An Adjoint-Based Parameter Identification Algorithm Applied to Planar Cell Polarity Signaling

Robin L. Raffard, Keith Amonlirdviman, Jeffrey D. Axelrod, and Claire J. Tomlin

***Abstract*—This paper presents an adjoint-based algorithm for performing automatic parameter identification on differential equation models of biological systems. The algorithm locally solves an optimization problem, in which the cost reflects the deviation between the observed data and the output of the parameterized mathematical model, and the constraints are the governing parameterized equations. The tractability and the speed of convergence (to local minima) of the algorithm are strongly favorable to numerical parameter search algorithms which do not make use of the adjoint. Furthermore, initializing the algorithm with different instantiations of the parameters allows one to effectively search the parameter space. Results of the application of this algorithm to a previously presented mathematical model of planar cell polarity (PCP) signaling in the wings of *Drosophila melanogaster* are presented, and some new insights into the PCP mechanism that are enabled by the algorithm are described.**

***Index Terms*—Adjoint method, differential equations, optimal control, optimization methods, parameter estimation, systems biology.**

I. Introduction

A KEY GOAL of systems biology is the development of mathematical models, at the appropriate level of abstraction, to help understand biological processes. This development usually proceeds in iterative fashion. First, a structure for the model is chosen based on hypotheses about how the system operates. Second, parameters are identified such that simulations of the model reproduce observations of the biological system. Often, the first experiment is to ask if a robust set of parameters exists so that the model reproduces all or most of the observed biological data. If such a parameter set cannot be found, the model structure may be brought into question, and the simulation results may be used to indicate how the structure could be changed [11], [23]. The model is tested against the actual data and for its predictive capabilities. As new data and/or new understanding arises, the structure of the model may be altered and new parameters selected [10].

In protein regulatory networks, the number of states to model is typically large and depends on the number of proteins of interest, the parameter spaces are large, and the most appropriate models are often nonlinear functions of the states and parameters. Due to this and the need to efficiently test the feasibility of different model structures, it is becoming increasingly important to develop fast, efficient, scalable methods for large-scale parameter identification.

In this paper, we present an algorithm for performing automatic parameter identification on differential equation models of biological systems. The algorithm attempts to minimize an objective function which encodes the deviation between the observed data of the system and the output of the parameterized model, with the governing parameterized equations forming the constraints of this optimization problem. The algorithm relies on the adjoint method, which efficiently calculates the gradient of the objective function with respect to the unknown parameters, essentially describing analytically how to minimize the objective by varying the parameters. We augment this gradient-based method by using additional information provided by the derivative of the gradient to give well-conditioned optimization even when the optimal parameter values are several orders of magnitude different from each other. While the adjoint method is a technique familiar to optimal control and has been used to great extent in areas such as aerodynamic design [12], its application to the kinds of ordinary differential equation (ODE) and partial differential equation (PDE) models of protein regulatory networks requires elucidation. We state and justify conditions on the model and the objective function so that the adjoint method may be applied. In addition, we discuss its implementation and derive results about its complexity.

The algorithm is described in the context of a previously presented mathematical model [2], [3] of the signaling network regulating planar cell polarity (PCP) of *Drosophila melanogaster* wing epithelial cells orthogonal to their apical-basal axes. Here, we demonstrate the ability of the adjoint method to significantly speed up the identification of parameters of this model. More importantly, we describe how the method helped lead us to a new hypothesis about how the PCP mechanism works. At each stage, we compare results of our identification with biological data.

This paper is organized as follows. Section II reviews PCP and its mathematical model [2], [3], as well as the current challenges suggested by this model and the data. Section III reviews

Manuscript received January 23, 2007. This work was supported by the National Institutes of Health Joint DMS/BIO/NIGMS Initiative to support research in the Area of Mathematical Biology, by a grant from the Defense Advanced Research Projects Agency BioComputation, and by a Stanford Bio-X award. Preliminary versions of this paper appeared in the *Proceedings of the International Symposium on Advanced Control of Chemical Processes (ADCHEM 2006)* and the *Proceedings of the 45th IEEE Conference on Decision and Control (CDC 2006)*.

R. L. Raffard was with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305-4035 USA. He is now with Barclays Global Investors, San Francisco, CA 94105 USA (e-mail: raffard@stanfordalumni.org).

K. Amonlirdviman was with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305-4035 USA. He is now with McKinsey and Company, Montreal, QC H3B 4W8, Canada (e-mail: amon@stanford.edu).

J. D. Axelrod is with the Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305-5324 USA (e-mail: jaxelrod@stanford.edu).

C. J. Tomlin is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA, and also with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305-4035 USA (e-mail: tomlin@stanford.edu).

Digital Object Identifier 10.1109/TAC.2007.911362

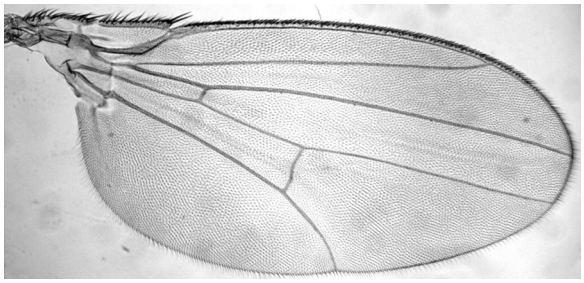


Fig. 1. *Drosophila* adult wing epithelium. The proximal edge is to the left, and the distal edge is to the right.

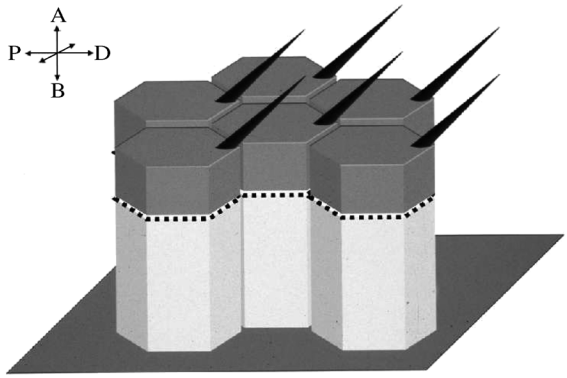


Fig. 2. Diagram shows that each epithelial cell constructs a hair that protrudes from its distal vertex and points *distally*, creating a virtually error-free parallel array.

the parameter identification problem, using a format amenable to the adjoint method. Section IV presents the adjoint method for parameter identification, and Section V details the results of this method applied to PCP. The adjoint method is presented for both ODE and PDE models of PCP: the method for ODEs is developed in the text, and the additional details for PDEs are left to Appendix B.

II. Review of PCP

A. Description of PCP Phenotypes and PCP Proteins

Many epithelial tissues are polarized along an axis orthogonal to the apical-basal axis. For instance, in adult *Drosophila*, each epithelial cell on the wing produces a single hair or trichome; the hairs grow from the *distal* edge (edge of the cell closest to the wing tip) of each cell and all point distally, parallel to the long axis of the wing, as shown in Figs. 1 and 2 (note that all images in this paper follow the convention that the *proximal* side of the cell/wing is to the left of the image and distal is to the right). Genetic analyses have identified a group of proteins that are required to correctly polarize these arrays [1], [20]. These proteins form the PCP signaling pathway. The regular array of hairs is caused by spatially asymmetric distributions of PCP proteins within each cell in the epithelium. The wing epithelial cells aggregate in a hexagonal close-packed array (Fig. 2).

In the presence of cell clones mutant for some PCP genes, the hair polarity in neighboring wild-type cells is disrupted, a phenomenon termed *domineering nonautonomy*. Domineering nonautonomy reverses hair orientation on either the proximal or distal side of the clone in a manner characteristic to the particular mutant protein. Based on the available biological data, a three-tiered model was hypothesized in order to mediate PCP signaling and explain domineering nonautonomy [3], [14]. A diagram of this model is presented in Fig. 3. The first tier is composed of three proteins: the transmembrane protein four-jointed (Fj), and the atypical caherins Dachsous (Ds) and Fat (Ft), providing a global directional signal, or cue, to the second tier. The hypothesis for this mechanism is shown in the bottom-left subfigure of Fig. 3. This diagram indicates that Ft and Ds form a heterodimer across a cell boundary and that Fj promotes Ft and inhibits Ds. While it is known that Fj protein appears in a gradient across the wing, higher at the distal than the proximal side [14], and it is known that Ft affects the localization Fz, the exact mechanism by which the first-tier proteins interact with Fz is not well understood. The second tier, which forms the core PCP components, amplifies the first-tier signal and converts it into cellular asymmetric localization of PCP core proteins. This process, which tends to align neighboring cell polarity, is explained via a feedback-loop mechanism [4], [21], for which the signaling diagram is drawn schematically in the upper right of Fig. 3. This diagram, in which an arrow indicates a positive influence and a line indicates a negative influence, shows the following: Frizzled (Fz), a membrane protein, promotes the localization of Disheveled (Dsh), a cytoplasmic protein, to a membrane; Dsh stabilizes Fz location; Fz promotes the localization of Van Gogh (Vang), a membrane protein, and Prickle (Pk), a cytoplasmic protein, on the membrane of a neighboring cell; Pk and Vang inhibit the recruitment of Dsh to a membrane. Experimentally, it has been observed that, in the steady state, Dsh and Fz proteins localize to the distal edge and Pk and Vang to the proximal edge of all cells in the array, thus the large font in Fig. 3 indicates that the wild-type protein localizes at this location. Thus, the first-tier proteins cause Fz to localize on the distal side of each cell. The third tier directs the tissue specific readout of the cell polarity. Because in Dsh mutants the hair grows from the center of the cell in all null alleles, whereas in other mutants there exist some cells in which the hair grows from the periphery, the hair is assumed to grow at the site of the greatest concentration of Dsh protein [2], [3].

