Global optimization in systems biology: stochastic methods and their applications

E. Balsa-Canto*, J.R. Banga, J.A. Egea, A. Fernandez-Villaverde and G.M. de Hijas-Liste

Abstract Mathematical optimization is at the core of many problems in systems biology: i) as the underlying hypothesis for model development, ii) in model identification or iii) in the computation of optimal stimulation procedures to synthetically achieve a desired biological behaviour. These problems are usually formulated as non-linear programming problems with dynamic and algebraic constraints. However the nonlinear and highly constrained nature of systems biology models, together with the usually large number of decision variables, can make their solution a daunting task, therefore calling for efficient and robust optimization techniques.

Here, we present novel global optimization methods and software tools such as cooperative eSS, AMIGO or DOTcvpSB, and illustrate their possibilities in the context of modelling including model identification and stimulation design in systems biology.

1 Introduction

The use of optimization has allowed biologists not only to describe patterns or mechanisms but to predict, from first principles, how organisms should be designed [41, 6]. In particular, mathematical optimization i) is the underlying hypothesis for model development in for example flux balance analysis [21] or the activation of metabolic pathways [22, 47, 29], ii) is at the core of model identification, including parameter estimation and optimal experimental design [7] or iii) enables the computation of optimal stimulation procedures to synthetically achieve a desired biological behaviour [25, 35].

E. Balsa-Canto, A. Fernandez-Villaverde, G.M. de Hijas-Liste, J.R. Banga
Bioprocess Eng. Group, IIM-CSIC, Vigo-Spain e-mail: ebalsa@iim.csic.es

J.A. Egea
Dept. Appl. Math. & Stat., Tech. Univ. Cartagena UPCT, Cartagena-Spain

Most of these problems are formulated as non-linear programming problems (NLPs) where the objective is to find a set of decision variables (or functions) in order to minimize or maximize a given cost function (or functional) subject to a set of dynamic and algebraic constraints. The solution of such problems requires the use of advanced numerical optimization methods. In this regard, hundreds of different methods are at hand: from deterministic local methods to sophisticated metaheuristics. One aspect that should be taken into account at the time of selecting the most appropriate method is the nature of the problem under consideration.

Whereas convex problems present a unique solution which may be found with deterministic local methods, finding the global optimum for multimodal problems, i.e. those presenting multiple local optima, including noisy problems, typical in dynamic systems due to the numerical integration of complex partial and ordinary differential equations, requires robust and efficient global optimization methods.

Some of these methods have been incorporated in software tools devoted to modelling, model analysis, simulation and parameter estimation such as: COPASI [18], SBToolbox2 [37] or PottersWheel [26].

In this work we present novel global optimization methods and software tools developed at our group which are devoted to handle not only parameter estimation but different optimization problems in the context of systems biology. In this regard we will introduce:

- DOTcvpSB [17], which is the first toolbox for dynamic optimization problem in Systems Biology, i.e. offering the possibility of handling dynamic FBA problems, optimal enzymatic activation problems or the optimal design of stimulation profiles to achieve certain desired biological behaviours.
- AMIGO, that covers all the steps of the iterative identification procedure [2]: local and global sensitivity analysis, parameter estimation, identifiability analysis and optimal experimental design. The robust identifiability analysis, parameter estimation and optimal experimental design problems are formulated and solved as general (dynamic) optimization problems.
- A multi-thread cooperative scatter search approach based on eSS [14] is presented here as a means to handle large scale multimodal problems. We remark that the cooperative eSS could be incorporated as an optimizer in both DOTcvpSB and AMIGO.

Four illustrative examples have been selected to show their applicability.

DOTcvpSB [17] is used to solve a problem related to the enzyme activation in a branched reaction network. The advantages of the cooperative scatter search approach are illustrated through the solution of a parameter estimation problem related to the modelling of the central carbon metabolism in *E. coli*. AMIGO is used to solve an optimal experimental design problem related to a three-step metabolic pathway model. And the last example illustrates how hybrid optimization methods incorporated in DOTcvpSB are able to solve a highly multimodal problem related to the computation of the optimal stimulation conditions to obtain a given multicellular structure in bacterial chemotaxis.

2 Optimization Problem formulation

Consider a general dynamic and possibly distributed system described by the following state space equations:

$$\dot{\mathbf{y}} = \Xi(\mathbf{x}, \mathbf{y}, \mathbf{u}, \theta, t); \qquad \mathbf{x}_t = \Psi(\mathbf{x}, \mathbf{x}_\xi, \mathbf{x}_{\xi\xi}, \mathbf{y}, \mathbf{u}, \theta, t) \tag{1}$$

where $\xi \in \Omega \subset \Re^3$ are the spatial variables, $\mathbf{x}(\xi, t) \in X \subset \Re^\nu$ is the subset of state variables depending on both time and spatial location, $\mathbf{y}(t) \in Y \subset \Re^\mu$ is the subset of time dependent variables, $\mathbf{x}_\xi = \partial\mathbf{x}/\partial\xi$, $\mathbf{x}_{\xi\xi} = \partial^2\mathbf{x}/\partial\xi^2$, $\mathbf{x}_t = \partial\mathbf{x}/\partial t$, $\dot{\mathbf{y}} = d\mathbf{y}/dt$, $\mathbf{u} \in U \subset \Re^\sigma$ are the control variables and $\theta \in \Theta \subset \Re^\eta$ are time independent parameters.

