate the synthetic noiseless data and the estimates.

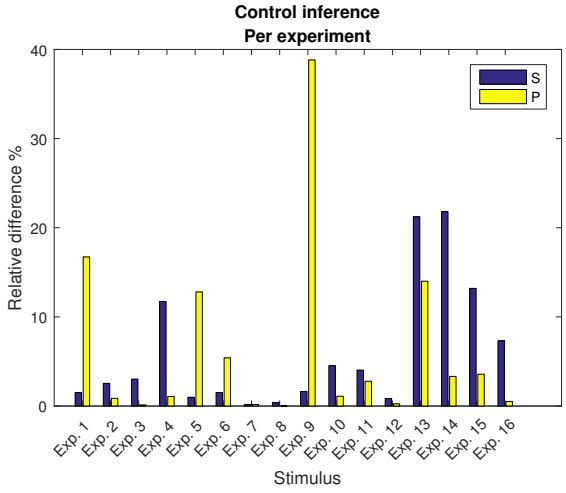

Fig. S20. Case TSP5: stimuli inference per experiment. In this case study, synthetic multi-experimental datasets were used for the inverse optimal control problem. Here the relative difference between the true input values used to generate the noisy data for each experiment and the input estimates is analyzed, per experiment, for both inputs.

1.9 Case study LPN3B

This case study is given as a first illustrative example of the more general IOCP-2 problem class. The problem is a generalization of the one studied by de Hijas-Liste *et al.* (2014), where it was considered as a standard optimal control problem (OCP). Here, we take the solution reference of the inner problem as the multi-objective OCP described in de Hijas-Liste *et al.* (2014), selecting a specific point of the resulting Pareto front. The nature of this case study illustrates well the ability of our approach to consider dynamic models with path constraints on the states and the inputs. In Figure S24 we show the network representation considered.

1.9.1 Problem formulation

The mathematical formulation of inverse optimal control problem is stated below:

$$\min_{\mathbf{e}(t),\mathbf{k},t_f} \mathrm{J}_{\mathrm{outer}}[\mathbf{S},\mathbf{e},\mathbf{k}] \tag{S.54}$$

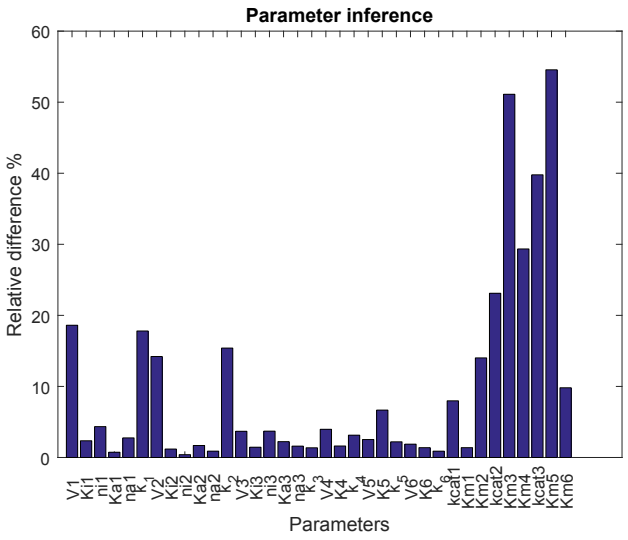

Fig. S21. Case TSP5: parameter inference as relative difference between the true values used to generate the synthetic noisy data and the estimates.

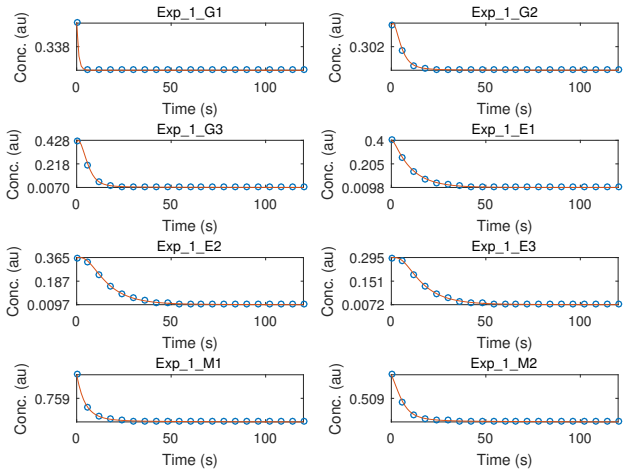

Fig. S22. Case TSP0: example of resulting fit for the observables (experiment 1; the results for the other experiments are included as part of the software distribution).

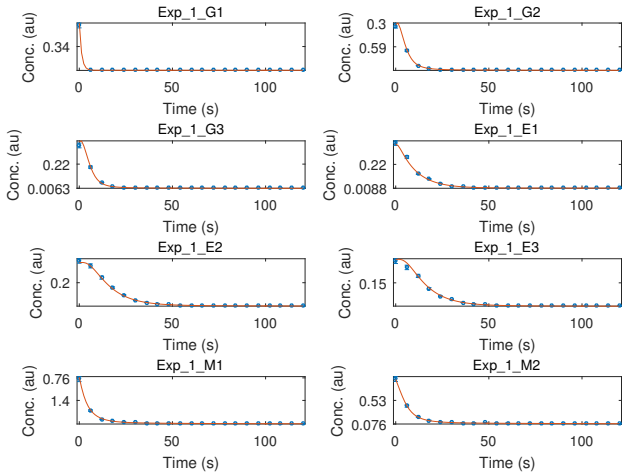

Fig. S23. Case TSP5: example of resulting fit for the observables (experiment 1; the results for the other experiments are included as part of the software distribution).

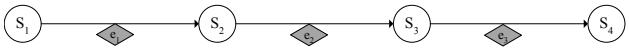

Fig. S24. Case LPN3B: Network representation for the LPN3B case study.

Where:

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

Subject to:

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

Where:

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

Subject to the system dynamics:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\boldsymbol{\upsilon} \tag{S.58}$$

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

Where the states' vector is:

$$\mathbf{S} = [S_1, S_2, S_3, S_4] \tag{S.60}$$

While the stoicheiometric matrix **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:

$$\upsilon_\mathbf{i} = k_i \cdot S_i \cdot e_i \tag{S.61}$$

With the following end-point constraint:

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

and path constraint:

$$\sum_{i=1}^{3} e_i \leq E_T \tag{S.63}$$

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.* (2011a,b), available at github.com/genssideveloper/GenSSI).