B. PCP Mathematical Modeling

A mathematical model based on the three-tiered model hypothesis was used to demonstrate, through simulation, the feasibility of this hypothesis in its ability to reproduce all of the most characteristic PCP phenotypes [2], [3]. The mathematical program simulates the PCP network from shortly after puparium formation (APF) to about 34 h APF. It mechanistically encodes the logic of the feedback loop of the second tier, while abstracting the effect of the first tier and shortcutting the behavior of the third. In this program, the first tier is represented by a directional cue which biases the direction towards which the feedback loop orients. Positive interactions between proteins are represented as these proteins binding to form protein complexes. For example, the interaction between Fz and Dsh is represented as a reaction forming the complex DshFz, which can interact with other proteins and complexes, and it can undergo a backward reaction that separates it back into its components

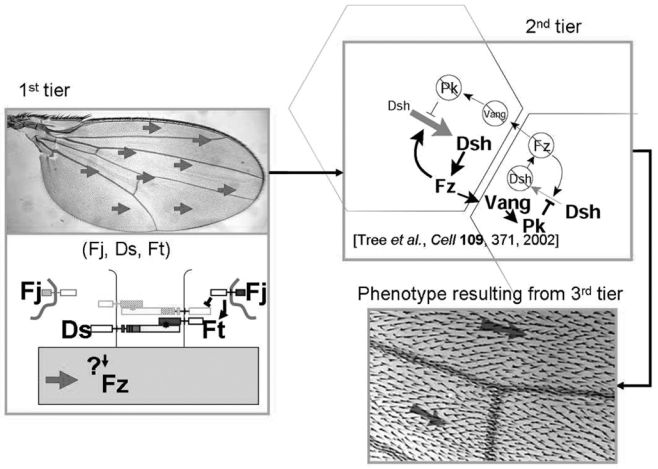


Fig. 3. Three-tiered model for PCP signaling network. The first tier is composed of three proteins: Fj, Ds, and Ft. It provides, via a hypothetical mechanism represented in the bottom-left diagram, a global asymmetry signal to the second tier. This asymmetry signal is represented by arrows in the upper left image. It is read and then amplified by the second-tier proteins (Dsh, Fz, Vang, and Pk), which results in asymmetric localization of these proteins via a mechanism represented in the upper right diagram. The third tier converts the asymmetric accumulation of the second-tier proteins into hair polarity.

Fz and Dsh. The mathematical model includes the four original proteins, as well as six complexes, the last four of which form across the cell boundary with the adjacent cell: DshFz, VangPk, FzVang, DshFzVang, FzVangPk, and DshFzVangPk. While positive influences are encoded by complex formation, negative influences are encoded as terms that aid the reverse reaction. The ten reactions governing the feedback loop mechanism read as follows [2], [3]:

$$\mathrm{Dsh} + \mathrm{Fz} \underset{SB\lambda_1}{\overset{R_1}{\rightleftarrows}} \mathrm{DshFz} \tag{1}$$

$$\mathrm{Fz}^{\dagger} + \mathrm{Vang} \underset{\lambda_2}{\overset{R_2}{\rightleftarrows}} \mathrm{FzVang} \tag{2}$$

$$\mathrm{Vang} + \mathrm{Pk} \underset{\lambda_3}{\overset{R_3}{\rightleftarrows}} \mathrm{VangPk} \tag{3}$$

$$\mathrm{DshFz}^{\dagger} + \mathrm{Vang} \underset{\lambda_4}{\overset{R_4}{\rightleftarrows}} \mathrm{DshFzVang} \tag{4}$$

$$\mathrm{Dsh}^{\dagger} + \mathrm{FzVang} \underset{S^{\dagger}B^{\dagger}\lambda_5}{\overset{R_5}{\rightleftarrows}} \mathrm{DshFzVang} \tag{5}$$

$$\mathrm{Fz}^{\dagger} + \mathrm{VangPk} \underset{\lambda_6}{\overset{R_6}{\rightleftarrows}} \mathrm{FzVangPk} \tag{6}$$

$$\mathrm{FzVang} + \mathrm{Pk} \underset{\lambda_7}{\overset{R_7}{\rightleftarrows}} \mathrm{FzVangPk} \tag{7}$$

$$\mathrm{Dsh}^{\dagger} + \mathrm{FzVangPk} \underset{S^{\dagger}B^{\dagger}\lambda_8}{\overset{R_8}{\rightleftarrows}} \mathrm{DshFzVangPk} \tag{8}$$

$$\mathrm{DshFz}^{\dagger} + \mathrm{VangPk} \underset{\lambda_9}{\overset{R_9}{\rightleftarrows}} \mathrm{DshFzVangPk} \tag{9}$$

$$\mathrm{DshFzVang} + \mathrm{Pk} \underset{\lambda_{10}}{\overset{R_{10}}{\rightleftarrows}} \mathrm{DshFzVangPk}. \tag{10}$$

In (1)–(10), R_i, λ_i, $i = 1, \ldots, 10$, represent, respectively, the forward and backward rates of reaction. The daggered ($\dagger$) variables indicate that the reaction occurs with a protein across the cell membrane in a neighboring cell. S represents the strength of the asymmetry signal. This depends on the location of the cell in which the reaction takes place. For instance, in wild-type cells far away from any first-tier protein clone, S decreases linearly from the proximal edge to the distal edge of each cell. The value by which S decreases in wild-type cells is a parameter of the model. In *fat* clones, S is uniformly equal to 1, indicating that there is no first-tier asymmetry signal when the Ft function is absent. Finally, the quantity B accounts for the inhibition, by Pk and Vang, of the recruitment of Dsh to a membrane. B is a nonlinear function of second-tier protein concentrations and is parameterized by two unknown constants [2], [3]. The state variables of the mathematical model are the local concentrations of the second-tier proteins and complexes, all of which are assumed to be continuous signals in time.

Two instances of the mathematical program have been developed. The first program is based on PDEs [2], [3]. The second program is a simplified version of the first one and is based on ODEs [2].

1) PDE Model: The PDE mathematical model assumes that protein molecules move by diffusion: Dsh and Pk diffuse within the cell interior, while Vang, Fz, and all of the complexes diffuse in the membrane (or shared membranes). The mathematical model is therefore represented by ten reaction–diffusion PDEs. The complete development of the model and results of this analysis are available in [2] and [3]. For example, the rate of change of Dsh concentration is

$$\frac{\partial[\mathrm{Dsh}]}{\partial t} = -P_1 - P_5^{\dagger} - P_8^{\dagger} + \mu_{\mathrm{Dsh}}\Delta[\mathrm{Dsh}] \tag{11}$$

in which

$$\begin{aligned} P_1 &= R_1[\mathrm{Dsh}][\mathrm{Fz}] - SB\lambda_1[\mathrm{DshFz}] \\ P_5 &= R_5[\mathrm{Dsh}]^{\dagger}[\mathrm{FzVang}] - S^{\dagger}B^{\dagger}\lambda_5[\mathrm{DshFzVang}] \\ P_8 &= R_8[\mathrm{Dsh}]^{\dagger}[\mathrm{FzVangPk}] - S^{\dagger}B^{\dagger}\lambda_8[\mathrm{DshFzVangPk}]. \end{aligned} \tag{12}$$

Therefore, by noting $X(t,s) = ([\mathrm{Dsh}](t,s), \ldots, [\mathrm{DshFzVangPk}](t,s))$, the vector of all protein and protein complex concentrations at time t and location s, the PCP PDE model reads

$$\frac{\partial X(t,s)}{\partial t} = P(X(t,s),\theta) + \mu(\theta)\Delta X(t,s) \tag{13}$$

in which $\theta \in \mathbb{R}^{37}$ represents the vector of all parameters. It includes the 20 reaction rates, the ten diffusion constants, three initial protein concentrations (all concentrations are normalized with respect to Dsh initial concentration), the asymmetry signal strength parameter, the two constants parameterizing the Vang and Pk inhibition, and one parameter describing the initial protein concentration in overexpression clones. Finally, Neumann boundary conditions are enforced at the boundary of the computational domain.

2) ODE Model: In [2], it is shown that all rates of diffusion in the PDE model are either very large (for all complexes except DshFz) or very low (for DshFz). Therefore, the PDE model can be reduced to an ODE model by assuming that the concentrations of all of the complexes, except DshFz, are uniform in the domain in which they diffuse. In order to approximate its low

diffusion, the localization of the complex DshFz is restricted to individual cell edges. To briefly describe the ODE model, let N denote the number of cells. For each cell $i = 1, \ldots, N$, let N_i denote the number of edges of cell i. Since Dsh and Pk are assumed to have a uniform concentration in the interior of each cell, single variables $[\mathrm{Dsh}]_i$ and $[\mathrm{Pk}]_i$ represent their concentration in the interior of each cell. Similarly, Fz, Vang, and VangPk are assumed to be homogeneously distributed in the boundary of the cell. Therefore, single variables $[\mathrm{Fz}]_i$, $[\mathrm{Vang}]_i$, and $[\mathrm{VangPk}]_i$ denote their concentration in the boundary of each cell. Each of the remaining complexes is represented by a separate variable for each edge of the cell. For example, we denote $[\mathrm{DshFzVangPk}]_{i,j}$ as the concentration of DshFzVangPk in the j^{th} edge of the ith cell.

Geometrically, each cell i is thus divided into $N_i + 2$ compartments: one compartment representing the interior of the cell, one representing the boundary of the cell (the N_i edges of the cell, all together), and N_i others representing the N_i individual edges of the cell.

The concentration of each molecule is then governed by an ODE. For example, for Dsh, we have

$$\frac{d[\mathrm{Dsh}]_i}{dt} = \sum_{j=1}^{N_i} \frac{l_j}{C_i} \frac{A^* C_i}{A_i C^*} \left(-P_{1,i,j} - P^{\dagger}_{5,i,j} - P^{\dagger}_{8,i,j} \right) \quad (14)$$

in which

$$\begin{aligned} P_{1,i,j} &= R_1 [\mathrm{Dsh}]_i [\mathrm{Fz}]_i - SB\lambda_1 [\mathrm{DshFz}]_{i,j} \\ P_{5,i,j} &= R_5 [\mathrm{Dsh}]^{\dagger}_i [\mathrm{FzVang}]_{i,j} - S^{\dagger} B^{\dagger} \lambda_5 [\mathrm{DshFzVang}]_{i,j} \\ P_{8,i,j} &= R_8 [\mathrm{Dsh}]^{\dagger}_i [\mathrm{FzVangPk}]_{i,j} \\ &\quad - S^{\dagger} B^{\dagger} \lambda_8 [\mathrm{DshFzVangPk}]_{i,j} \end{aligned} \quad (15)$$

and where l_j, A_i, C_i, A^*, C^* are respectively the length of edge j, the area and the total edge length of cell i, and the area and total edge length of a hexagonal cell of unit edge length.