In addition, state variables are subject to initial and boundary conditions:

$$\mathbf{y}(t_0) = \Xi_0(\mathbf{x}(t_0), \mathbf{u}(t_0), \theta, t_0) \tag{2}$$

$$\mathbf{x}(t_0) = \Psi_0(\mathbf{y}(t_0), \mathbf{u}(t_0), \theta, t_0); \quad \mathscr{B}(\mathbf{x}, \mathbf{x}_\xi, \mathbf{u}, \theta, \xi, t) = 0; \quad \xi \in \Omega \tag{3}$$

Note that the formulation in Eqns. (1-3) can be used to model many biological systems such as biochemical pathways, e.g., cell signalling or metabolic pathways; diffusion reaction systems, e.g. pattern formation or persistence and extinction of species; etc.

State and control variables may be also subject to algebraic constraints which force the satisfaction of particular biological conditions at particular time points or throughout the process:

$$\mathbf{r}_k^{eq}(\mathbf{x}(\xi, t_k), \mathbf{y}(t_k), \mathbf{u}(t_k), \theta, t_k) = 0; \quad \mathbf{r}_k^{in}(\mathbf{x}(\xi, t_k), \mathbf{y}(t_k), \mathbf{u}(t_k), \theta, t_k) \leq 0 \tag{4}$$

$$\mathbf{c}^{eq}(\mathbf{x}(\xi, t), \mathbf{y}(t), \mathbf{u}(t), \theta, t) = 0; \quad \mathbf{c}^{in}(\mathbf{x}(\xi, t), \mathbf{y}(t), \mathbf{u}(t), \theta, t) \leq 0 \tag{5}$$

Control variables and parameters may be also subject to bound constraints:

$$\mathbf{u}^L \leq \mathbf{u}(t) \leq \mathbf{u}^U; \quad \theta^L \leq \theta \leq \theta^U \tag{6}$$

The last element in the problem definition will be the objective functional that quantifies the quality of a solution:

$$J = \phi(\mathbf{x}(\xi, t_f), \mathbf{y}(t_f), \theta, t_f) + \int_{t_0}^{t_f} L(\mathbf{x}(\xi, t), \mathbf{y}(t), \mathbf{u}(t), \theta, \xi, t) dt \tag{7}$$

where the scalar functions ϕ (Mayer term) and L (Lagrangian term) are continuously differentiable with respect to all of their arguments, and the final time t_f can be either fixed or free.

This objective functional may be related to, for example, the quantity of metabolites produced in a metabolic pathway, to the distance among experimental data and model predictions in the case of parameter estimation or to the information provided by an experimental scheme in the case of optimal experimental design.

The general **dynamic optimization** (DO) problem considered here can be then formulated as: Find the controls $\mathbf{u}(t)$ and the time-invariant parameters θ subject to the system dynamics in Eqns. (1-3) and the algebraic constraints in Eqns. (4-6) so as to minimize (or maximize) the objective functional in Eqn. (7).

3 Numerical methods

There are several alternatives for the solution of DO problems from which the indirect and the direct methods are the most widely used. The *indirect methods* make use of the Pontryagin's maximum principle so as to obtain the optimality necessary conditions. The method relies on the formulation of the Hamiltonian by summing the cost functional, the product of multiplier functions (co-states) with the dynamic equations in (1) and the product of the Lagrange multipliers with algebraic constraints and the subsequent derivation of the corresponding first and second order derivatives on the decision variables. The result will be a two or multi-point boundary value problem which must be solved for the state and co-state variables [11]. However, the complexity of the numerical solution of such boundary value problems has motivated the use of direct methods for most realistic applications.

Direct methods such as the complete parameterization (CP, [9]), multiple shooting (MS, [10]) or control vector parameterization (CVP, [44]) transform the DO problem into a non-linear programming (NLP) problem. These methods discretize and approximate either the control variables or both the control and state variables in such a way that the decision variables for the NLP are related to the given parameterization scheme. The three alternatives basically differ in the resulting number of decision variables, in the presence or absence of parameterization related constraints and in the necessity of using a boundary value problem solver. While the CP or the MS approaches may become prohibitively expensive in computational terms, the CVP approach allows handling large scale dynamic optimization problems without solving very large NLPs and without dealing with extra junction constraints.

3.1 Control vector parameterization

The CVP method proceeds dividing the duration of the process into a number ρ of control intervals and the control function is approximated using a low order polynomial form over each interval. Each control variable approximation may be expressed using Lagrange polynomials as follows:

$$u_j(t) = \sum_{i=1}^{M_j} u_{ij} \Phi_i^{(M_j)}(\tau) \tag{8}$$

where, $j = 1,...,\rho$, $t \in [t_0, t_f]$, and τ is normalized time given by,

$$\tau = \frac{t - t_0}{t_f - t_0} \tag{9}$$

and the Lagrange polynomials of order M, $\Phi_i^{(M_j)}$ are defined in the standard form,

if M=1

$$\Phi_i^{(M)}(\tau) \equiv 1 \tag{10}$$

if $M \geq 2$

$$\Phi_i^{(M)}(\tau) \equiv \prod_{i'=1, i \neq 1}^{M} \frac{\tau - \tau_{i'}}{\tau_i - \tau_{i'}} \tag{11}$$

The parameters of these polynomials, u_{ij}, will be used as decision variables in the optimization process together with time independent parameters.