We first needed to parameterize the solution for the activation profiles for the different enzymes $e_i(t)$, $i = 1\ldots3$. 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:

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

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:

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

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. $k_{ei} = k_{ei1} = k_{ei2}$, $i = 1\ldots3$; 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

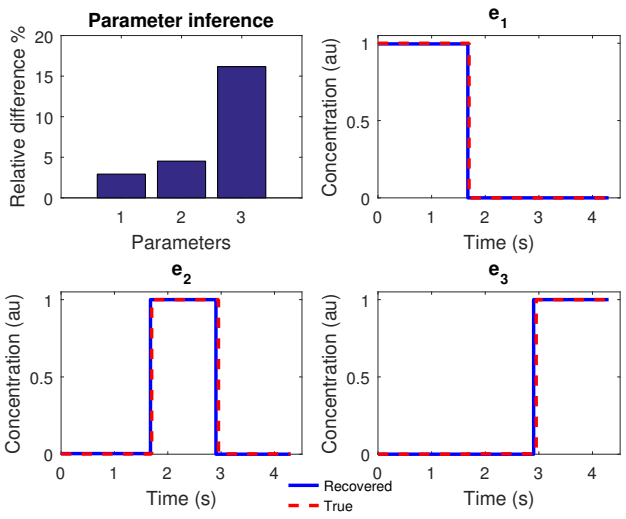

Fig. S25. Case LPN3B0: stimuli inference in 3rd re-optimization of the inverse optimal control problem and the relative difference between the true parameter values used to generate the synthetic noiseless data and the estimates.

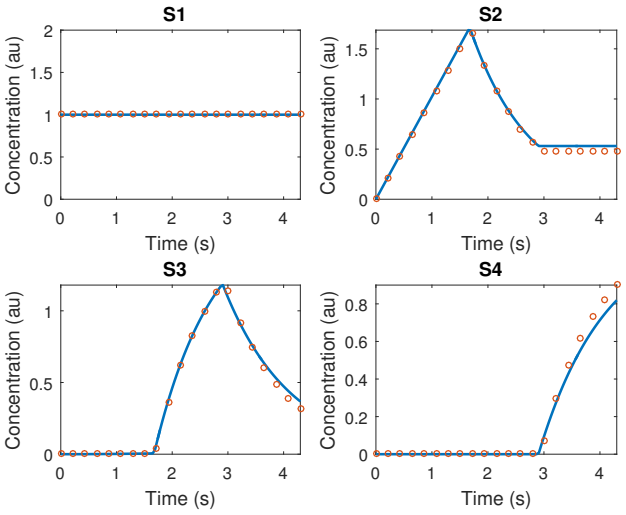

Fig. S26. Case LPN3B0: fit for the observables in 3rd re-optimization of the inverse optimal control problem for the noiseless case.

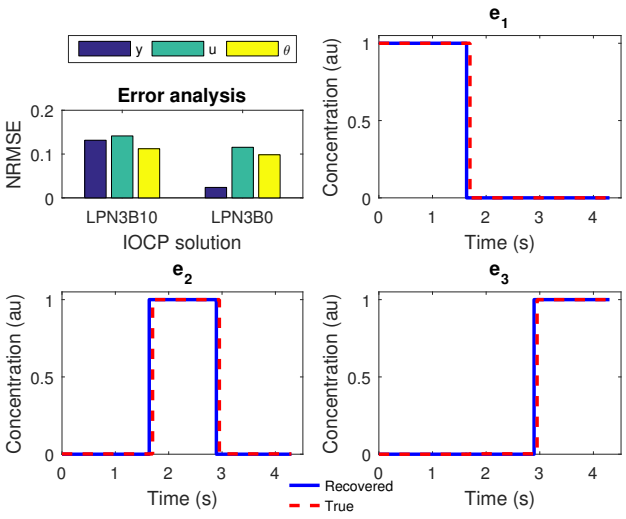

Fig. S27. Case LPN3B10: goodness of the estimation for the noisy and noiseless subcase in comparison, showing the normalized root mean square error and the reconstructed stimuli in the noisy subcase (versus the true value used for the pseudo-experimental noisy data generation).

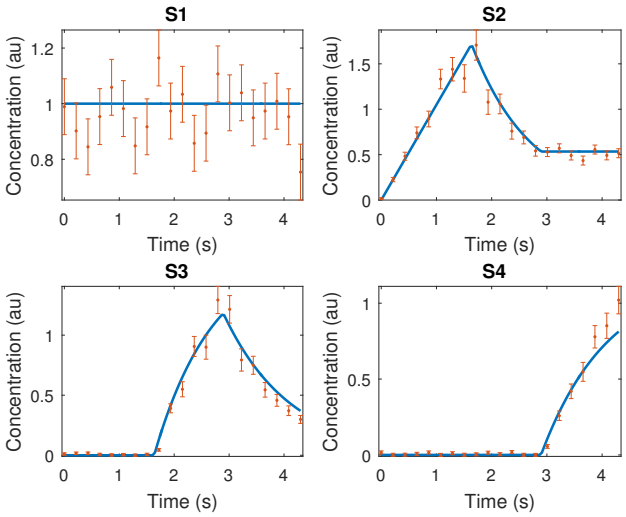

Fig. S28. Case LPN3B10: fit to the observables in the inverse optimal control problem for the noisy case.