Denoting $X_i = ([\mathrm{Dsh}]_i, [\mathrm{Pk}]_i, \ldots, [\mathrm{DshFzVangPk}]_{i,N_i})$, the vector of all concentrations in each cell i, and $X = (X_1, \ldots, X_N) \in \mathbb{R}^n$ the vector of all concentrations in the network, the ODE model is represented as

$$\frac{dX(t)}{dt} = f(X(t), \theta), \quad X(0) = X_0(\theta) \quad (16)$$

in which $\theta \in \mathbb{R}^{27}$ is the vector of all parameters, containing all the parameters of the PDE model except the ten diffusion rates.

In conclusion, the ODE model requires solving $5+5N_i$ ODEs (one ODE for [Dsh], [Pk], [Fz], [Vang], [VangPk] and N_i for each of the other five complexes) in each cell i. In the case of a hexagonal cell, for instance, this represents 35 ODEs. With as many as 600 cells being simulated for certain *fat* phenotypes [2], [3], the model requires solving systems of ODEs containing more than 20 000 variables.

C. Previous Results and Current Challenges

In [2] and [3], parameter values were identified using images of hair polarity, provided at final time $T = 34$ h of the signaling process. Namely, parameters were constrained to result in the desired qualitative features of the hair pattern of 17 phenotypes. These phenotypes included overexpression and loss-of-function clones of the four second-tier proteins as well as four *fat* clones. The Nelder–Mead simplex method [16] was used to attempt to minimize an objective function composed of quadratic penalty functions corresponding to these feature constraints, to produce a feasible solution set of parameters. Almost all of the characteristic PCP phenotypes were reproduced by the ODE model. However, in the simulations based on irregular geometry, many of the fat clones were not accurately reproduced. Recall that Fat (Ft) is a protein belonging to the first-tier group of PCP proteins, which are responsible for the localization of the asymmetry signal. Fig. 4(a) and (b) shows the localization of the asymmetry signal (represented by red arrows) as it was encoded in [2] and [3]. The strength of the signal is part of the parameters which were identified by the Nelder–Mead algorithm. Fig. 4(c) shows the experimental image of one of the four *fat* clones which were simulated. Note that this clone exhibits characteristic swirling patterns, resulting from the cells aligning their polarity with that of their neighboring cells, even when global signaling input is lost. Fig. 4(d) presents the result of its simulation. Swirling patterns were not reproduced.

At this stage, the challenges are twofold. The first challenge is to understand the reasons for this mismatch. Namely, is it due to an inexact parameter identification or is it due to an incorrect model structure? The second challenge is to derive a mathematical model able to reproduce all the phenotypes exhibited by the *fat* clones as well as all the phenotypes previously reproduced.

We will see that using an adjoint-based parameter identification algorithm which significantly speeds up the parameter identification task has allowed us to overcome these two challenges.

III. Parameter Identification Problem

This section describes the parameter identification problem. Recall that the unknown parameters of the model are the forward and backward rates of reaction, the parameter encoding the asymmetry signal, initial concentrations of the proteins, and parameters representing the strength of the Pk and Vang inhibition. Also recall that experimental data consist of images of hair polarity which are provided at final time T of the signaling process. In our PCP model, hair polarity is predicted based on the Dsh concentration in the cells and is stored in a vector $Y^{\mathrm{model}} \in [-1, 1]^N$ comprising as many entries as simulated cells and calculated by

$$Y^{\mathrm{model}} = g(X(T)) \quad (17)$$

in which $g : \mathbb{R}^n \to \mathbb{R}^N$ is a twice differentiable function, which gives a score of 1 to a cell with Dsh localization only on the distal side, -1 to a cell with Dsh localization only on the proximal side, and an intermediate score to a cell with Dsh localized on both sides. Similarly, the data $Y^{\mathrm{obs}} \in \mathbb{R}^N$ is a vector with entries ranging from -1 to 1: -1 for cells with reverse polarity and 1 for cells with correct polarity. The problem of identifying the unknown parameters is one of finding, among our parameterized set of PCP models, the model which best explains the experimental data. Therefore, it consists of minimizing the prediction error, i.e., the deviation between the observed data and

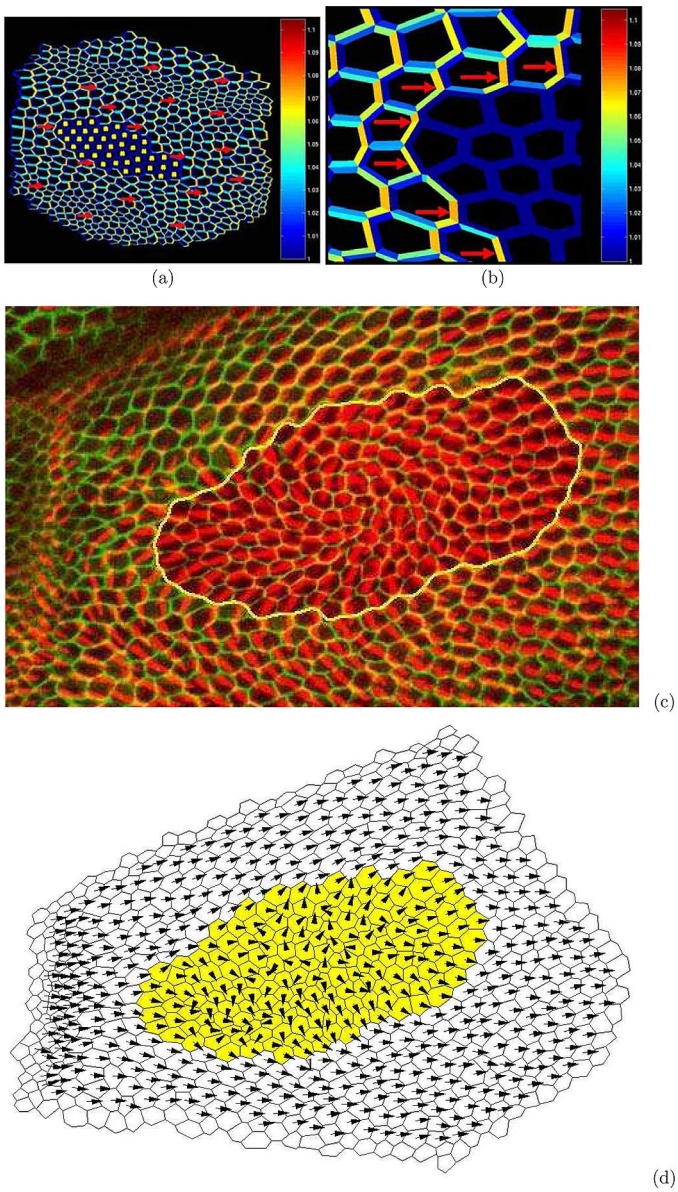


Fig. 4. (a) Direction of the asymmetry signal (red arrow) used in the simulation of the *fat* clones as in [2]. "Cool colors" indicate low Ft concentrations, and "warm colors" high Ft concentrations. The cells in the clone are indicated by yellow dots. The asymmetry signal points from proximal to distal sides in the background and is zero inside the clone. (b) Enlargement of (a) at the clone boundary. The edges of the *fat* clone mutant cells are uniformly colored in blue. (c) Clone exhibiting swirling patterns (courtesy of Dr. D. Ma). The boundary of the clone is colored in yellow. (d) Simulation of the *fat* clone using the former asymmetry signal model shown in (a) and using parameters identified by the Nelder–Mead simplex method. The cells colored in yellow indicate the clone.

the output of the parameterized model. Mathematically, it reads in the case of the ODE model

$$\begin{aligned} &\text{minimize} \quad J(\theta) = \left\| g\left(X(T)\right) - Y^{\text{obs}} \right\|_2^2 \\ &\text{subject to} \quad \frac{dX(t)}{dt} = f\left(X(t), \theta\right), \quad X(0) = X_0(\theta) \end{aligned} \tag{18}$$

in which $\theta \in \mathbb{R}^d$ represents the vector of parameters. In (18), the prediction error is measured in the L^2 norm. Besides its mathematical convenience, such a norm is often chosen because it recovers the maximum-likelihood criterion in the case in which data are altered by white noise [13]. In the case in which the noise is Gaussian with covariance Σ, then the norm L^2 in (18) should be replaced by the norm $\|\cdot\|_{\Sigma^{-1}}$, where $\|x\|_{\Sigma^{-1}} = (x^\top \Sigma^{-1} x)^{1/2}$. Without loss of generality, we assume in this paper that Σ is the identity matrix.

If the PDE model is used, the parameter identification problem reads

$$\begin{aligned} &\text{minimize} \quad J(\theta) = \left\| g\left(X(T)\right) - Y^{\text{obs}} \right\|_2^2 \\ &\text{subject to} \quad \frac{\partial X(t,s)}{\partial t} = P\left(X(t,s), \theta\right) + \mu(\theta) \Delta X(t,s). \end{aligned} \tag{19}$$

In conclusion, the parameter identification task for PCP has been posed as an optimization program involving differential equations (DEs). This control-based approach is widely used in the sciences and in engineering [7], [15], [18], [22]. In Section IV, we will present a fast and scalable algorithm to solve a wide class of DE optimization programs, thus allowing us to perform parameter identification on general DE-driven systems.

IV. Solution Method Via the Adjoint Method

Problems (18) and (19) belong to the class of nonlinear optimization programs involving DEs. These problems are generally written in the following form:

$$\begin{aligned} &\text{minimize} \quad J(U) = E\left[\int_0^T k(X(t), U(t))\, dt + l\left(X(T)\right) \right] \\ &\text{subject to} \quad \text{ODE}(X, U) = 0 \\ &\qquad\qquad \text{or PDE}(X, U) = 0 \\ &\qquad\qquad \text{or SDE}(X, U) = 0. \end{aligned} \tag{20}$$

The parameter vector θ has been replaced by an input function $U : [0, T] \to \mathbb{R}^d$ to stress the fact that the control input of DE optimization programs are in general functions of time. $E[\cdot]$ indicates expectation and is relevant only in the stochastic differential equation (SDE) case. If inequality constraints are added on the state variable or on the control parameters, these constraints can be easily handled using a barrier method in which the constraints are incorporated into a logarithmic term added to the cost function [6]. Provided the DEs of (20) can be solved numerically, it is usually possible to (locally) solve the optimization program (20) efficiently using an adjoint-based quasi-Newton method. Although this method is applicable to all of ODEs, PDEs, and SDEs [17], we present the algorithm on our PCP ODE model, and we show in Appendix B how it extends to the PDE model. From a control theory point of view, the algorithm is based on the Pontryagin maximum principle, as it attempts to iteratively solve the necessary conditions for optimality. From an optimization point of view, the algorithm consists of a quasi-Newton descent algorithm, in which the gradient of the cost function is efficiently computed via the adjoint method [12].