The generalization of the CVP approach for the case of optimal experimental design may be found in [1].

3.2 Boundary value problem solution

The solution of the nonlinear dynamic, sometimes distributed, models describing biological systems (Eqn. 1) requires the use of suitable numerical techniques. For the most general case involving partial differential equations (PDEs) numerical methods use some type of space parameterization approach to transform the PDEs into an equivalent set of ordinary differential equations (ODEs) [36]. The numerical method of lines and the finite element method are the most widely used approaches for this transformation. The underlying idea is to discretize the domain of interest into many smaller subdomains and use local spatial functions to approximate the distributed variables in each subdomain. As a result a large-scale, usually stiff, set of ODEs is obtained which may be solved with a sparse implicit initial value problem solver.

3.3 Nonlinear programming methods

Nonlinear programming methods may be largely classified in two main groups: local and global. Local methods are designed to generate a sequence of solutions, using some type of pattern search or gradient and Hessian information, that will converge to a local optimum, usually the closest to the provided initial guess. However the NLPs with non-linear dynamic constraints (such as in parameter estimation or the ones resulting from the application of the CVP approach) are frequently multimodal (i.e. presenting multiple local optima) [6, 7]. Therefore, local methods may converge

to local solutions, especially if they are started far away from the global optimum. In order to surmount these difficulties, global methods must be used.

3.3.1 Global optimization methods

Global methods have emerged as the alternative to search the global optimum [30]. The successful methodologies combine effective mechanisms of exploration of the search space and exploitation of the previous knowledge obtained by the search. Depending on how the search is performed and the information is exploited the alternatives may be classified in three major groups: deterministic, stochastic and hybrid.

Global deterministic methods [31, 16] in general take advantage of the problem's structure and guarantee global convergence for some particular problems that verify specific smoothness and differentiability conditions. Although they are very promising and powerful, there are still limitations to their application, particularly for non-linear dynamic systems, since the computational cost increases rapidly with the size of the considered dynamic system and the number of decision variables.

Global stochastic methods do not require any assumptions about the problem's structure. They make use of pseudo-random sequences to determine search directions toward the global optimum. This leads to an increasing probability of finding the global optimum during the run time of the algorithm, although convergence may not be guaranteed. The main advantage of these methods is that, in practice, they rapidly arrive to the proximity of the solution.

The most successful approaches lie in one (or more) of the following groups: pure random search and adaptive sequential methods, clustering methods or metaheuristics. Metaheuristics are a special class of stochastic methods which have proved to be very efficient in recent years. They include both population (e.g., genetic algorithms) or trajectory-based (e.g., simulated annealing) methods. They can be defined as guided heuristics and many of them try to imitate the behaviour of natural or social processes that seek for any kind of optimality [42]. Some of these strategies have been successfully applied to, for example, parameter estimation [27, 28, 40] or optimal experimental design [1] in the context of systems biology.

Despite the fact that many stochastic methods can locate the vicinity of global solutions very rapidly, the computational cost associated to the refinement of the solution is usually very large. In order to surmount this difficulty, hybrid methods and metaheuristics have been recently presented for the solution of dynamic optimization problems [4, 13] or parameter estimation problems [34, 33, 3]. They speed up these methodologies while retaining their robustness and, provided a gradient based local method is used, they guarantee convergence to a gradient zero solution.

In particular, the Scatter Search metaheuristic [15] is an evolutionary hybrid optimization method that has been successfully applied to the solution of parameter estimation problems [34, 46, 32] but also dynamic optimization [13] and optimal experimental design [2] problems. The newest version, the enhanced scatter search method (eSS, `www.iim.csic.es/~gingproc/ssmGO.html`, [14]), presents

a simpler but more effective design which helps to overcome typical difficulties of nonlinear dynamic systems optimization such as noise, flat areas, non-smoothness, and/or discontinuities.

4 Illustrative examples

4.1 Optimal enzyme activation in metabolic networks

An example of the insight that optimization can provide concerns the enzyme activation in metabolic networks. Several authors have shown that the genetic regulation of metabolic networks may follow an optimality principle such as the minimization of the transition time or the maximization of the production of a given metabolite. For example, the optimal "just-in-time" activation pattern in enzyme expression for the case of unbranched pathways has been formulated and solved as a non-linear optimization problem with dynamic constraints [22, 47]. More recently the problem has been considered as a general dynamic optimization one and was solved through the use of the Pontryagin's maximum principle for linear pathways [29, 8]. However the difficulty (or impossibility) of analytically solving other more realistic cases, such as those considering non-linear dynamics for the enzyme expression, or other arbitrarily complex networks, calls for the use of robust numerical DO approaches.