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 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

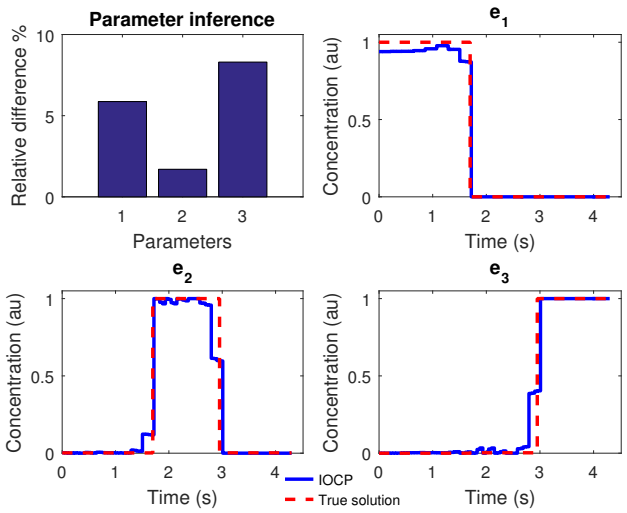

Fig. S29. Case LPN3B0re: stimuli inference in re-optimization 4_3 and the relative difference between the true parameter values used to generate the synthetic noiseless data and the estimates. This subcase corresponds to the solution of the noiseless case without any initial assumption of input discretization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

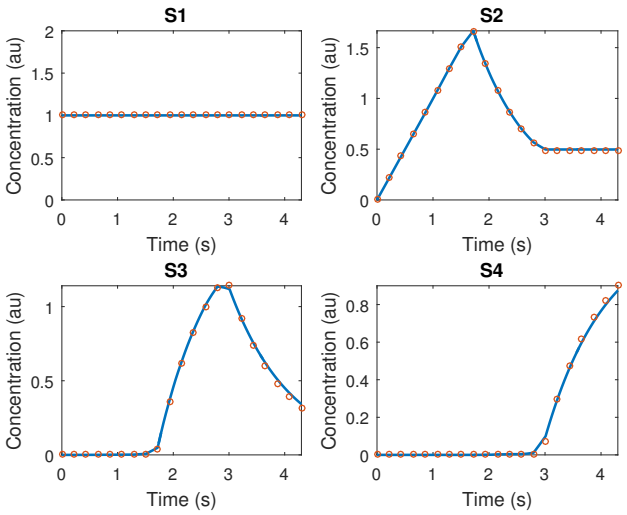

Fig. S30. Case LPN3B0re: fit on the observables in re-optimization 4_3 in the inverse optimal control problem. This subcase corresponds to the solution of the noiseless case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

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.

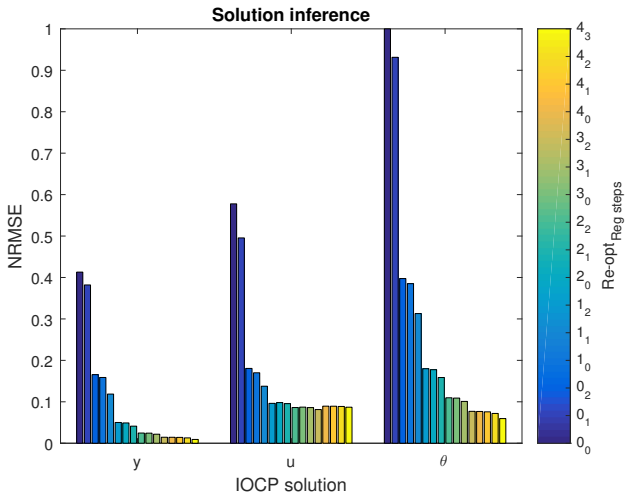

Fig. S31. Case LPN3B0re: Quality of fit in the solution of the inverse optimal control problem presented in Fig. S29, corresponding to the noiseless case, in terms of normalized root mean square (with respect to the observables (y), inputs (u) and parameters (θ))and how it changes by the iterative use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

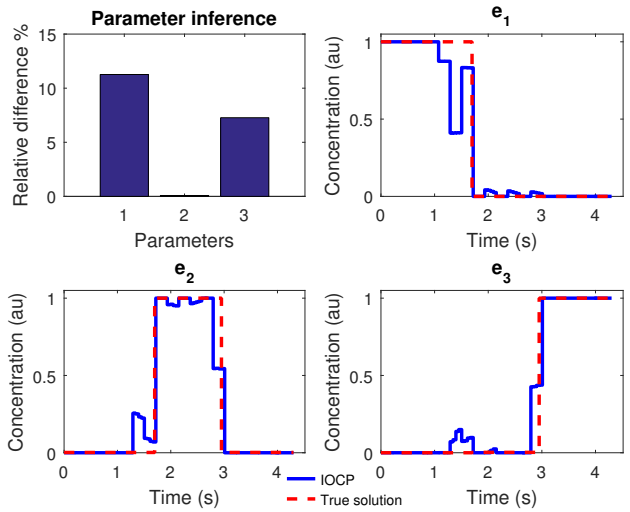

Fig. S32. Case LPN3B10re: stimuli inference in re-optimization 4_3 and the relative difference between the true parameter values used to generate the synthetic noisy data and the estimates. This subcase corresponds to the solution of the noisy case without any initial assumption of input discretization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

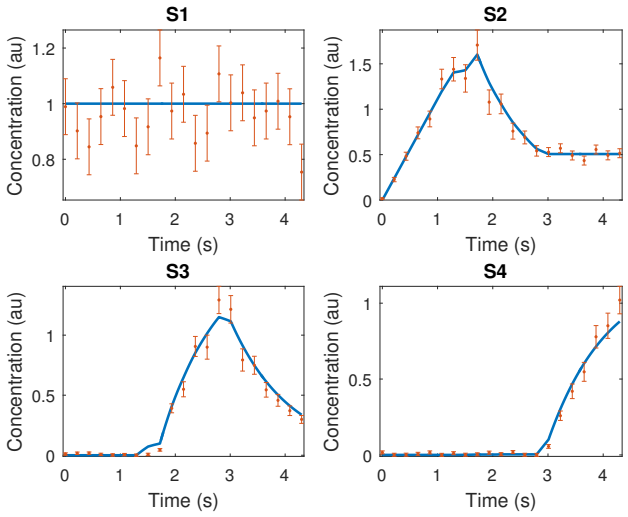

Fig. S33. Case LPN3B10re: fit on the observables in re-optimization 4_3 in the inverse optimal control problem. This subcase corresponds to the solution of the noisy case without any initial assumption of input discretization and iteratively using regularized re-optimization as described in-text.