A. Computation of the Gradient

Although rarely used directly as a descent direction, the gradient plays a key role in the implementation of the descent algorithm. The method developed in this section allows for fast, scal-

able computation of the gradient and consists of two steps. First, it consists of computing the first order variation (or derivative) of the cost function when the control input $\theta \in \mathbb{R}^d$ or $U : [0,T] \to \mathbb{R}^d$ varies. Application of the calculus of variations to the cost function and to the ODE model leaves the variation of the cost function expressed as a function of the variation of the control input and as a function of the variation of the state variable, which cannot be directly manipulated. In order to derive the expression of the gradient of the cost function, we thus need to replace the variation of the state variable by the variation of the control. The adjoint method performs this task very efficiently.

1) Derivative of the Cost Function: For systems driven by ODEs, there exists a general theorem proving differentiability of the cost function in (20). A similar theorem exists for systems governed by SDEs but not for systems governed by PDEs, which require case-by-case analyses. For ODEs, the differentiability theorem reads as follows and requires the following regularity conditions.

Assumption 1 (Regularity Conditions on the Dynamics): Let the ODE be expressed in the general form $\dot{X}(t) = f(X(t),U(t))$, $X(0) = X_0$. $U : [0,T] \to \mathbb{R}^d$ is a measurable, bounded function, representing the control. $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ is continuously differentiable with bounded derivative: $\forall (x,u) \in \mathbb{R}^n \times \mathbb{R}^d$, $|f_x(x,u)| \leq K_{f_x}$ and $|f_u(x,u)| \leq K_{f_u}$ in which K_{f_x} and K_{f_u} are positive constants.

Assumption 2 (Regularity Conditions on the Cost Function): k and l in (20) are locally continuously differentiable, which means that k and l are continuously differentiable on an open subset of $\mathbb{R}^n \times \mathbb{R}^d$ containing $\{(X(t),U(t)) : t \in [0,T]\}$.

Theorem 1 (Gâteaux Derivative of (20) Cost Function): Under Assumptions 1 and 2, the cost function J is differentiable with respect to the control U and the Gâteaux derivative of J at U in any direction $\widehat{U} : [0,T] \to \mathbb{R}^d$, bounded, is

$$\begin{aligned}\lim_{h\to 0} \frac{J(U+h\widehat{U}) - J(U)}{h} &= \nabla l\left(X(T)\right)\widehat{X}(T) + \int_0^T k_x\left(X(t),U(t)\right)\widehat{X}(t) \\ &\quad + k_u\left(X(t),U(t)\right)\widehat{U}(t)dt \end{aligned} \tag{21}$$

in which $\widehat{X}$ is the derivative of X with respect to U and is the solution of

$$\begin{aligned}\frac{d\widehat{X}(t)}{dt} &= f_x\left(X(t),U(t)\right)\widehat{X}(t) + f_u\left(X(t),U(t)\right)\widehat{U}(t), \\ \widehat{X}(0) &= 0. \end{aligned} \tag{22}$$

Proof: See [5], [17], or [19]. ∎

Unfortunately, the dynamics of the PCP ODE, f, is quadratic and therefore does not satisfy the growth condition of Assumption 1. Nevertheless, we can rely on Theorem 2, which is an adaptation of Theorem 1 to the case in which the dynamics do not necessarily satisfy Assumption 1, but in which the state variable is bounded.

Assumption 3 (Bound on the Concentrations): The state variable is bounded: there exists $M > 0$, such that for all $U : [0,T] \to \mathbb{R}^d$, measurable and bounded, $X(t) \leq M$, $t \in [0,T]$.

Assumption 3 is satisfied for the PCP ODE model. The model indeed conserves the total mass in the cell network and therefore the concentrations in each compartment remain bounded no matter what the input parameters θ are. Algebraically, if $n(t)$ denotes the cumulated number of Dsh, Vang, Pk, and Fz molecules in the cell networks, then $n(t)$ is expressed as a weighted sum of protein concentrations: $n(t) = \sum_{i=1}^n \alpha_i X_i(t)$, in which $\alpha_i > 0$ is the size of the compartment corresponding to the state variable indexed by i. Therefore, because $n(t)$ is constant equal to $n(0)$, there exists $M \in \mathbb{R}_+$ such that for all $\theta \in \mathbb{R}^d$, $|X(T)| \leq M$.

Theorem 2 (Gâteaux Derivative of (18) Cost Function): Under the conditions that f and X_0 are continuously differentiable, the result of Theorem 1 holds in the setting of problem (20). Applied to problem (18), the cost function J is differentiable with respect to $\theta \in \mathbb{R}^d$ and the directional derivative of J at θ in any direction $\widehat{\theta} \in \mathbb{R}^d$ is

$$\begin{aligned}\lim_{h\to 0} \frac{J(\theta+h\widehat{\theta}) - J(\theta)}{h} &= 2\left(g\left(X(T)\right) - Y^{\text{obs}}\right)^\top \nabla g\left(X(T)\right)\widehat{X}(T) \end{aligned} \tag{23}$$

in which $\widehat{X}$ is the derivative of X with respect to θ and is the solution of

$$\begin{aligned}\frac{d\widehat{X}(t)}{dt} &= f_x\left(X(t),\theta\right)\widehat{X}(t) + f_\theta\left(X(t),\theta\right)\widehat{\theta} \\ \widehat{X}(0) &= \nabla X_0(\theta)\widehat{\theta}. \end{aligned} \tag{24}$$

Proof: See Appendix A. ∎

2) Rewriting the Derivative With the Adjoint Method: The expression of the derivative of the cost function (23) is not convenient as it includes the variation of the state variable with respect to the parameters, which cannot be directly manipulated. The adjoint method is an ingenious technique which allows us to replace this expression by an expression including only the parameter variation $\widehat{\theta}$. It proceeds as follows. Let us take the inner product of an arbitrary $\mathbb{R}^n$-valued continuously differentiable function of time p, called an adjoint process, with the linearized dynamics (24), given as follows:

$$\begin{aligned}\int_0^T p(t)^\top \frac{d\widehat{X}(t)}{dt}dt &= \int_0^T p(t)^\top f_x\left(X(t),\theta\right)\widehat{X}(t)dt \\ &\quad + \int_0^T p(t)^\top f_\theta\left(X(t),\theta\right)\widehat{\theta}dt. \end{aligned} \tag{25}$$

Integrating the left-hand side by parts yields

$$\begin{aligned} p(T)^\top \widehat{X}(T) - p(0)^\top \widehat{X}(0) - \int_0^T \widehat{X}(t)^\top \frac{dp(t)}{dt}dt \\ = \int_0^T \widehat{X}(t)^\top f_x\left(X(t),\theta\right)^\top p(t)dt \\ + \int_0^T p(t)^\top f_\theta\left(X(t),\theta\right)\widehat{\theta}dt. \end{aligned} \tag{26}$$

Making the following choice for the adjoint process p:

$$\begin{aligned} p(T) &= \nabla h\left(X(T)\right)^\top \\ \frac{dp(t)}{dt} &= f_x\left(X(t),\theta\right)^\top p(t) \end{aligned} \tag{27}$$

the directional derivative can be rewritten as

$$\begin{aligned} &\lim_{h\to 0} \frac{J(\theta + h\widehat{\theta}) - J(\theta)}{h} \\ &\quad = \int_0^T p(t)^\top f_\theta\left(X(t),\theta\right) dt\, \widehat{\theta} + p(0)^\top \nabla X_0(\theta)\widehat{\theta} \end{aligned} \tag{28}$$

and therefore the gradient is expressed as

$$\nabla J = \int_0^T p(t)^\top f_\theta\left(X(t),\theta\right) dt + p(0)^\top \nabla X_0(\theta). \tag{29}$$

Besides its importance in the implementation of the descent algorithm, the gradient also provides useful robustness information: it is equal to the sensitivity of the cost function with respect to the parameters. Therefore, if the ith component of the gradient is large, then the cost function will be very sensitive to a change in the ith parameter. Reversely, if the ith component of the gradient is close to zero, then the simulated phenotype is robust with respect to a change in the ith parameter. The adjoint variable p provides complementary information. It represents the sensitivity of the cost function with respect to the state variable.

3) Implementation: The code running the ODE-based simulation of PCP is composed of several routines. A schematic representation of the code is shown in Fig. 5. Besides a routine solving the ODE (called DE), a routine is responsible for setting the initial protein levels (initialConditions) and another routine post-processes the final protein concentrations, computes cell polarities, and evaluates the cost function (costFunction). When computing the gradient of the cost function with respect to the parameters, the sensitivity of the output of each routine needs to be computed with respect to the parameters as well as with respect to the state variable. Therefore, associated to each routine is an additional routine performing this sensitivity computation. The results of these computations are then collected and used to evaluate the gradient

$$\nabla J = \frac{\partial g(X_T,\theta)}{\partial \theta} + \frac{\partial g(X_T,\theta)}{\partial X_T}\frac{\partial X_T}{\partial \theta} + \frac{\partial g(X_T,\theta)}{\partial X_T}\frac{\partial X_T}{\partial X_0}\frac{dX_0}{d\theta}. \tag{30}$$

The routines sensitivityCostFunction and sensitivityInitialConditions, respectively, evaluate the terms $\partial g(X_T,\theta)/\partial\theta$ and $dX_0/d\theta$. These routines perform analytical differentiation of the output of their associated routine. The routine sensitivityDE evaluates the terms $(\partial g(X_T,\theta)/\partial X_T)(\partial X_T/\partial\theta)$ and $(\partial g(X_T,\theta)/\partial X_T)(\partial X_T/\partial X_0)$. This routine uses the adjoint method.

4) Computational Complexity: The adjoint method drastically reduces the complexity of the gradient computation in comparison with other classical methods such as gradient calculation via finite differences or via sensitivity propagation. These differences are given here.