Here we consider one of such examples and approach its solution using the DOTcvpSB toolbox (`http://www.iim.csic.es/~dotcvpsb`, [17]) which combines the CVP approach with global stochastic and hybrid methods to solve dynamic optimization problems.

The pathway considered is depicted in Fig. 1. It consists of four enzymatic reactions with one branch where the products are accumulated to be consumed later.

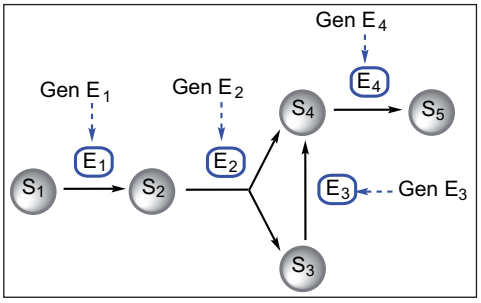

Fig. 1 Schematic representation of the branched pathway considered. The pathway consists of four enzymatic reactions catalyzed by a specific enzyme (E_i) where S_1 is the substrate, $S_2 - S_4$ are intermediates and S_5 is the product. The enzyme dynamics are considered to be linear with a reaction rate r_i.

The hypothesis is that the pathway activation minimizes the time from the substrate to the product. The activation profile may then be found by computing $\mathbf{r}_i(t)$ over $t \in [t_0, t_f]$ to minimize $J = t_f$ subject to the system dynamics:

$$\frac{d\mathbf{S}_i}{dt} = N\upsilon \qquad \frac{d\mathbf{E}_i}{dt} = r_i - \lambda \mathbf{E}_i \tag{12}$$

where:

$$\upsilon = \frac{k_{cat}\mathbf{S}_i}{K_M + \mathbf{S}_i} \qquad N = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

and the following end point and path constraints:

$$S_5\left(t_f\right) = P_{tf} \qquad \sum_{i=1}^{4} \mathbf{E}_i \leq E_T \tag{14}$$

with $K_M = 1mM$, $k_{cat} = 1s^{-1}$, $P_{tf} = 0.75mM$, $\lambda = 0.5$.

The optimal activation profile corresponding to an optimal final time $t_f = 10.4s$ obtained with an evolutionary approach is presented in Fig. 2.

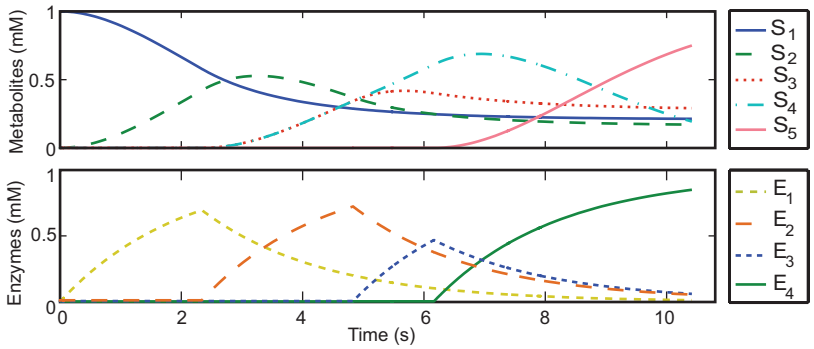

Fig. 2 Optimal enzyme activation profile and the corresponding metabolite dynamics. The optimal profiles for the expression rates follow a switching pattern that matches with the pathway topology leading to enzyme profiles that follow a sequential activation with protein degradation to synthesize another protein. The substrate (S_1) is converted into the product (S_5) through the intermediates (S_2, S_3, S_4), the intermediate (S_4) is accumulated and consumed in the last section of the pathway.

4.2 Parameter estimation in complex systems biology models

The problem of parameter estimation in biochemical pathways, formulated as a non-linear programming problem where the objective is to compute the model parameter values that maximize the fit to the experimental data, has received substantial attention [19, 28, 33]. Many difficulties found during parameter estimation are not only due to the highly non-linear nature of the models and their size, but also due to the quality and quantity of experimental data. These result in poor practical identifiabil-

ity, i.e. in the difficulty or impossibility to compute unique values for the parameters given a set of data, or the presence of suboptimal solutions.

The presence of suboptimal solutions may be tackled with global optimization methods. Recent works show how hybrid stochastic-deterministic methods handle small-to-medium size problems with reasonable computational efforts [34, 33, 3]. However, further developments are necessary to enhance, in so far is possible, the efficiency of the optimization while keeping robustness for large scale complex biological models. Multi-thread approaches, i.e. those running several computations in parallel in different processors, seem to be the most suitable for this purpose.

Here we present a new *cooperative* strategy for the parallelization of the enhanced Scatter Search (eSS) algorithm [14]. The central idea is to run, in parallel, several *threads* of eSS, which may have different settings and/or random initializations, and exchange information among them as shown in Fig. 3. Taking into account the classification of cooperation schemes proposed in [43], the cooperative eSS can be described as follows:

1. There are η concurrent programs.
2. The best solution found and the eSS reference set, which contains valuable information about the diversity of solutions, are available for sharing.
3. All threads share the information.
4. The threads exchange information at a fixed time interval τ.