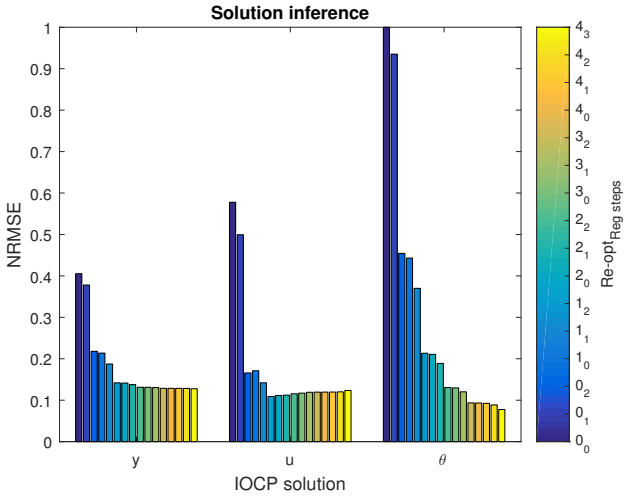

Fig. S34. Case LPN3B10re: Quality of fit in the solution of the inverse optimal control problem presented in Fig. S32, corresponding to the noisy case, in terms of normalized root mean square (with respect to the observables (y), inputs (u) and parameters (θ))and how it changes by the iterative use of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

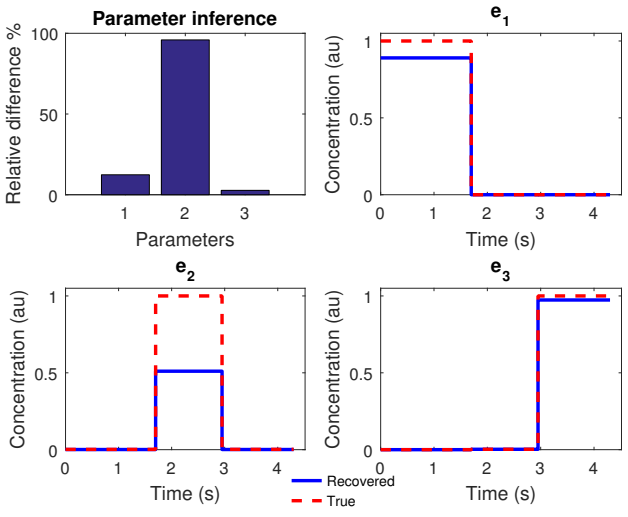

Fig. S35. Case LPN3B0noreg: two different optimal solutions with the same quality of fit to the observables as a result of lack of structural identifiability. The red dashed line is the true solution used to generate the synthetic noiseless data and the blue line is the estimate computed without the 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.

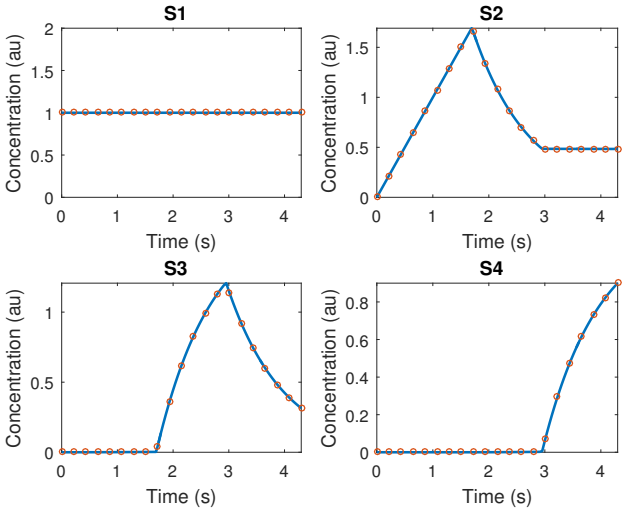

Fig. S36. Case LPN3B0noreg: fit to the observables for the optimal solutions depicted in Fig.S35.

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, LPN3B0p2l2 and LPN3B0norege2), 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 respecto 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 LPN3B0norege2 we show how introducing prior information (in the bounds and using regularization)

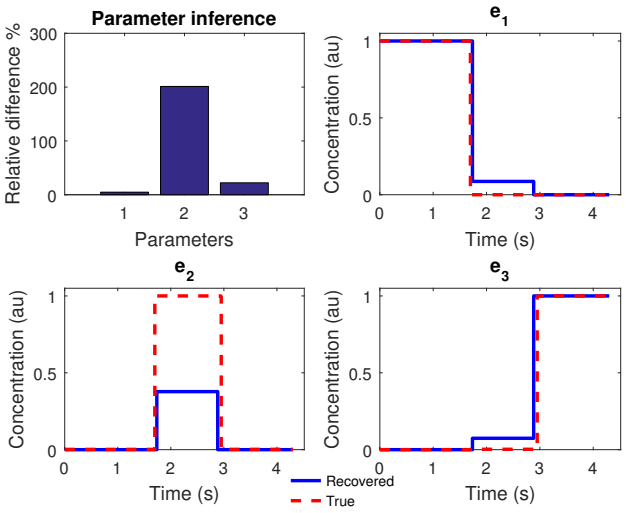

Fig. S37. Case LPN3B0p2l2: two different optimal enzyme profiles which result in the same quality of fit. Since the optimal p_2 value is twice its true value, the computed enzyme profiles compensate for this modification. This result illustrates the lack of structural identifiability and corresponds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

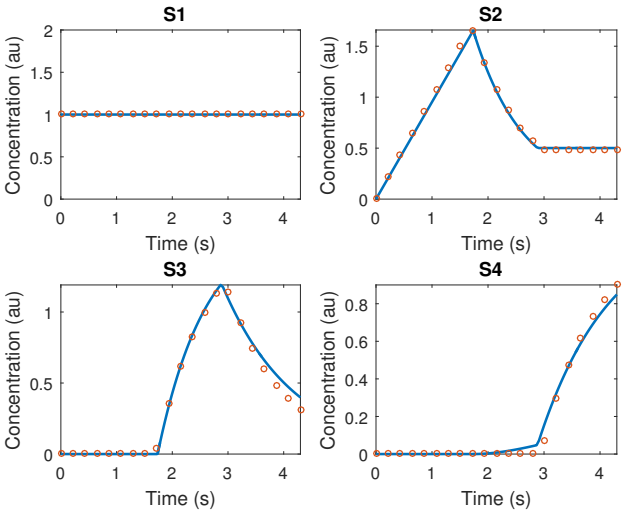

Fig. S38. Case LPN3B0p2l2: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S37.