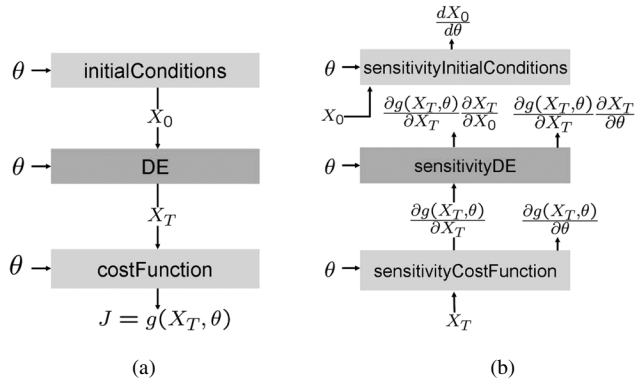


Fig. 5. (a) Schematic representation of the code running the PCP ODE model. (b) Diagram of the code performing sensitivity analyses for the gradient evaluation. The routine sensitivityDE uses the adjoint method.

- The adjoint method only requires two DE calculations: the governing DE and the adjoint DE. In principle, the complexity of the adjoint equation is similar to the complexity of the governing DE. However, in practice, solving the adjoint equation is slower than solving the governing equation, as the adjoint equation requires the entire time history of the state variable which is only available at the time steps at which the governing DE is simulated. Therefore, interpolations of the state variable values often need to be performed in order to run the adjoint equation, which typically makes the gradient computation three to five times as slow as a governing DE simulation.
- The method of finite differences consists of calculating $\nabla_{\theta_i} J \simeq (J(\theta + he_i) - J(\theta - he_i))/2h$, in which $e_i \in \mathbb{R}^d$, $i = 1,\ldots,d$, is the unit vector with component i equal to 1 and all other components equal to 0. It requires at least $2d$ DE computations. In practice, the complexity of the gradient computation via finite differences is larger than $2d$ times the complexity of a DE simulation. Indeed, if a precision of ϵ is required on the gradient, then, based on $\nabla_{\theta_i} J = (J(\theta+he_i) - J(\theta - he_i))/2h + \mathcal{O}(h^2)$, a precision of $\epsilon^{3/2}$ is required on the evaluation of the cost function, and therefore the DEs need to be computed with greater precision, which adds to the computational time.
- The method of sensitivity propagation consists of analytically differentiating the ODE with respect to θ and solving for $\partial X(t)/\partial\theta \in \mathbb{R}^{n\times d}$ as

$$\begin{aligned} \frac{d\frac{\partial X(t)}{\partial\theta}}{dt} &= f_x\left(X(t),\theta\right)\frac{\partial X(t)}{\partial\theta} + f_\theta\left(X(t),\theta\right) I_d, \\ \frac{\partial X(0)}{\partial\theta} &= \nabla X_0. \end{aligned} \tag{31}$$

  This method has roughly d times the complexity of a DE computation.

B. Second-Order Method

The gradient algorithm is numerically efficient when the problem is well conditioned, meaning that the derivatives in all

directions have the same order of magnitude. In the case of PCP, the parameters are unknown and may range over several orders of magnitude. Therefore, the problem is likely to be poorly conditioned, in which case a second-order method is preferable. A second-order method, such as the Newton method, rescales the variables so that in the new system of variables the problem is well conditioned and, consequently, the descent algorithm is fast, yet no tractable method currently exists for executing the Newton method in optimization programs involving general DEs. However, it is possible to implement a quasi-Newton method [8], [9], in which the second-order derivative of the objective function, called the Hessian, is computed via finite differences on the gradient.

By so doing, we can form an approximate Hessian H and the descent direction is taken as the one which minimizes the quadratic approximation of the objective function: $\delta\theta = -H^{-1}\nabla J$.

C. Summary of the Algorithm

Algorithm 1 (Second-order parameter id. algorithm)

Start *with an initial guess for the parameters* θ^{guess} *and an initial guess for the Hessian* H^{guess}.

Repeat

1. *Solve the governing* (16) *for* X, *using the current parameter vector* θ.
2. *Solve the adjoint equation for* p (27), *using the current* θ *and* X.
3. *Form the gradient* ∇J (29).
4. *Update the Hessian* H *via finite difference between the current gradient and the previous ones.*
5. *Form the descent direction* $\Delta\theta = -H^{-1}\nabla J$.
6. *Line search: compute* $\beta > 0$ *so that* $J(\theta + \beta\Delta\theta)$ *is minimized.*
7. *Update* $\theta := \theta + \beta\Delta\theta$.

until $|\nabla J^{\top} H^{-1} \nabla J|$ *is small.*

Return $\theta^* = \theta$.

As mentioned in Section II-C, the cost function J incorporates 17 phenotypes. The simulation as well as the gradient computation of these 17 phenotypes is performed in parallel. Namely, Algorithm 1 runs on a 33 node Linux cluster (one central node and 32 subnodes). The central node runs the quasi-Newton algorithm (main loop of Algorithm 1). Each time the quasi-Newton algorithm requires gradient or function evaluations, it launches 17 jobs in parallel, corresponding to the 17 phenotypes. The simulations of these 17 phenotypes are independent of each other; however, because the cost function is nonlinear, the computation of the gradient requires collecting all phenotype state variables at the final time, setting the adjoint variable terminal conditions (which depend on the variables of all phenotypes), and then running the adjoint equations in parallel.

Each iteration of Algorithm 1 consists of a coarse one-dimensional minimization (line search of step 6) of the cost function, which is typically terminated after three to six DE system [(16)] computations. Taking into account the computation of the gradient, each iteration of the algorithm requires eight to ten times the computational time of running the governing DE system [(16)]. In terms of convergence, the algorithm generally terminates after 50 iterations; therefore, the algorithm requires on the order of 50×10 objective function evaluations, each one being performed on 17 processors.

Finally, since Algorithm 1 only guarantees convergence to a local minimum, it is usually run for different initial parameter guesses, uniformly spread in the parameter space. For this purpose, several instances of Algorithm 1 are run in parallel.

V. Results

Here, we show that the adjoint-based quasi-Newton method adds significant value to the parameter identification task for PCP in that it enables the use of gradient-based search algorithms, such as the fast converging quasi-Newton method, in a tractable fashion. We compare the convergence of the Nelder–Mead simplex method [2], which avoids the difficulties of calculating the gradient, to the convergence of the adjoint-based quasi-Newton method, thus demonstrating the benefit of using the adjoint-based method.

We also show that this speed-up helps us to better understand the first tier of the PCP signaling network, in particular the localization of the initial asymmetry signal.

A. Adjoint-Based Quasi-Newton Method Versus Nelder–Mead Simplex Method

The first validation of the adjoint-based algorithm is to verify that it searches the parameter space more efficiently than does the parameter identification algorithm used previously [2]. For this purpose, Fig. 6 shows a convergence comparison between the adjoint-based quasi-Newton method and the Nelder–Mead simplex method, which was used in [2]. The goal of both algorithms was to identify a set of parameters reproducing the wild-type phenotype. After 30 function evaluations (ODE computations), the adjoint-based method identified a set of parameters reproducing the wild-type phenotype to a good precision, whereas the Nelder–Mead simplex method failed to identify a satisfactory set of parameters.

B. Matching the fat Clones

Although it is a key part of PCP signaling, the interaction between the first-tier and second-tier proteins is not well understood. Little is known about the role of Ft and Ds in this interaction either, and open questions are legion: what is the wild-type orientation of the heterodimer between Ft and Ds? Do *fat* clones influence the localization of this heterodimer in wild-type cells? What role does this Ft, Ds heterodimer play in the generation of the asymmetry signal? To what extent is the asymmetry signal generated in wild-type cells amplified and then propagated by the second tier in *fat* clone mutant cells? The goal of this section is to shed some light on these questions by efficiently testing, via simulation, possible hypotheses. We start our analysis with the following understanding. In mutant cells, because of the absence of the *fat* function, the second-tier feedback loop propagates the

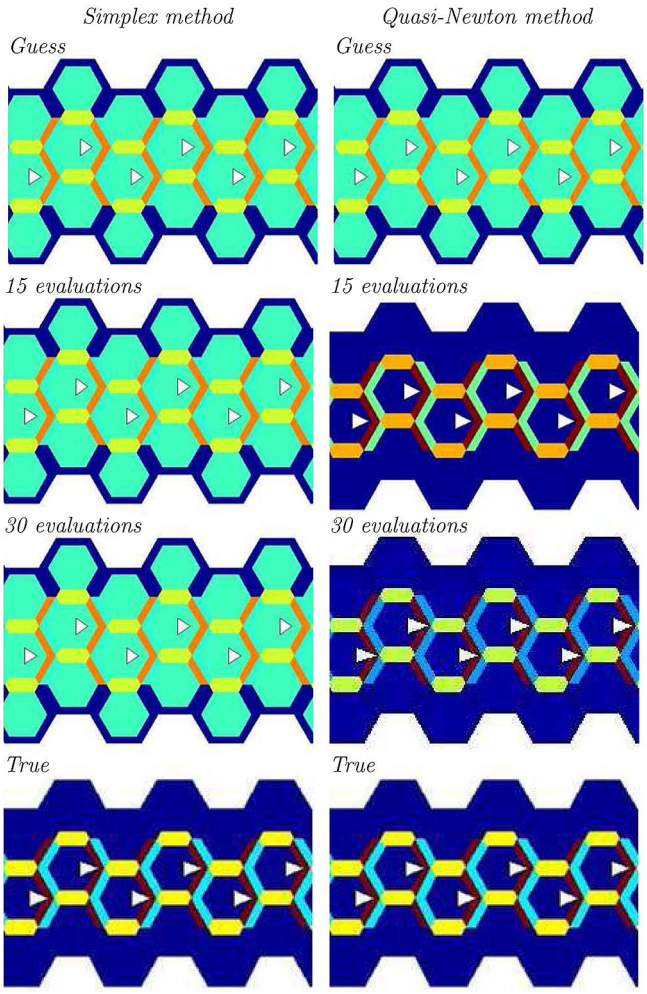


Fig. 6. Comparison between the simplex method and the quasi-Newton method for the parameter identification problem. "Cool colors" indicate low Dsh concentrations, and "warm colors" indicate high Dsh concentrations. After 30 ODE system computations, the quasi-Newton method has almost recovered the wild-type phenotype.

asymmetry signal generated in wild-type cells by the first tier, and this signal gradually attenuates as it penetrates deeper into the clone. Therefore, in small-size clones [see Fig. 10(c)], the polarity is not disrupted as it is easily propagated from wild-type cells to all mutant cells. However in large size clones [see Fig. 10(a)], wild-type asymmetry information does not manage to penetrate deep enough inside the clone: the second tier locally aligns polarity among neighboring cells, but polarity is progressively disrupted. By challenging our model against various types of *fat* clones, we will thus test the modeling of the second-tier feedback loop propagation mechanism inside the clone as well as the strength of the asymmetry signal in wild-type cells. Furthermore, we expect that the detailed *fat* clone patterns will provide useful information regarding the localization of Ft and its effect on the asymmetry signal generation.