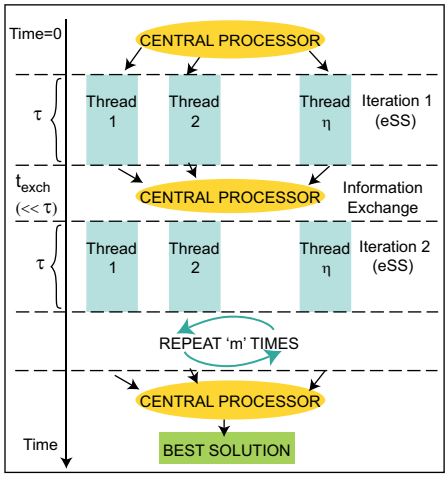

Fig. 3 Schematic representation of the cooperative eSS.

Each of the η threads has a fixed degree of "aggressiveness". "Conservative" threads are used for increasing the probabilities of finding a feasible solution, even if the parameter space is "rugged" or weakly structured. "Aggressive" threads may speed up the calculations in "smoother" areas. Communication, which takes place at fixed time intervals, enables each thread to benefit from the knowledge gathered by the others. This knowledge includes not only information about the best solution found so far, but also about the sets of diverse parameter vectors that may be worth trying for improving the solution.

It should be noted that cooperation produces more than just speed-up since it can change the systemic properties of an algorithm and therefore its macroscopic behavior [43]. To illustrate this point an example related to the parameter estimation of a model describing the central carbon metabolism of *E. coli* that takes into account the enzymatic and transcriptional regulation layer [23] is considered.

The model consists of 47 non-linear ODEs with 193 unknown parameters (affinity constants, specific activities, Hill coefficients, growth rates, expression rates, etc). The objective is to compute those parameters so as to predict a given system

behaviour. Due to the stiff character of the equations and the time required for their solution, one evaluation of the least squares function takes a few seconds.

The cooperative and non cooperative multi-thread implementations of eSS are compared by launching ten threads in both cases. In the cooperative case, the ten threads exchange information as explained. In the non-cooperative case, they simply run until the maximum computation time is reached. Fig. 4 presents the corresponding convergence curves showing how the cooperative version outperforms the non-cooperative one, being capable of finding a better value of the objective function while reducing computation time by 70%.

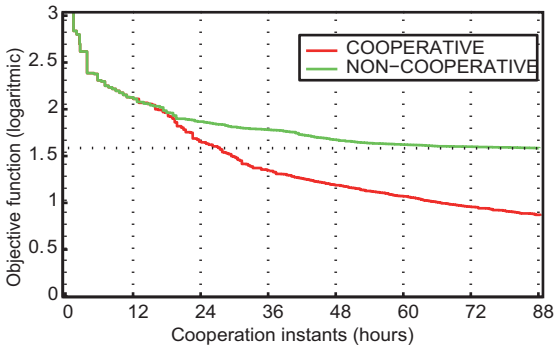

Fig. 4 Comparison of the performance of the parallel and cooperative eSS implementations in the solution of a large scale parameter estimation problem. Each curve represents, at every time instant, the best value found by any of its 10 threads.

4.3 Optimal experimental design for parameter estimation

As mentioned above, poor practical identifiability has to do with the type of experimental scheme being used and the quality of the corresponding experimental data in terms of experimental noise. The purpose of optimal experimental design is to devise the necessary dynamic experiments in such a way that the parameters are estimated from the resulting experimental data with the best possible statistical quality, which is usually a measure of the accuracy and/or decorrelation of the estimated parameters. In this way, the model and a close-to-optimal solution for the parameters are being used to design new more informative experiments which in general will result in better practical identifiability properties. The information provided by the measurements is often quantified by means of the Fisher information matrix [1, 5].

AMIGO (`http://www.iim.csic.es/~amigo`, [2]) is a multi-platform toolbox which apart from covering model simulation, local and global sensitivity analysis, parameter estimation and identifiability analysis, incorporates the optimal experimental design as a general dynamic optimization problem.

Here we illustrate its possibilities in the context of optimal experimental design with an example related to a model describing a pathway consisting of three enzy-

matic steps including the enzymes and the mRNAs explicitly [28]. Previous works [28, 34, 33] considered a factorial plan consisting of 16 experiments under different amounts of substrate and product to estimate all 36 model parameters. We will consider here the case of estimating: na_2, na_3, k_1, k_2, k_3, k_4, k_6, V_1, V_2, V_3, V_5, K_5.

In a few seconds with eSS as implemented in AMIGO the global optimum is achieved corresponding to the following parameter values: $k_1 = 1.0\pm4.4$, $k_2 = 0.1\pm6.9$, $k_3 = 1.0\pm8.8$, $k_4 = 0.1\pm0.01$, $k_6 = 0.1\pm0.02$, $V_1 = 1.0\pm4.4$, $V_2 = 0.1\pm6.9$, $V_3 = 1.0\pm8.8$, $V_5 = 0.1\pm0.07$, $na_2 = 2.0\pm0.7$, $na_3 = 2.0\pm0.7$, $K_5 = 1.0\pm1.2$. Note that even though the global solution was found, the confidence regions for some of the parameters are rather large and in many cases (k_1, k_2, k_3, V_1, V_2, V_3, K_5) they are over the 100%.