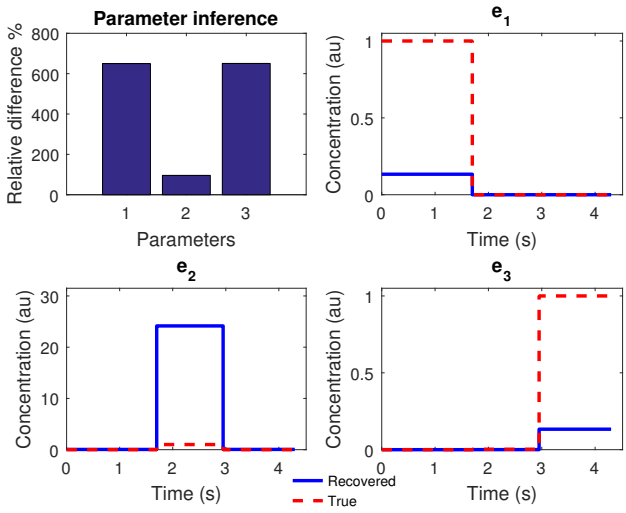

Fig. S39. Case LPN3B0norege2: two different optimal enzyme profiles which result in the same quality of fit. The computed enzyme profiles compensate for the massive differences in the parameter estimates. This result illustrates the lack of structural identifiability and corresponds to the solution of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

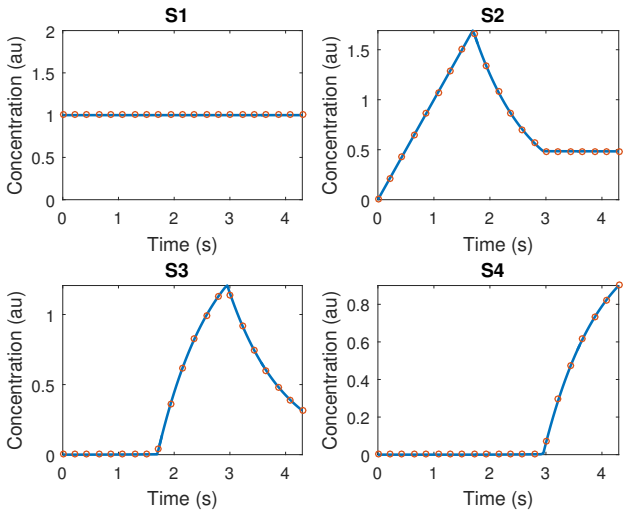

Fig. S40. Case LPN3B0norege2: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S39.

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.

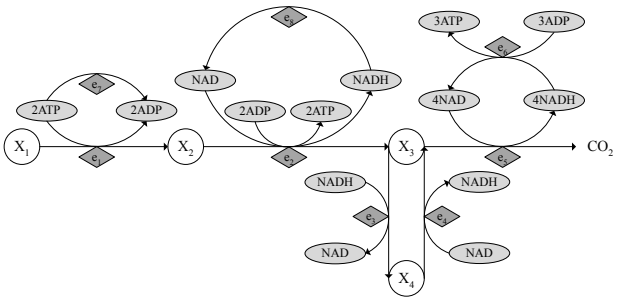

Fig. S41. Case SC: Network representation for the SC case study. Glucose and triose phosphates are represented in this model by X_1 and X_2, while pyruvate and ethanol, respectively, by X_3 and X_4.

1.10.1 Problem formulation

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

$$\min_{\mathbf{e}(t),\mathbf{k},t_f} \mathrm{J}_{\mathrm{outer}}[\mathbf{S},\mathbf{e},\mathbf{k}] \tag{S.66}$$

Where:

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

Subject to:

$$\max_{\mathbf{e}(t),t_f} \mathbf{J_{inner}}[\mathbf{S},\mathbf{e},\mathbf{k}] \tag{S.68}$$

Where:

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

Subject to:

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \mathbf{N}\cdot\mathbf{v}, \mathbf{S}\{t_0\} = \mathbf{S_0} \tag{S.70}$$

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

$$\sum_{i=1}^{6} e_i \leq E_T \tag{S.72}$$

$$NADH \geq NADH_c \tag{S.73}$$

$$ATP \geq ATP_c \tag{S.74}$$

Where the states' vector is:

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

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

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

$$\begin{aligned} v_1 &= k_1 \cdot e_1 \cdot X_1 \cdot ATP \\ v_2 &= k_2 \cdot e_2 \cdot X_2 \cdot NAD \cdot ADP \\ v_3 &= k_3 \cdot e_3 \cdot X_3 \cdot NADH \\ v_4 &= k_4 \cdot e_4 \cdot X_4 \cdot NAD \\ v_5 &= k_5 \cdot e_5 \cdot X_3 \cdot NAD \\ v_6 &= k_6 \cdot e_6 \cdot NADH \cdot ADP \\ v_7 &= k_7 \cdot ATP \\ v_8 &= k_8 \cdot NADH \end{aligned} \tag{S.77}$$

With the given initial values of:

$$\mathbf{S_0} = \begin{bmatrix} 1 & 1 & 1 & 10 & 0.7 & 0.8 & 0.3 & 0.2 \end{bmatrix} \tag{S.78}$$

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:

$$\min_{\tilde{\mathbf{e}}(t),\mathbf{k},t_f} \mathrm{J}_{\mathrm{outer}}[\mathbf{S},\tilde{\mathbf{e}},\mathbf{k}] \tag{S.79}$$

Where:

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

Subject to:

$$\max_{\tilde{\mathbf{e}}(t),t_f} \mathbf{J_{inner}}[\mathbf{S},\tilde{\mathbf{e}},\mathbf{k}] \tag{S.81}$$

Where:

$$\mathbf{J_{inner}}[\mathbf{S},\tilde{\mathbf{e}},\mathbf{k}] = \left[t_f, -\textstyle\int_{t_0}^{t_f} (\sum_{i=1}^{6} \tilde{e_i})dt\right]^T \tag{S.82}$$

