Supplementary 2: Analysis of different approaches

Let us consider the following E2F/DP dimmer model [4],

![Diagram of the double inhibitor–activator module](image)

The ODEs model is shown as following

\[
\frac{d[pRB]}{dt} = k_i \frac{[E2F1]}{K_{m1} + [E2F1] + [pRB]} - \phi_{pRB}[pRB]
\]

\[
\frac{d[E2F1]}{dt} = k_p + k_2 \frac{a^2 + [E2F1]^2}{K_{m1} + [E2F1]^2 + [pRB]} - \phi_{E2F1}[E2F1]
\]

(1)

The parameter values are shown in Table S1. Figure S2 is a dynamic profile of the system with noise free data.

<table>
<thead>
<tr>
<th>(k_i)</th>
<th>(K_{m1})</th>
<th>(J_{11})</th>
<th>(\phi_{pRB})</th>
<th>(k_p)</th>
<th>(k_2)</th>
<th>(a)</th>
<th>(K_{m2})</th>
<th>(J_{12})</th>
<th>(\phi_{E2F1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.005</td>
<td>0.05</td>
<td>1.6</td>
<td>0.04</td>
<td>4</td>
<td>5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

![Graph of the dynamic profile](image)

Fig. S2 The dynamic profile of a trial with observation data subject to 0% noise
S1 The direct optimization method or Single shooting method

The parameter estimation problem can be formulated as a nonlinear programming problem (NLP) with differential-algebraic constraints, which finds a set of optimal parameters $\hat{\theta} \in \mathbb{R}^n$, so that the responses from the ODEs model is in best alignment with the measurement data:

$$
\min_{\hat{\theta}} \sum_{j=0}^{N-1} (y(t_j) - \hat{y}(t_j|\hat{\theta}))^T W_j (y(t_j) - \hat{y}(t_j|\hat{\theta})) ,
$$

subject to:

(i) $\dot{x}(t) = f(\dot{x}(t), x(t), \theta)$, $x(t_0) = x_0$,

(ii) $\dot{y}(t) = g(\dot{x}(t), \theta)$, $j = 1, 2, \cdots, N-1$,

(iii) $C_{eq} (\hat{x}, \dot{x}, \hat{\theta}) = 0$,

(iv) $C_{ineq} (\hat{x}, \dot{x}, \hat{\theta}) \leq 0$,

(v) $\theta_L \leq \hat{\theta} \leq \theta_U$.

Here, $p_0$ minimizes a cost function that measures the fitness of the model with respect to a given set of experiment data subjecting to a set of constraints, where $\hat{\theta} \in \mathbb{R}^n$ is the set of parameters to be estimated, $\hat{x}_0$ is the estimated initial condition, $\hat{x} \in \mathbb{R}^n$ are the estimated system states ($\dot{x}(t|\hat{\theta})$ represents the estimated variable at time $t_j$ with parameter $\hat{\theta}$ and initial condition $\hat{x}_0$), $W_j$ is the weighting matrix, $\hat{y}$ is the estimated measured data. In some applications, additional constraints are introduced to impose special structural properties of a given system; it can be implemented in the form of the equality and inequality constraints $C_{eq}$ and $C_{ineq}$ (for instance the system performance and the mass balance constraints). Finally, $\theta_L$ and $\theta_U$ are simple structural constraints such as the parameter’s upper/lower bounds (it can be part of the $C_{ineq}$).

The basic idea of single shooting method is to estimate a set of parameters $\hat{\theta}$ and the initial condition $\hat{x}_0$. Then, putting $\hat{\theta}$ and $\hat{x}_0$ into the ODEs model and the estimated system states $\hat{x}$ can be computed. Also the error between the estimated system states $\hat{x}$ and the measurement data $y(t)$ can be expressed by (shown as Fig. S3)

$$
e(t_j) = y(t_j) - \hat{y}(t_j|\hat{\theta}) = y(t_j) - g(\hat{x}(t_j|\hat{\theta})) ,
$$

and we have the following cost function:

$$
C(\hat{\theta}) = \sum_{j=0}^{N-1} (y(t_j) - \hat{y}(t_j|\hat{\theta}))^T W_j (y(t_j) - \hat{y}(t_j|\hat{\theta}))
$$
In order to simplify the analysis for exposition purpose, we only vary parameters $p_{RB}$ and $\phi$ over the ranges $p_{RB} \in [0, 0.025]$ and $\phi \in [0, 0.5]$ respectively, and fixed all other parameters at their nominal values. The following displays the cost function surface of $P_0$.