In order to improve the practical identifiability we implemented in AMIGO the design of a parallel-sequential experimental scheme. In particular two experiments were designed under the following conditions:

- Experiment 17: pulsed stimulation of the substrate. The location and duration of the pulses as well as the number and location of sampling times and experiment duration were to be optimized.
- Experiment 18: a 5 step-wise stimulation of the substrate is allowed within the maximum and minimum values.

Even allowing for limited flexibility in the design of the experiments, results (Fig. 5) reveal a substantial reduction in the confidence regions for the parameters: $k_1 = 1.0\pm1.3(-70\%)$, $k_2 = 0.1\pm3.2(-54\%)$, $k_3 = 1.0\pm4.1(-53\%)$, $V_1 = 1.0\pm1.3(-70\%)$, $V_2 = 0.1\pm3.2(-54\%)$, $V_3 = 1.0\pm4.1(-53\%)$, $K_5 = 1.0\pm0.2(-83\%)$. Further improvements may be achieved by either allowing for further flexibility in the designs or by adding new experiments.

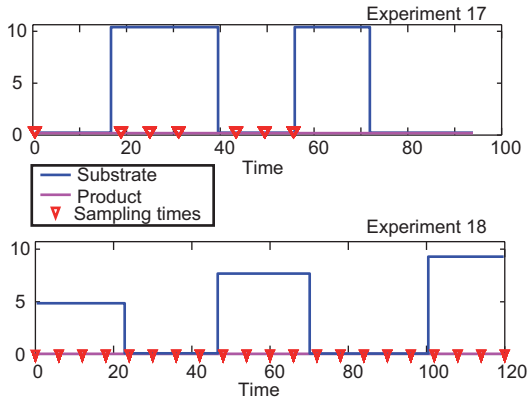

Fig. 5 Optimally designed experiments. For the experiment 17 the number and location of sampling times was optimally designed, resulting in a much reduced number of necessary sampling times. It is also remarkable that optimal step-wise experiment 18 results in a pulse-up type stimulation.

4.4 Stimulation design

For some particular systems, once reliable models have been developed it is possible to design stimulation conditions so as to achieve a certain goal. In this context, it is possible, for example, to optimally design medical treatments or immune responses [12, 20], or to obtain particular behaviours such as in the case of pattern formation [25, 24, 35]. However the solution of such problems often results in the presence of suboptimal solutions [25]. Here, we illustrate with an example related to pattern formation in bacterial chemotaxis [24] how hybrid global optimization methods may successfully solve these problems.

Some types of cells are able to sense the presence of chemical signals (chemoattractants) and guide their movement in the direction of the concentration gradient of these signals. This process is called chemotaxis. The chemotaxis of the bacteria *E. coli* is one of the best understood chemotaxis processes. These bacteria, under given stress conditions, secrete chemoattractants. Other cells respond to these secreted signalling molecules by moving up their local concentration gradients and forming different types of multicellular structures.

The system may be described by a two-component diffusion reaction model:

$$\frac{\partial z}{\partial t} = D\frac{\partial^2 z}{\partial \xi} + \mu \frac{\partial}{\partial \xi}\left(\frac{z}{(1+c)^2}\frac{\partial c}{\partial \xi}\right) \qquad \frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial \xi} + \frac{z^2}{(1+z^2)} \tag{15}$$

with homogenous first order boundary conditions and initial conditions $z(\xi,0) = 1$; $c(\xi,0) = 0$. Where $z(\xi,t)$ and $c(\xi,t)$ represent the cell density and the concentration of the chemoattractant, respectively.

Lebiedz and co-workers [25, 24] considered the problem of externally manipulating the process so as to achieve a particular Gaussian cell distribution. With this aim, a non-zero chemoattractant flux is introduced in the boundary $\frac{\partial c}{\partial \xi}(\xi = L,t) = -c(\xi = L,t) + u(t)$. The problem was formulated as DO problem where the objective is to find $u(t)$ so as to minimize the distance between the distribution of bacteria at final time ($z(\xi,t_f)$) and the desired Gaussian distribution, subject to the system dynamics (Eqns. 15) and bounds on the concentration of chemoattractant. These authors made use of the multiple shooting approach with a local optimization method to solve the problem reporting some difficulties due to the presence of local optima and the large computational cost associated. Here the problem is solved by means of the CVP approach in combination with a sequential hybrid-deterministic method [4]. The optimal solution (Fig. 6) was found in a few seconds.