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

$$\sum_{i=1}^{6} \tilde{e_i} \leq \tilde{E_T} \tag{S.83}$$

$$\begin{aligned} v_1 &= \tilde{e_1} \cdot X_1 \cdot ATP \\ v_2 &= \tilde{e_2} \cdot X_2 \cdot NAD \cdot ADP \\ v_3 &= \tilde{e_3} \cdot X_3 \cdot NADH \\ v_4 &= \tilde{e_4} \cdot X_4 \cdot NAD \\ v_5 &= \tilde{e_5} \cdot X_3 \cdot NAD \\ v_6 &= \tilde{e_6} \cdot NADH \cdot ADP \\ v_7 &= k_7 \cdot ATP \\ v_8 &= k_8 \cdot NADH \end{aligned} \tag{S.84}$$

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

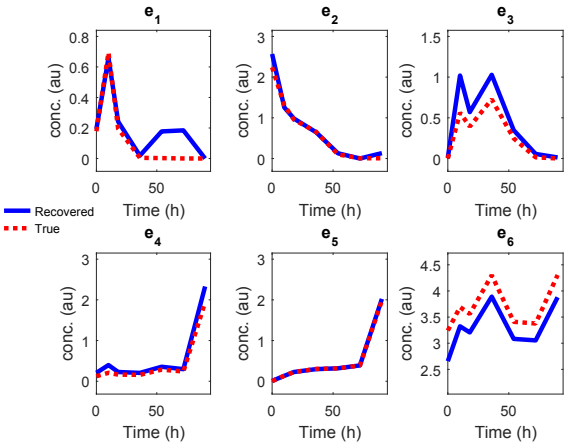

Fig. S42. Case SC0A: stimuli inference in the inverse optimal control problem. This solution corresponds to the noiseless case and scenario A, where we assume that we can measure the total enzyme concentration to be measured.

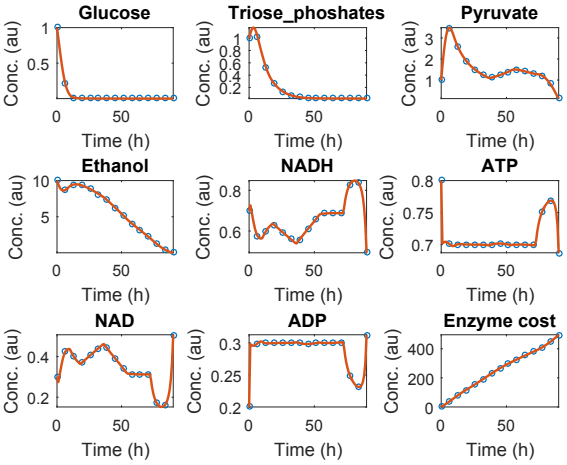

Fig. S43. Case SC0A: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S42.

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

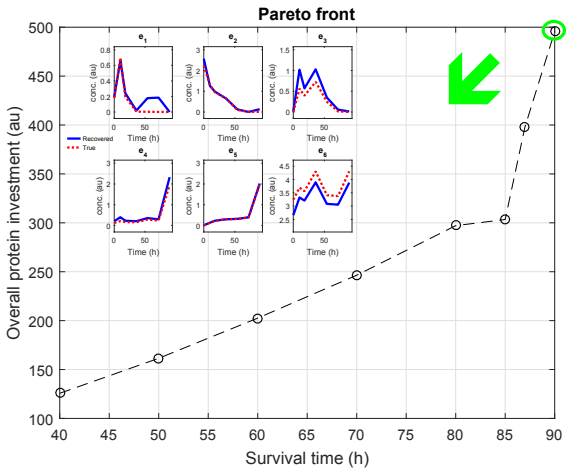

Fig. S44. Case SC0A: estimated inputs from the inverse optimal contro problem being matched to the corresponding point in the Pareto front. This corresponds to the noiseless case (scenario A) optimal solution presented in Fig. S42.

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-possedness 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

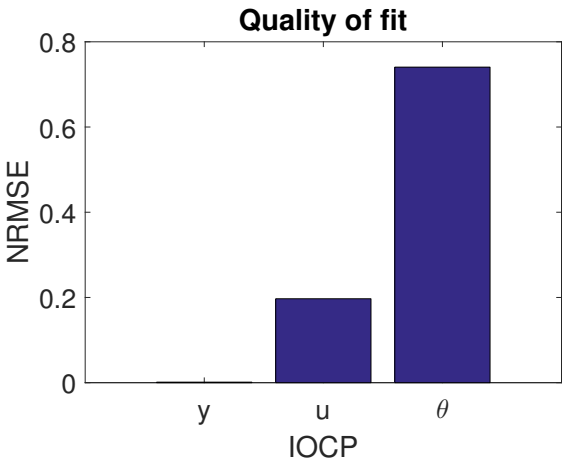

Fig. S45. Case SC0A: normalized root mean square error with respect to the observables (y), the inputs (u) and the true parameters (θ), corresponding to the noiseless case optimal solution presented in Fig. S42.

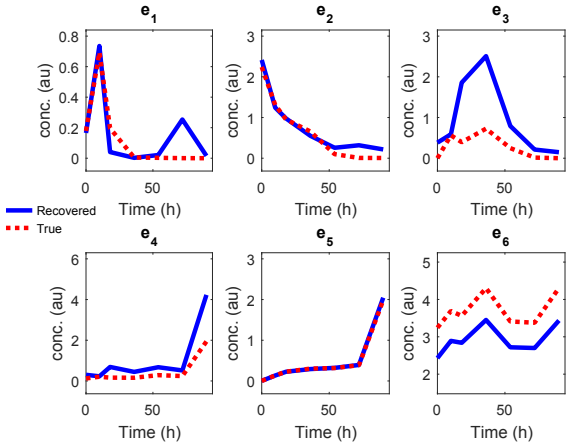

Fig. S46. Case SC0B: stimuli inference in the inverse optimal control problem. This solution corresponds to the noiseless case and scenario B, where we assume we can not measure the total enzyme concentration.