1) Testing the Former Asymmetry Signal Model: In a first step, we would like to test the asymmetry signal model which we previously used in Section II-C and which was introduced in [2]. According to [2], we suppose that Ft preferentially localizes on the proximal side of each cell and that the asymmetry signal points away from Ft and toward Ds accumulation. This model is

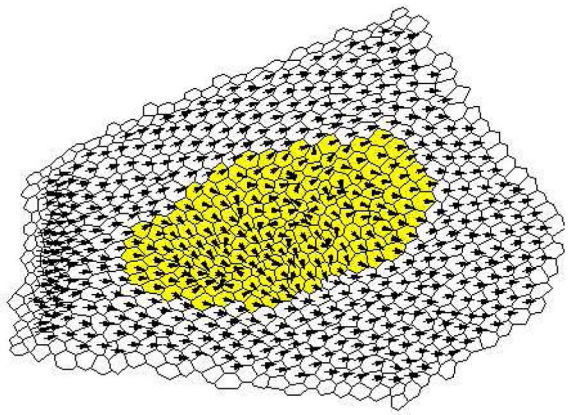


Fig. 7. Simulation of the *fat* clone using the former asymmetry signal model shown in Fig. 4(a) and using parameters identified by the adjoint-based quasi-Newton method. The simulated phenotype does not present any improvement with respect to Fig. 4(d) and suggests that the model structure of the asymmetry signal in Fig. 4(a) should be revised.

in agreement with the hypothetical diagram of Fig. (9a). Therefore, in wild-type cells, the asymmetry signal points from the proximal side to the distal side of each cell. In mutant cells, there is no asymmetry signal, and no specific clone boundary effect is implemented. Fig. 4(a) and (b) shows the localization of the asymmetry signal in this model.

The adjoint-based algorithm was run in order to try to reproduce the 17 training phenotypes of [2]. The simulation, after convergence of the adjoint-based algorithm, of the clone of Fig. 4(c), is shown in Fig. 7. Results show that swirling patterns were still not reproduced.

In conclusion, the adjoint-based quasi-Newton algorithm did not succeed in reproducing the desired *fat* clone phenotype based on the asymmetry signal model structure shown in Fig. 4(a) and (b). Note that the adjoint-based algorithm only guarantees local convergence; therefore, it may have been trapped in a local minimum. Nevertheless, because we performed the search algorithm with various initial parameter values, we conjectured that no parameter could properly explain the desired phenotype, and we concluded that the asymmetry signal model structure had to be modified.

2) New Asymmetry Signal Model: Discussions including contributions from Prof. Mike Simon in the Department of Biological Sciences at Stanford University lead us to the conjecture that the cause of the problem lay in our incorrect implementation of the asymmetry signal at the clone boundary. Inspection of fat clone boundaries proved that polarity tended to be orthogonal to the boundary of the clones, as shown in Fig. 8, which was in disagreement with the result of our simulation (Fig. 7). The second step of our analysis was therefore to try refining the implementation of the clone boundary conditions. Fig. 9(a)–(e) illustrates our reasoning. As hypothesized in the diagram of Fig. 9(a), Ft and Ds bind across the cell boundary. In mutant cells located at the boundary of the clone, Ds can only bind with Ft from neighboring wild-type cells. Therefore, in mutant cells at the clone boundary, Ds accumulates at the boundary of the clone and recruits Ft from wild-type cells, in order to form a heterodimer at the clone boundary. The excessive accumulation of this heterodimer at the clone boundary is

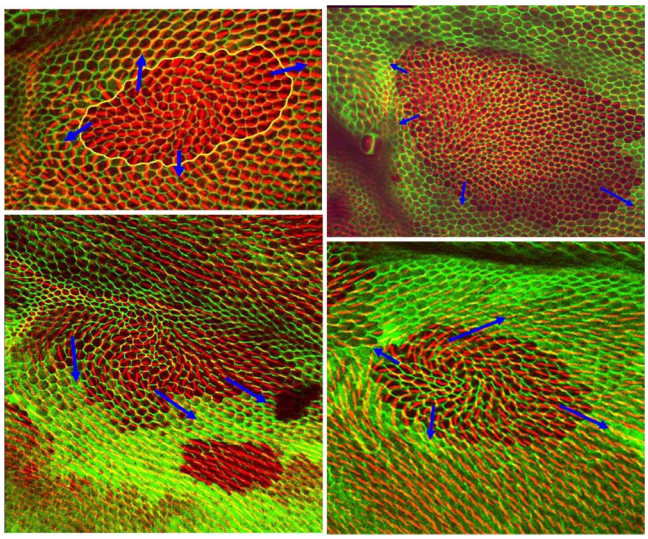


Fig. 8. Four images of large *fat* clones (courtesy of Dr. D. Ma). The hair polarity seems to be orthogonal to the clone boundary and to point outward from the clone, as indicated by the black arrows.

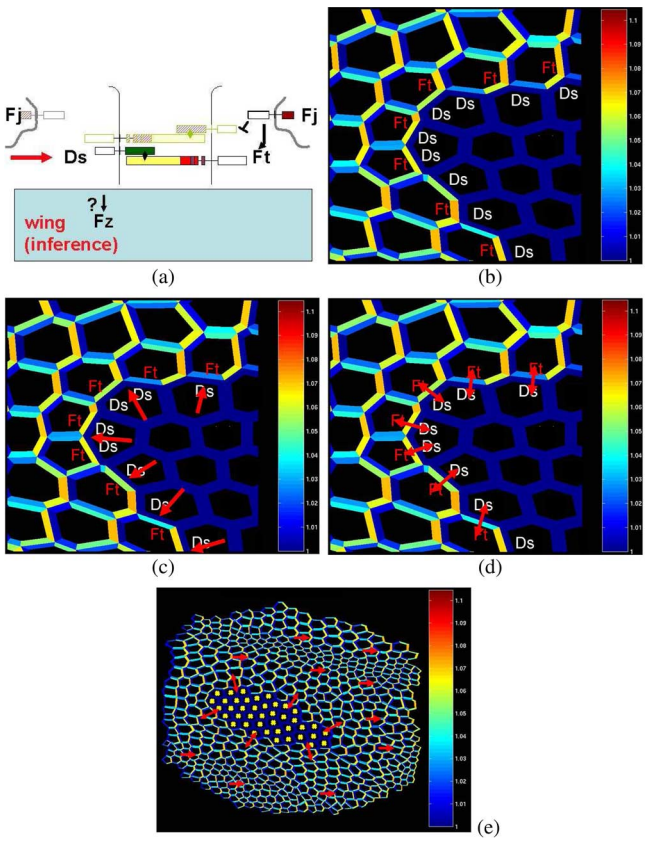


Fig. 9. (a) Hypothetical inference diagram between first-tier PCP proteins. The asymmetry signal localizes by the cell edge at which Ds binds with Ft across the cell boundary to form a heterodimer. (b) Accumulation of the heterodimer between Ds and Ft at the boundary of the *fat* clone. Because of the absence of the *fat* function in the clone, Ds is promoted to the clone boundary in order to bind with Ft from the wild-type cells. (c) Based on (a), the asymmetry signal should point away from the clone orthogonally to the clone boundary. (d), (e) Model structure allowing the asymmetry signal to point either towards or outwards the clone at the clone boundary.

predicted to distort the asymmetry signal in the wild-type cells at the clone interface such that it would tend to be orthogonal to the clone boundary and point away from the clone. We chose to test this inference by modifying the asymmetry signal so that

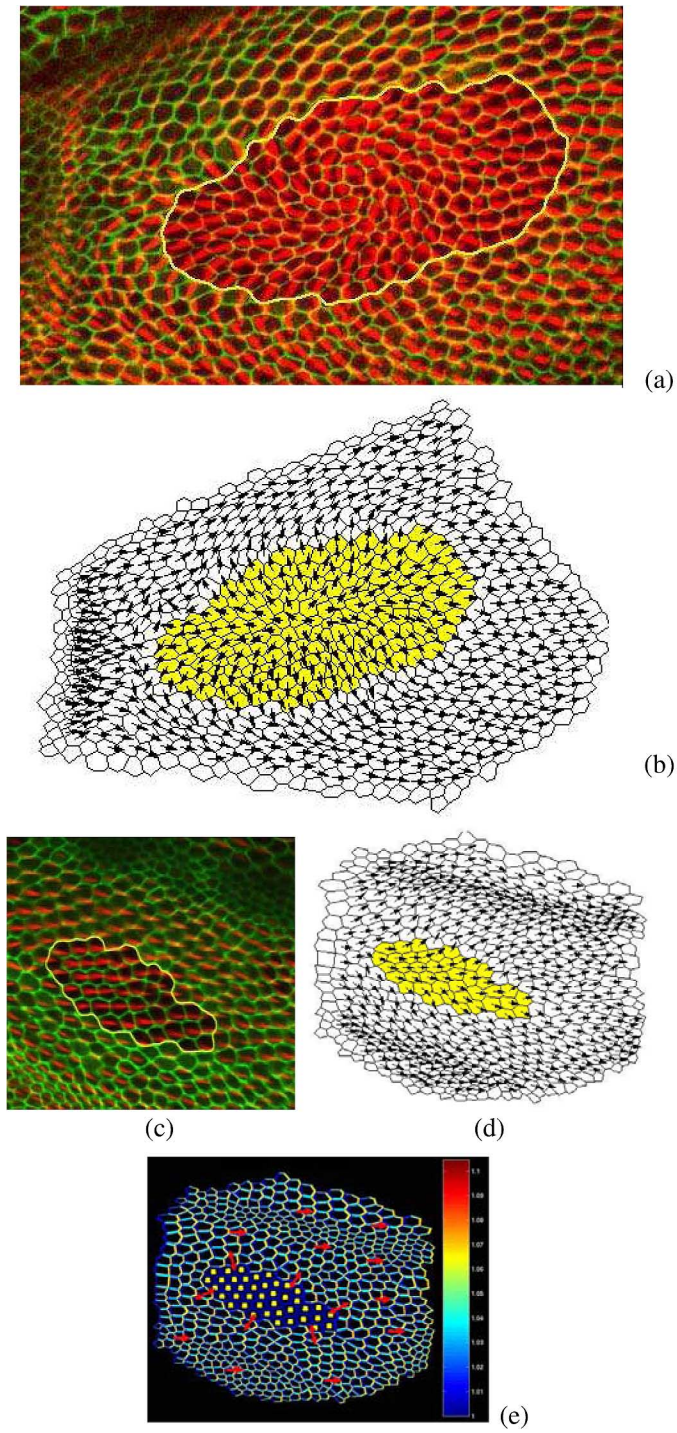


Fig. 10. (a) *fat* clone exhibiting swirling patterns (courtesy of Dr. D. Ma). (b) Simulation of the clone in (a), using new asymmetry signal model structure. (c) *fat* clone exhibiting wild-type phenotype (courtesy of Dr. Dali Ma). (d) Simulation of the clone in (c), using new asymmetry signal model structure. (e) Direction of asymmetry signal obtained by the adjoint-based parameter identification algorithm with new model structure.

it reflects this orthogonal orientation and by allowing it to vary in magnitude in either direction.