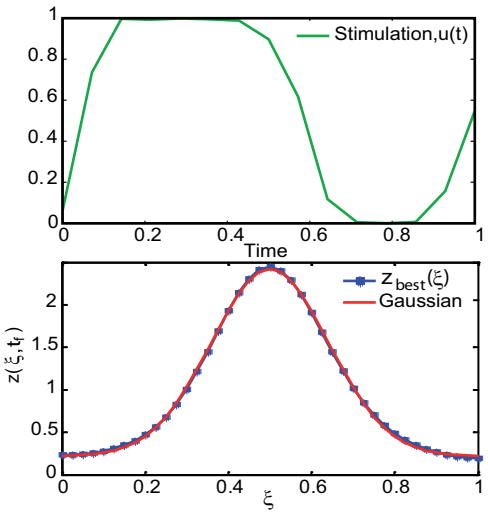

Fig. 6 Optimal stimulation profile and corresponding optimal behaviour for the bacterial chemotaxis problem.

The model in Eqns. (15) with the corresponding boundary and initial conditions was numerically solved using the numerical method of lines with fourth order formulae and a mesh of 41 elements.
In a first approximation to the DO problem the CVP approach with linear control interpolation was combined with a multi-start of a local method (i.e. the solution of the problem with a local method from 500 different initial guesses) which revealed the presence of several local optima.
The global solution was found with a sequential hybrid-deterministic method [4].

5 Conclusions

In this work we have focused on typical optimization problems in systems biology and how their solution may be approached with novel global optimization methods and software tools developed in our group. In particular, a novel optimization approach, the multi-thread eSS, for the solution of large scale optimization problems was presented, together with the AMIGO toolbox devoted to model identification and the DOTcvpSB toolbox devoted to dynamic optimization.

As illustrative examples we considered the optimal enzyme activation in a branched metabolic pathway, the parameter estimation of a large scale dynamic model, an optimal experimental design problem to improve identifiability and the design of optimal stimulation conditions to achieve a given desired result in a reaction-diffusion system.

It should be noted that these software tools can be easily extended to handle multiobjetive optimization problems following the methods described in [38, 39, 45].

Acknowledgements This work was supported by the Spanish MICINN project "MultiSysBio" (ref. DPI2008-06880-C03-02), and by CSIC intramural project "BioREDES" (ref. PIE-201170E018).

References

1. Balsa-Canto, E., Alonso, A.A., Banga, J.R.: Computational procedures for optimal experimental design in biological systems. IET Syst. Biol. **2(4)**, 163–172 (2008)

2. Balsa-Canto, E., Alonso, A.A., Banga, J.R.: An iterative identification procedure for dynamic modeling of biochemical networks. BMC Systems Biology **4:11** (2010)
3. Balsa-Canto, E., Peifer, M., Banga, J.R., Timmer, J., Fleck, C.: Hybrid optimization method with general switching strategy for parameter estimation. BMC Systems Biology **2**, 26– (2008)
4. Balsa-Canto, E., Vassiliadis, V.S., Banga, J.R.: Dynamic optimization of single- and multi-stage systems using a hybrid stochastic-deterministic method. Ind. Eng. Chem. Res. **44**(5), 1514–1523 (2005)
5. Bandara, S., Sclöder, J., Eils, R., Bock, H., Meyer, T.: Optimal experimental design for parameter estimation of a cell signaling model. Plos Comput. Biol. **5**(11), 1–12 (2009)
6. Banga, J.R.: Optimization in computational systems biology. BMC Systems Biology **2**, 47–53 (2008)
7. Banga, J.R., Balsa-Canto, E.: Parameter estimation and optimal experimental design. Essays in Biochemistry **45**, 195–210 (2008)
8. Bartl, M., Li, P., Schuster, S.: Modelling the optimal timing in metabolic pathway activation-use of pontryagin's maximum principle and role of the golden section. Biosystems **101**(1), 67–77 (2010)
9. Biegler, L.T., Cervantes, A., Wätcher, A.: Advances in simulaneous strategies for dynamic process optimization. Chem. Eng. Sci. **57**(4), 575–593 (2002)
10. Bock, H., Plitt, K.: A multiple shooting algorithm for direct solution of optimal control problems. In: Proc. 9th IFAC World Congress, pp. 242–247. Pergamon Press: New York (1984)
11. Bryson, A.E., Ho, Y.C.: Applied optimal control. Hemisphere Pub. Corp, New York (1975)
12. Castiglione, F., Piccoli, B.: Cancer immunotherapy, mathematical modeling and optimal control. J. Theor. Biol. **247**(4), 723–732 (2007)
13. Egea, J.A., Balsa-Canto, E., Garcia, M.S.G., Banga, J.R.: Dynamic optimization of nonlinear processes with an enhanced scatter search method. Ind. & Eng. Chem. Res. **48**(9), 4388–4401 (2009)
14. Egea, J.A., Martí, R., Banga, J.R.: An evolutionary method for complex-process optimization. Comp. & Op. Res. **37**(2), 315–324 (2010)
15. Egea, J.A,, Rodriguez-Fernandez, M., Banga, J.R., Marti, R.: Scatter search for chemical and bio-process optimization. J. Global Optim. **37**(3), 481–503 (2007)
16. Floudas, C.: Deterministic Global Optimization: Theory, Methods and Applications. Kluwer Academics, The Netherlands (2000)
17. Hirmajer, T., Balsa-Canto, E., Banga, J.R.: DOTcvpSB, a software toolbox for dynamic optimization in systems biology. BMC Bioinformatics **10**, 199–213 (2009)
18. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U.: COPASI- A COmplex PAthway SImulator. Bioinformatics **22**(24), 3067–3074 (2006)
19. Jaqaman, K., Danuser, G.: Linking data to models: data regression. Nat. Rev. Mol. Cell Bio. **7**(11), 813–819 (2006)
20. Joly, M., Pinto, J.: Role of mathematical modeling on the optimal control of hiv-1 pathogenesis. AiChe J. **52**(3), 856–884 (2006)
21. Kauffman, K., Prakash, P., Edwards, J.: Advances in flux balance analysis. Current opinion in Biotechnology **14**(5), 491–496 (2003)
22. Klipp, E., Heinch, R., Holzhtte, H.: Prediction of temporal gene expression. metabolic opimization by re-distribution of enzyme activities. Eur. J. Biochem. **269**, 5406–5413 (2002)
23. Kotte, O., Zaugg, J., Heinemann, M.: Bacterial adaptation through distributed sensing of metabolic fluxes. Mol. Sys. Biol. **6**, 355 (2010)
24. Lebiedz, D.: Exploiting optimal control for target-oriented manipulation of (bio)chemical systems: A model-based approach to specific modification of self-organized dynamics. Int. J. Mod. Physics B **19**, 3763–3798 (2005)
25. Lebiedz, D., Maurer, H.: External optimal control of self-organisation dynamics in a chemotaxis reaction diffusion system. IEE Syst. Biol. **2**, 222–229 (2004)
26. Maiwald, T., Timmer, J.: Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics **24**(18), 2037–2043 (2008)