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 in 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 S54-S57 we present typical convergence

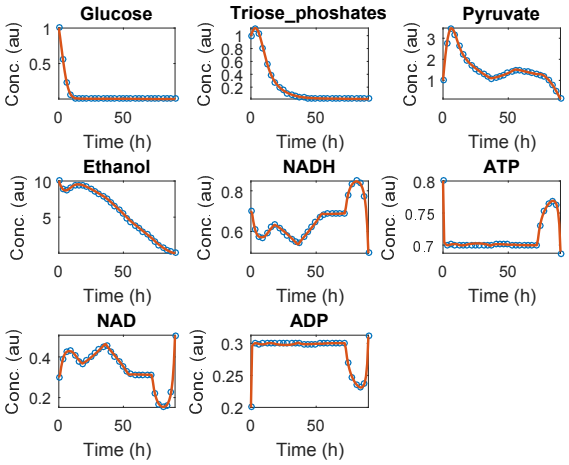

Fig. S47. Case SC0B: fit to the observables corresponding to the noiseless case optimal solution presented in Fig. S46.

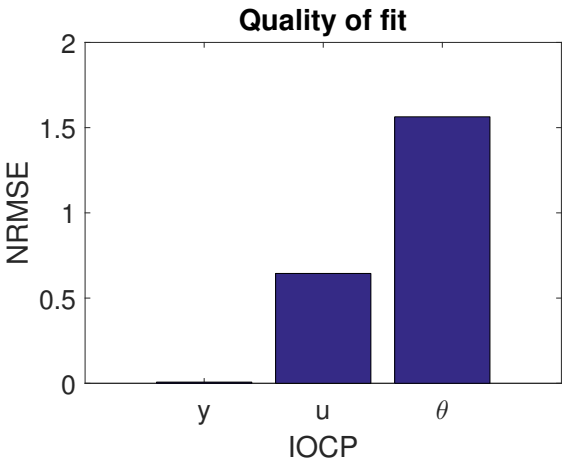

Fig. S48. Case SC0B: normalized root mean square error with respect to the observables (y), the inputs (u) and the true parameters (θ), corresponding to the noiseless case optimal solution presented in Fig. S46.

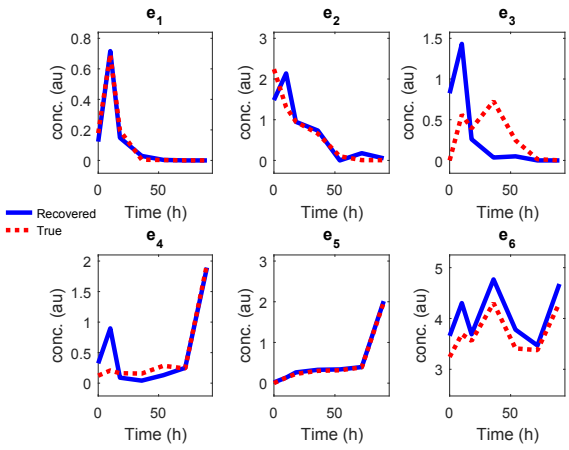

Fig. S49. Case SC10: stimuli inference in the inverse optimal control problem. This solution corresponds to the noisy case.

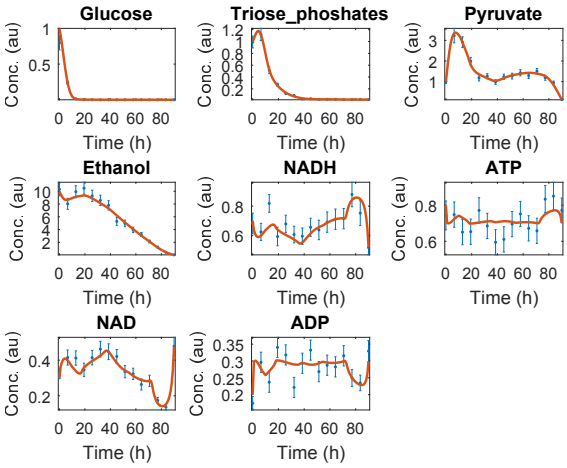

Fig. S50. Case SC10: fit to the observables corresponding to the noisy case optimal solution presented in Fig. S49.

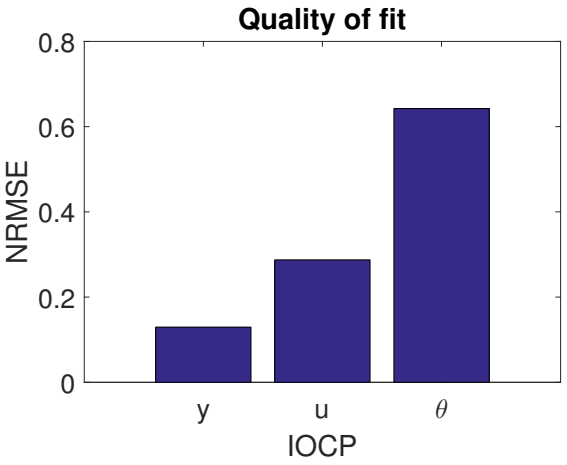

Fig. S51. Case SC10: normalized root mean square error with respect to the observables (y), the inputs(u) and the true parameters (θ) corresponding to the noisy case optimal solution presented in Fig. S49.

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.T6 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).