In order to reflect our hypothetical understanding of the first tier and our *fat* clone observations, we modify the asymmetry signal model as follows. The direction of the asymmetry signal is set orthogonal to the clone boundary in order to account for the probable accumulation of the heterodimer between Ds and

Ft, but the value of the signal at the boundary is unknown and consists of a new parameter to be identified. Therefore, within the new model structure, the asymmetry signal could point towards or away from the clone.

Based on this new model, the adjoint-based algorithm was run in order to try matching the 17 phenotypes of [2]. All phenotypes were reproduced to a good precision. Since the reproduction of all loss-of-function and overexpression clones of the second-tier proteins were already acquired in [2], [3], we only present the results of the *fat* clones. Fig. 10(b) and (d) show the results of our simulations for two of these *fat* clones, one of large size and one of small size. Both are good reproductions of the biological data. Fig. 10(e) shows the resulting orientation of the asymmetry signal within the model structure of Fig. 9(e). This orientation corroborates our observations of Figs. 8. Following this dialectic of theory, simulation and experiment, the asymmetry signal has therefore been incorporated to reflect a more detailed prediction of the first tier model as shown in Fig. 9(a). The ability of this more detailed implementation to better reproduce the experimental results provides additional evidence for this structure of the first-tier model.

VI. Conclusion

The construction of mathematical models in systems biology usually involves two phases: 1) the construction of the model structure based on biological insights and 2) the identification of the unknown constants parameterizing the model structure. The more rapidly the parameter identification can be performed, the more iterations on the model structure can be accomplished, thus allowing for better testing of biological hypotheses.

Through the example of PCP, this paper has shown the importance of using adjoint-based control in the parameter identification of protein networks driven by DEs. The simulation of such complex systems is indeed computationally very expensive, requiring the parameter identification task to be performed with as few system simulations as possible. Because of its fast convergence (in terms of system simulation numbers), the adjoint-based quasi-Newton algorithm meets this requirement.

Finally, we have shown that using this parameter identification algorithm allowed us to quickly test hypothetical model structures for the first-tier proteins of PCP.

Appendix A
ODE Model Gradient Computation

Proof of Theorem 2: Let us denote by $\widetilde{\theta} = \theta + h\widehat{\theta} \in \mathbb{R}^d$ the perturbed control and with similar notation $\widetilde{X} \in L^2(0,T,\mathbb{R}^n)$ the perturbed state variable: $\dot{\widetilde{X}}(t) = f(\widetilde{X}(t), \widetilde{\theta}(t))$, $\widetilde{X}(0) = X_0(\widetilde{\theta})$. Furthermore, let $\epsilon_X(t) = \widetilde{X}(t) - X(t)$ be the deviation between the perturbed and the nominal state variables. Since f is continuously differentiable, then

$$\dot{\epsilon_X}(t) = \int_0^1 f_x\left(X(t) + \lambda\epsilon_X(t), \theta + \lambda h\widehat{\theta}\right)\epsilon_X(t)d\lambda + \int_0^1 f_\theta\left(X(t) + \lambda\epsilon_X(t), \theta + \lambda h\widehat{\theta}\right)h\widehat{\theta}d\lambda. \quad (32)$$

Furthermore, $\epsilon_X(0) = \int_0^1 \nabla X_0(\theta + \lambda h\widehat{\theta})h\widehat{\theta}d\lambda$. Since $X(t)$ and $\widehat{X}(t)$ are bounded by M according to Assumption 3, and since f_x, f_θ and ∇X_0 are continuous, there exist K_{f_x}, K_{f_θ}, and $K_{\nabla X_0}$, positive constants, such that for all $h \in]0,1]$, we have

$$\begin{aligned} f_x\left(X(t) + \lambda\epsilon_X(t), \theta + \lambda h\widehat{\theta}\right) &\le K_{f_x} \\ f_\theta\left(X(t) + \lambda\epsilon_X(t), \theta + \lambda h\widehat{\theta}\right) &\le K_{f_\theta} \\ \nabla X_0(\theta + \lambda h\widehat{\theta}) &\le K_{\nabla X_0}. \end{aligned} \quad (33)$$

Therefore

$$|\epsilon_X(t)| \le K_{f_x}\int_0^t |\epsilon_x(t)|\,dt + K_{f_\theta}Th|\widehat{\theta}| + K_{\nabla X_0}h|\widehat{\theta}| \quad (34)$$

and, thus, by Gronwall's lemma, we have

$$|\epsilon_X(t)| \le \left(K_{f_\theta}T + K_{\nabla X_0}\right)h|\widehat{\theta}|\exp\left(K_{f_x}t\right) \quad (35)$$

which proves that ϵ_X uniformly converges to 0 when h tends to 0 and therefore that the state variable X is continuous with respect to the control variable θ. Furthermore, let us note that ϵ_X/h is bounded: $\exists K_1 > 0$ such that $\forall t \in [0,T]$, $|\epsilon_X(t)|/h < K_1$.

We now need to obtain a result on the convergence of ϵ_X/h towards $\widehat{X}$. For this purpose, using (32), we have

$$\begin{aligned} \left|\frac{\epsilon_X(t)}{h} - \widehat{X}(t)\right| \le{}& \int_0^t K_{f_x}\left|\frac{\epsilon_X(t)}{h} - \widehat{X}(t)\right|dt \\ &+ \int_0^t\int_0^1 \left(f_x\left(X_\lambda(t),\theta_\lambda\right) - f_x\left(X(t),\theta\right)\right) \\ &\qquad \times \frac{\epsilon_X(t)}{h}d\lambda dt \\ &+ \int_0^t\int_0^1 \left(f_\theta\left(X_\lambda(t),\theta_\lambda\right) - f_\theta\left(X(t),\theta\right)\right) \\ &\qquad \times \widehat{\theta}d\lambda dt \\ &+ \int_0^1 \left(\nabla X_0(\theta_\lambda) - \nabla X_0(\theta)\right)\widehat{\theta}d\lambda \end{aligned} \quad (36)$$

in which $X_\lambda(t) = X(t) + \lambda\epsilon_X(t)$ and $\theta_\lambda = \theta + \lambda h\widehat{\theta}$. Because of the bound on ϵ_X obtained in (35) and because of the bounds on f_x, f_θ, and ∇X_0 obtained in (33), the three last integrands in (36) are bounded. Furthermore, since f_x, f_θ, and ∇X_0 are continuous, these three integrands converge to 0 almost everywhere when $h \to 0$. By the theorem of dominated convergence, we have

$$\left|\frac{\epsilon_X(t)}{h} - \widehat{X}(t)\right| \le \int_0^t K_{f_x}\left|\frac{\epsilon_X(t)}{h} - \widehat{X}(t)\right|dt + o(1). \quad (37)$$

Using Gronwall's lemma, we deduce that

$$\sup_{t\in[0,T]}\left\{\left|\frac{\epsilon_X(t)}{h} - \widehat{X}(t)\right|\right\} \underset{h\to 0}{\longrightarrow} 0. \quad (38)$$

Finally, posing $\phi(X(T)) = \|g(X(T)) - Y^{\text{obs}}\|_2^2$, the calculus of variations on J leads to

$$\begin{aligned}\frac{\widetilde{J}-J}{h} &- 2\left(g\left(X(T)\right)-Y^{\text{obs}}\right)^{\top}\nabla g\left(X(T)\right)\widehat{X}(T)\\ &= \int_0^1 \phi_x\left(X_\lambda(T)\right)\left(\frac{\epsilon_X(T)}{h}-\widehat{X}(T)\right)d\lambda\\ &\quad+\int_0^1\left(\phi_x\left(X_\lambda(T)\right)-\phi_x\left(X(T)\right)\right)\widehat{X}(T)d\lambda. \qquad (39)\end{aligned}$$

Since ϕ is continuously differentiable, the two integrands are bounded and converge to zero almost everywhere when $h \to 0$. By the theorem of dominated convergence, the proof is complete. ∎

Appendix B
Adjoint Equations for PCP PDE Model

Let us explicit the dependence in space of the cost function by rewriting $J = \|g(X(T)) - Y^{\text{obs}}\|$ as $J = \|\int_\Omega k(X(s,T),s)ds - Y^{\text{obs}}\|^2$, in which Ω is the computational domain. Furthermore, let us specify the boundary conditions for the PDE (13) as follows:

$$\mu(\theta)\partial_s X(t,s)\cdot n = P\left(s,X(t,s),\theta\right), \quad \forall s\in\partial\Omega \qquad (40)$$

in which n is the unit vector normal to the boundary $\partial\Omega$. The regularity of the PCP PDEs provides us with enough technical conditions to compute the derivative of J as follows:

$$\begin{aligned}\lim_{h\to 0}\frac{J(\theta+h\widehat{\theta})-J(\theta)}{h} =& 2\left(\int_\Omega k\left(X(s,T),s\right)ds - Y^{\text{obs}}\right)^{\top}\\ &\times\int_\Omega k_x\left(X(s,T),s\right)\widehat{X}(s,T)ds\end{aligned} \qquad (41)$$

in which $\widehat{X}$ is the solution of the following linear PDE:

$$\begin{aligned}\frac{\partial\widehat{X}}{\partial t}(t,s) &= P_x\left(s,X(t,s),\theta\right)\widehat{X}(t,s)+\mu(\theta)\Delta\widehat{X}(t,s)\\ &\quad+\left(P_\theta\left(s,X(t,s),\theta\right)+\Delta X(t,s)\nabla\mu(\theta)\right)\widehat{\theta} \qquad (42)\end{aligned}$$

with linear boundary conditions

$$\begin{aligned}&\nabla\mu(\theta)\widehat{\theta}\partial_s X(t,s)\cdot n+\mu(\theta)\partial_s\widehat{X}(t,s)\cdot n\\ &= C\left(P_x\left(s,X(t,s),\theta\right)\widehat{X}(t,s)+P_\theta\left(s,X(t,s),\theta\right)\widehat{\theta}\right). \qquad (43)\end{aligned}$$