Fig. S4-(a) shows that the cost function surface of $P_0$ is ridge shape, which drops suddenly from $10^6$ to 0 (the red dot, in Fig. S4-(b), is the minimal point). For the NLP-$P_0$, the disadvantages of the single shooting method are: (i) single shooting method requires solving the nonlinear dynamic model (1) for $\hat{x}$ in order to compute the cost function. The common method to estimate $\hat{x}(t_i)$ and $\hat{\dot{x}}(t_i)$ is using ODEs solvers, which perform the
numerical integration with $\hat{\theta}$ fixed at each iteration. During the process of identification, the integration has to be executed thousands even millions of times. Hence, the computation time spent on the $P_0$ can be hours even days. (ii) $P_0$ is a nonlinear optimization problem subject to a set of linear and non-linear differential equation constraints. Hence, $P_0$ is often multimodal (non-convex) and has many local minima. In a high-noise environment and combined with the non-convex property, the situation becomes more difficult. (iii) The estimated state may diverge away from the true trajectory, in some cases it may cause numerical overflow[1]. On the contrary, the single shooting method is simple in programming and there is no need to measure all states of the system.

S2. Decomposition method [2]

Single shooting method took a much longer time frame (horizon) to compute the estimated trajectory and hence requires more iteration for the estimation to converge to its true values. The computation time can be drastically reduced if one interprets the derivatives $\dot{x}$ at all measured time points $t_i$ as slopes, estimates these slopes from the observation data $y(k_i)$. Then, the system model becomes:

$$\hat{x}(t_j) = f(\hat{x}(t_j), u(t), \hat{\theta})$$

If the metabolic profile consists of metabolites $\{x_1, x_2, ..., x_n\}$, which are measured at $N$ instances $\{t_1, t_2, ..., t_N\}$. The inverse problem can thus be reformulated from one involving $n$ differential equations into a larger system of $n \times N$ algebraic equations as follows:

$$\dot{x}_1(t_j) = f(\hat{x}(t_j), u(t), \hat{\theta}_1)$$

$$\dot{x}_2(t_j) = f(\hat{x}(t_j), u(t), \hat{\theta}_2)$$

$$\vdots$$

$$\dot{x}_n(t_j) = f(\hat{x}(t_j), u(t), \hat{\theta}_n)$$

The error between the estimated system states $\dot{x}_i(t)$ and its derivative based on the ODEs model $f(\dot{x}(t), u(t), \hat{\theta})$ (shown as Fig. S5–S7) becomes

$$e_i(t_j) = \dot{x}_i(t_j) - f(\hat{x}(t_j), u(t), \hat{\theta})$$

Then, the parameter estimation problem can be separated into $n$ small subproblems. Hence, the optimization function becomes

$$P_{\text{decoup},i} : \min_{\hat{\theta}} \sum_{j=0}^{N-1} \| \dot{x}_i(t_j) - f(\hat{x}(t_j), \hat{\theta}) \|_2$$

s.t. (i) $\theta_L \leq \hat{\theta} \leq \theta_U$, 

$$i = 1, 2, \cdots, n_i$$

$$i = 1, 2, \cdots, n$$
Fig. S5 observed dynamical profile and estimated dynamic profile in noise free condition

Then the cost function surface of (8) is shown in Fig. S6

Compared with the cost function surface of $P_j$, the cost function surface of (8) is now bowl-shaped and local convex, which is “more smooth”. However, the reliability of the decomposed methods depends on the accuracy of the estimated states $\hat{x}(t_j)$ and their derivative $\dot{\hat{x}}(t_j)$. In practice, these data are subject to significant observation noise. Without proper regularization pre-processing, the overfitting problem may occur (shown in Fig. S7).
Fig. S7 The observed dynamical profile and estimated dynamic profile in 10% noise condition