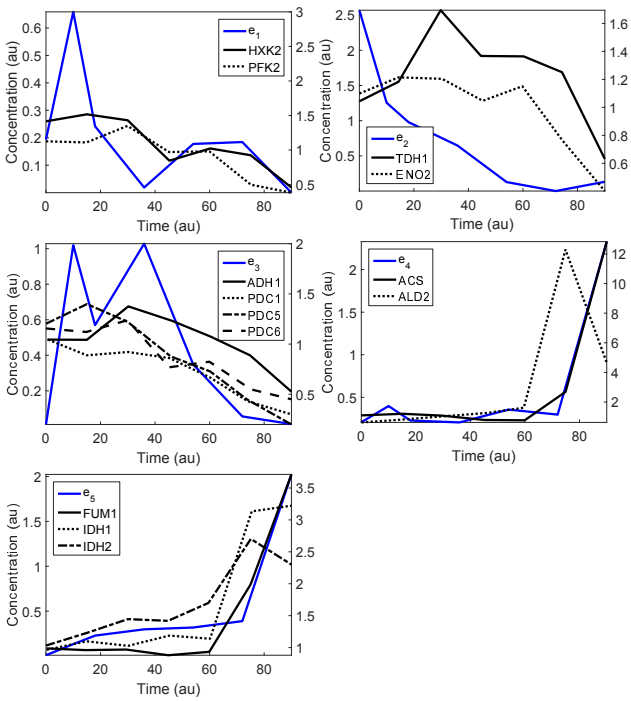

Fig. S52. Case SC0A: qualitative comparison of the estimated inputs with relevant experimental enzyme data. Our estimated inputs (blue line), presented in Fig. S42, correspond to the left axis while the experimental data (black lines) correspond to the right axis.

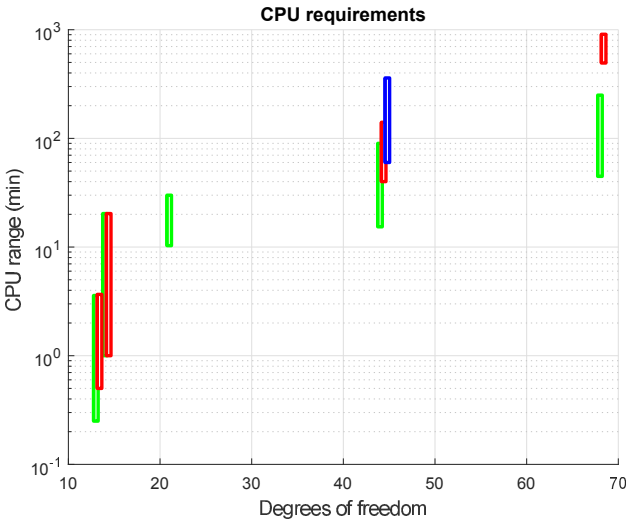

Fig. S53. Computation (CPU) time ranges for the case studies considered in this paper as a function of the degrees of freedom of the optimization problems. The different colors correspond to different subcases.

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, $CPUrange$: 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}	$CPUrange(min)$
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

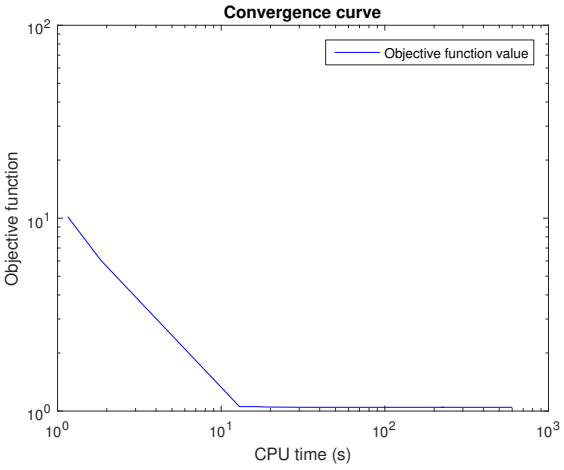

Fig. S54. Case JAKSTATreal: Convergence curve for the reported solution corresponding to the real experimental data.

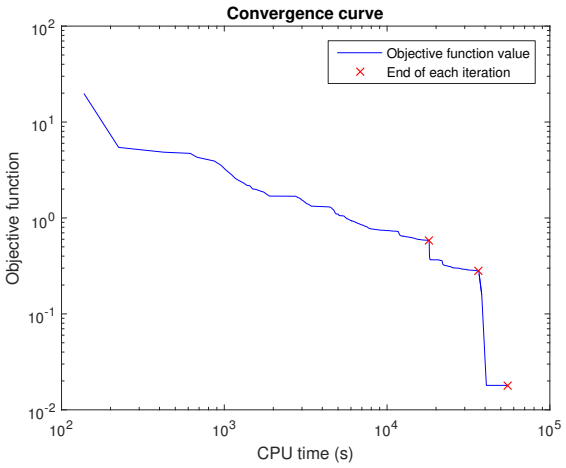

Fig. S55. Case TSP5: Convergence curve for the reported noisy solution.

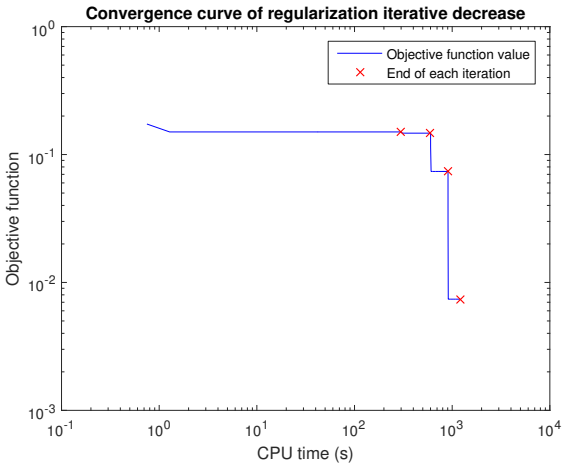

Fig. S56. Case LPN3B10: Convergence curve for the reported noisy solution.

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.

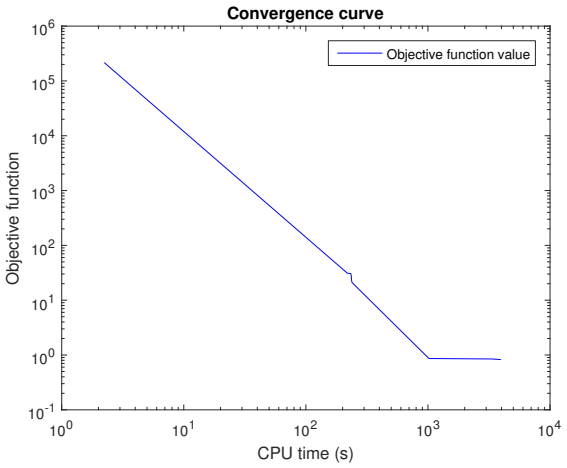

Fig. S57. Case SC10: Convergence curve for the reported noisy solution.

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