Taking the inner product of this linear PDE with an arbitrary costate p yields

$$\begin{aligned}\int_\Omega\int_0^T p^{\top}\frac{\partial\widehat{X}}{\partial t}(t,s) &= \int_\Omega\int_0^T p^{\top}\left(P_x\widehat{X}+\mu(\theta)\Delta\widehat{X}(t,s)\right)\\ &\quad+\int_\Omega\int_0^T p^{\top}\left(P_\theta+\Delta X\nabla\mu(\theta)\right)\widehat{\theta}. \qquad (44)\end{aligned}$$

Integrating by parts, we obtain

$$\begin{aligned}\int_\Omega p^{\top}\widehat{X}(T) &= \int_\Omega\int_0^T\widehat{X}^{\top}\left(\frac{\partial p}{\partial t}+P_x^{\top}p+\mu(\theta)\Delta p\right)\\ &\quad+\int_\Omega\int_0^T p^{\top}\left(P_\theta+\Delta X\nabla\mu(\theta)\right)\widehat{\theta}\\ &\quad+\int_{\partial\Omega}\int_0^T\widehat{X}^{\top}\left(P_x^{\top}p-\mu(\theta)\partial_s p\cdot n\right)\\ &\quad+\int_{\partial\Omega}\int_0^T p^{\top}\left(P_\theta-\partial_s X\cdot n\nabla\mu(\theta)\right)\widehat{\theta}. \qquad (45)\end{aligned}$$

We are now in a position to extract the gradient of J. Provided that p solves the following linear PDE:

$$-\frac{\partial p}{\partial t} = P_x^{\top}p+\mu(\theta)\Delta p, \qquad (46)$$

with boundary conditions

$$\mu(\theta)\partial_s p\cdot n = P_x^{\top}p \qquad (47)$$

and terminal condition

$$p(s,T) = 2\left(\int_\Omega k\left(X(s,T),s\right)-Y^{\text{obs}}\right)^{\top}k_x\left(X(s,T),s\right) \qquad (48)$$

the gradient is

$$\begin{aligned}\nabla J &= \int_\Omega\int_0^T p^{\top}\left(P_\theta+\Delta X\nabla\mu(\theta)\right)\\ &\quad+\int_{\partial\Omega}\int_0^T p^{\top}\left(P_\theta-\partial_s X\cdot n\nabla\mu(\theta)\right). \qquad (49)\end{aligned}$$

References

[1] P. N. Adler, "Planar signaling and morphogenesis in *drosophila*," *Develop. Cell*, vol. 2, no. 5, pp. 525–535, 2002.

[2] K. Amonlirdviman, "Mathematical modeling of planar cell polarity signaling in the Drosophila Melanogaster wing," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Stanford Univ., Palo Alto, CA, Aug. 2005.

[3] K. Amonlirdviman, N. A. Khare, D. R. P. Tree, W.-S. Chen, J. D. Axelrod, and C. J. Tomlin, "Mathematical modeling of planar cell polarity to understand domineering nonautonomy," *Science*, vol. 307, no. 5708, pp. 423–426, Jan. 2005.

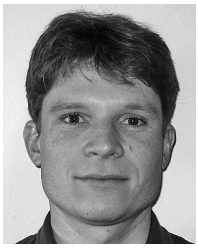
[4] J. D. Axelrod, "Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling," *Genes & Development*, vol. 15, no. 10, pp. 1182–1187, May 2001.

[5] A. Bensoussan, *Perturbation Methods in Optimal Control*. New York: Wiley/Gauthier-Villars, 1988, Series in Modern Applied Mathematics.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ., 2004.

[7] G. Evensen, D. P. Dee, and J. Schröter, "Parameter estimation in dynamical models," in *Ocean Modeling and Parameterizations*, E. P. Chassignet and J. Verron, Eds. Dordrecht, The Netherlands: Kluwer, 1998.

[8] P. E. Gill and M. W. Leonard, "Reduced-Hessian quasi-Newton methods for unconstrained optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 209–237, 2001.
[9] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. New York: Academic/Harcourt Brace, 1999.
[10] O. A. Igoshin, A. Goldbeter, D. Kaiser, and G. Oster, "A biochemical oscillator explains several aspects of *Myxococcus xanthus* behavior during development," *Proc Natl Acad Sci USA*, vol. 101, pp. 15760–15775, 2004.
[11] D. L. Ippolito, H. Shankaran, H. Resat, G. J. Newton, W. B. Chrisler, L. K. Opresko, and H. S. Wiley, "Analysis and modeling of the rapid ERK oscillations that occur in response to EGF receptor activation," in *Proc. Amer. Soc. Cell Biology 46th Annu. Meet. Computat. Applicat. Cell Biol.*, Dec. 2006.
[12] A. Jameson, "Aerodynamic design via control theory," *J. Sci. Comput.*, vol. 3, pp. 233–260, 1988.
[13] L. Ljung, *System Identication: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.
[14] D. Ma, C. Yang, H. McNeill, M. A. Simon, and J. D. Axelrod, "Fidelity in planar cell polarity signalling," *Nature*, vol. 421, pp. 543–547, 2003.
[15] B. Marcos and G. Payre, "Parameters estimation of an aquatic biological system by the adjoint method," *Math. Comput. Simulation*, vol. 30, no. 5, pp. 405–418, Nov. 1988.
[16] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
[17] R. L. Raffard, "Optimal control of systems governed by differential equations with applications in air traffic management and systems biology," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, Dec. 2006.
[18] E. D. Sontag, *Neural Networks for Control, in Essays on Control: Perspectives in the Theory and Its Applications*, H. L. Trentelman and J. C. Willems, Eds. Boston, MA: Birkhäuser, 1993, pp. 339–380.
[19] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 2nd ed. New York: Springer, 1998.
[20] D. I. Strutt, "The asymmetric subcellular localisation of components of the planar polarity pathway," *Seminars in Cell Developm. Biol.*, vol. 13, no. 3, pp. 225–231, 2002.
[21] D. R. Tree, J. M. Shulman, R. Rousset, M. P. Scott, D. Gubb, and J. D. Axelrod, "Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling," *Cell*, vol. 109, no. 3, pp. 371–381, May 2002.
[22] A.-A. M. Ussif, L. K. Sandal, and S. I. Steinshamn, "Estimation of biological and economic parameters of a bioeconomic sheries model using dynamical data assimilation," *J. Bioeconom.*, vol. 4, no. 1, pp. 39–48, Jan. 2002.
[23] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, "The segment polarity network is a robust developmental module," *Genes & Development*, vol. 406, pp. 188–292, 2000.

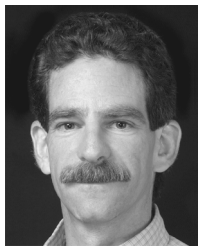


Robin L. Raffard received the M.Sc. degree in mechanical engineering from Ecole Centrale Paris, Paris, France, in 2002, and the Ph.D. degree in aeronautics and astronautics from Stanford University, Stanford, CA, in 2006.

He is a Research Associate with Fixed Income Structuring, Barclays Global Investors. His research interests are in control theory, specifically in stochastic control and in optimal control of differential equations.

Keith Amonlirdviman received the S.B. degree from the Massachusetts Institute of Technology, Cambridge, in 1999, and the Ph.D. degree from Stanford University, Stanford, CA, in 2005, both in aeronautics and astronautics.

He is currently with McKinsey and Company, Montreal, QC, Canada. At Stanford, his research focused on the mathematical modeling of biological cell networks. Prior to beginning his doctoral work, he was a co-founder of Epigraph, Inc., Guitar.com, Inc., and ePrairie.com, Inc., now Midwest Business, Inc.



Jeffrey D. Axelrod received the M.D. and Ph.D. degrees from Washington University.

He is currently an Associate Professor of Pathology with Stanford University School of Medicine, Stanford, CA. He did residency training at Brigham and Women's Hospital and held a postdoctoral fellowship with the Department of Genetics, Harvard Medical School, before joining the Stanford faculty in 1998. His laboratory studies problems related to morphogenesis during development, using genetic analyses in *Drosophila melanogaster* as a mainstay approach.

Claire J. Tomlin received the B.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1992, the M.Sc. degree from Imperial College, London, U.K., in 1993, and the Ph.D. degree from the University of California, Berkeley, in 1998, all in electrical engineering.

She is a Professor of Electrical Engineering and Computer Sciences with the University of California, Berkeley, and holds a joint appointment as a Professor with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA. Her research interests are in control systems, specifically hybrid control theory, and she works on air traffic control automation, flight-management system analysis and design, and modeling and analysis of biological cell networks. She joined Stanford in September 1998 as a Terman Assistant Professor. In July 2005, she joined the University of California, Berkeley, as an Associate Professor. She has held visiting research positions at NASA Ames and Honeywell Labs. During 2003–2006, she was a Visiting Associate, part time, in Control and Dynamical Systems at the California Institute of Technology.

Dr. Tomlin was a recipient of the Chancellor's Professorship at Berkeley (2007), the MacArthur Fellowship (2006), the Okawa Foundation Award (2006), the Eckman Award of the American Automatic Control Council (2003), MIT Technology Review's Top 100 Young Innovators Award (2003), the AIAA Outstanding Teacher Award (2001), the National Science Foundation Career Award (1999), and the Bernard Friedman Memorial Prize in Applied Mathematics (1998).