27. Mendes, P., Kell, D.: Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics **14**(10), 869–883 (1998)
28. Moles, C.G., Mendes, P., Banga, J.R.: Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Research **13**, 2467–2474 (2003)
29. Oyarzun, D.A., Ingalls, B., Middleton, R., Kalamatianos, D.: Sequential activation of metabolic pathways: a dynamic optimization approach. Bull. Math. Biol. **71**, 1851–1872 (2009)
30. Pardalos, P., Romeijn, H.E., Tuyb, H.: Recent developments and trends in global optimization. J. Comp. & App. Math. **124**, 209–228 (2000)
31. Pinter, J.: Global Optimization in Action. Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications. Kluwer Academics, Netherlands (1996)
32. Rateitschak, K., Karger, A., Fitzner, B., Lange, F., Wolkenhauer, O., Jaster, R.: Mathematical modelling of interferon-gamma signalling in pancreatic stellate cells reflects and predicts the dynamics of stat1 pathway activity. Cellular signalling **22**, 97–105 (2010)
33. Rodriguez-Fernandez, M., Egea, J.A., Banga, J.R.: Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC Bioinformatics **7**, 483 (2006)
34. Rodriguez-Fernandez, M., Mendes, P., Banga, J.R.: A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems **83**(2-3), 24 (2006)
35. Salby, O., Sager, S., Shaik, O., Kummer, U., Lebiedz, D.: Optimal control of self-organized dynamics in cellular signal transduction. Math. Comp. Model. Dyn. **13**, 487–502 (2007)
36. Schiesser, W.E.: Computational Mathematics in Engineering and Applied Science: ODEs, DAEs and PDEs. CRC Press, Inc. (1994)
37. Schmidt, H., Jirstrand, M.: Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics **22**(4), 514–515 (2006)
38. Sendin, J.O.H., Exler, O., Banga, J.R.: Multi-objective mixed integer strategy for the optimisation of biological networks. IET Systems Biology **4**(3), 236–248 (2010)
39. Sendin, J.O.H., Vera, J., Torres, N., Banga, J.R.: Model based optimization of biochemical systems using multiple objectives: A comparison of several solution strategies. Math. & Comp. Mod. Dyn. Syst. **12**(5), 469–487 (2006)
40. Sugimoto, M., Kikuchi, S., Tomita, M.: Reverse engineering of biochemical equations from time-course data by means of genetic programming. BioSystems **80**, 155–164 (2005)
41. Sutherland, W.: The best solution. Nature **435**, 569 (2005)
42. Talbi, E.G.: Metaheuristics: From Design to Implementation. Wiley Publishing (2009)
43. Toulouse, M., Crainic, T., Sansó, B.: Systemic behavior of cooperative search algorithms. Parallel Computing **30**, 57–79 (2004)
44. Vassiliadis, V.S., Pantelides, C.C., Sargent, R.W.H.: Solution of a class of multistage dynamic optimization problems. 1. problems without path constraints. Ind. Eng. Chem. Res. **33**(9), 2111–2122 (1994)
45. Vera, J., de Atauri, P., Torres, N., Banga, J.R.: Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by saccharomyces cerevisiae. Biotechnol. Bioeng. **83**(3), 335–343 (2003)
46. Vera, J., Balsa-Canto, E., Wellstead, P., Banga, J.R., Wolkenhauer, O.: Power-law models of signal transduction pathways. Cellular signalling **19**, 1531–1541 (2007)
47. Zaslaver, A., Mayo, A., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M., Alon, U.: Just-in-time transcription program in metabolic pathways. Nature Genetics **36**, 486–491 (2004)