Under this condition, the optimal solution can deviate from the “true” value [3]. One can see from Fig S8. The red point shown in Fig S8(b) deviates from the “true” value $\phi_{pRB} = 0.005$, $\phi_{E2F1} = 0.1$

Fig. S8 (Color online); (a) Cost function surface of the $P_{decoupl}$ as parameters $\phi_{pRB}$ and $\phi_{E2F1}$ are varied; (b) corresponding contours of the cost function (red dot is the minimal point).
S3. Combine spline theory and NLP (The proposed method $P_3$)

The parameter estimation problem is given as follows

$$P_3: \min_{\theta, p} \sum_{j=0}^{N-1} (y(t_j) - \hat{y}(t_j))^T w_j (y(t_j) - \hat{y}(t_j))$$

subject to

(i) $\hat{x}(t_j) = h(t_j) \cdot p$, 
(ii) $\hat{x}(t_j) = h(t_j) \cdot p$,
(iii) $\|\hat{x}(t_j) - f(\hat{x}(t_j), u(t), \hat{\theta})\|_2^2 = 0$,
(iv) $\hat{y}(t_j) = g(\hat{x}(t_j))$,
(v) $\theta_L \leq \hat{\theta} \leq \theta_U$,

$i = 1, 2, \ldots, n, \ j = 0, 2, \ldots, N-1$. \hspace{1cm} (9)

Now, the NLP-$P_3$ consists of algebraic equation constraints only. Hence, $P_3$ does not require ODEs solvers, which drastically ease the computation burden.

Then, we can turn $P_3$ into

$$P_4: \min_{\theta, p} \sum_{j=0}^{N-1} (y(t_j) - g(B(t_j) \cdot p))^T w_j (y(t_j) - g(B(t_j) \cdot p))$$

subject to

(i) $\|\hat{B} \cdot p - f(B \cdot p, B_t \cdot p, \hat{\theta})\|_2^2 = 0$,
(ii) $\theta_L \leq \hat{\theta} \leq \theta_U$, \hspace{1cm} (10)

where $p = (p_1, p_2, \ldots, p_L)^T$.

The Lagrangian function of $P_4$ is

$$L(\hat{\theta}, p, \lambda) = \sum_{j=0}^{N-1} (y(t_j) - B(t_j) \cdot p)^T w_j (y(t_j) - B(t_j) \cdot p) + \lambda \|\hat{B}(t_j) \cdot p - f(B(t_j) \cdot p, B_t(t_j) \cdot p, \hat{\theta})\|_2^2$$ \hspace{1cm} (11)

where $\lambda > 0$ is the Lagrange multiplier associated with the constraint (i) of (10). In conclusion, the parameter estimation problem becomes a minimization problem as follows:

$$P_3: \min_{\theta, p} \sum_{j=0}^{N-1} (y(t_j) - B \cdot p)^T w_j (y(t_j) - B \cdot p) + \lambda \|\hat{B} \cdot p - f(B \cdot p, \hat{\theta})\|_2^2$$ \hspace{1cm} (12)

over the search range $\theta_L \leq \hat{\theta} \leq \theta_U$. Compared with the decomposed methods which divide the estimation of the system states (and its derivative) and the parameter estimation into two separate steps; while $P_4$ computes the estimated states (and its derivative) and parameters at the same time. Note that constraint (iii) of $P_3$ governs the estimation of system’s states (and their derivative) so as to make sure these estimates belong to one solution of system (1). Thus, the dynamic system model itself servers as a filter performing regularization and the overfitting problem can be relieved. Hence, even in noisy condition, the position of minimal point do not derivate as badly as the decomposition method (shown as Fig S8 and S9).
Fig. S8 observed dynamical profile and estimated dynamic profile in 10% noise condition

Fig. S9 (Color online); (a) Cost function surface of the $P_{\text{decoup}}$ as parameters $\phi_{pRB}$ and $\phi_{E2F1}$ are varied; (b) corresponding contours of the cost function (red dot is the minimal point).

Hence, the augmented cost function surface of $P_i$ is smooth. This characteristic enables the optimization algorithm to arrive at the best solution more easily.
Reference


