MATHEMATICAL NOTES "WORKFLOW DESCRIPTION TO DYNAMICALLY MODEL β-ARRESTIN SIGNALING NETWORKS"

Contents

1. Network definition

A biochemical reaction network is defined by

- **Species**: A finite set $\mathcal{S} = \{S_1, \cdots, S_d\}$ of $d \geq 1$ species.
- **Reaction**: A finite set $\mathcal{R} = \{R^1, \cdots, R^n\}$ of $n \geq 1$ relation between linear combination of species, that is

$$R_i : \sum_{j=1}^{d} \alpha_j^i S_j \to \sum_{j=1}^{d} \beta_j^i S_j \,.$$

where α_j^i, β_j^i are non-negative integers and called respectively reactant and product stoichiometric coefficients for reaction i and species j. The set of reactions defines a directed graph between linear combination of species (called "complexes" in the reaction network community).

2. Deterministic mass-action model

A **deterministic** mass-action dynamic model associated to the biochemical reaction network $(\mathcal{S}, \mathcal{R})$ is given by the set of ordinary differential equations

$$\frac{dx_j}{dt} = \sum_{i=1}^{n} \left(\beta_j^i - \alpha_j^i\right) k_i \prod_{l=1}^{d} x_l^{\alpha_l^i}\,, \quad j = 1\,, 2\,, \cdots\,, d\,, \tag{1}$$

and a set initial conditions $x_j(0) \in \mathbb{R}_+$, $j = 1..d$. The law of mass-action assumes that the rate of reaction R^i is proportional to the concentration of its reactants to the power of their stoichiometry. The constant k_i is called the rate constant of reaction R^i.

Date: October 2, 2017.

3. Stochastic mass-action model

A **stochastic** mass-action dynamic model associated to the biochemical reaction network $(\mathcal{S}, \mathcal{R})$ is given by a continuous-time Markov chain, with transitions

$$(X_1\,, X_2\,, \cdots, X_d) \to \left(X_1 + \beta_1^i - \alpha_1^i\,, X_2 + \beta_2^i - \alpha_2^i\,, \cdots, X_d + \beta_d^i - \alpha_d^i\right), i = 1, \cdots, n\,, \quad (2)$$

occurring at rates $\kappa_i \prod_{j=1}^{d} \binom{X_j}{\alpha_j^i}$, and with initial conditions $X_j(0) \in \mathbb{N}_+$, $j = 1..d$. The stochastic law of mass-action assumes that the rate of reaction R^i is proportional to the number of possible combination of its reactants. The constant k_i is called the rate constant of reaction R^i.

4. Observation model

An obervation model is a function h that links the kinetic variable x to the data y, with possible additional measurement parameters and taken into account some noise term ε

$$y = h(x, \varepsilon)\,. \quad (3)$$

Additive noise model can be re-written as

$$y = h(x) + \varepsilon\,, \quad (4)$$

while multiplicative noise model can be re-written as

$$y = h(x)\varepsilon\,. \quad (5)$$

5. Modeling β-arrestin recruitment kinetics at the FSH receptor

5.1. **Topological model.** The species list is given by

$$\mathcal{S} = \{FSH, FSHR, FSHFSHR, \beta arrestin, FSHFSHR\beta arrestin\}\,. \quad (6)$$

The reaction list together with its rate constants are given by

- $FSH + FSHR \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} FSHFSHR$
- $\beta arrestin + FSHFSHR \xrightarrow{k^+} FSHFSHR\beta arrestin$
- $FSHFSHR\beta arrestin \xrightarrow{k^-} \emptyset$

5.2. **Dynamic model and experimental conditions.** After taking care of each experimental conditions (see main text), we obtain the full dynamic model that includes 7 experiments (state variables are indexed by a label $i = 1, \cdots, 7$ for each different doses) together with the

observation model:

$$
\begin{aligned}
\frac{d}{dt}FSH_i &= -k_{on}FSH_i \cdot FSHR_i + k_{off}FSHFSHR_i\,, && t \geq 0\\
\frac{d}{dt}FSHR_i &= -k_{on}FSH_i \cdot FSHR_i + k_{off}FSHFSHR_i\,, && t \geq 0\\
\frac{d}{dt}FSHFSHR_i &= k_{on}FSH_i \cdot FSHR_i - k_{off}FSHFSHR_i - k^{+}\beta arrestin_i \cdot FSHFSHR_i\,, && t \geq 0\\
\frac{d}{dt}\beta arrestin_i &= -k^{+}\beta arrestin_i \cdot FSHFSHR_i\,, && t \geq 0\\
\frac{d}{dt}FSHFSHR\beta arrestin_i &= k^{+}\beta arrestin_i \cdot FSHFSHR_i - k^{-}FSHFSHR\beta arrestin_i\,, && t \geq 0\\
FSH_i(t=0) &= Init\_FSH_i\,,\\
FSHR_i(t=0) &= Init\_FSHR\,,\\
FSHFSHR_i(t=0) &= 0\,,\\
\beta arrestin_i(t=0) &= Init\_\beta arrestin\,,\\
FSHFSHR\beta arrestin_i(t=0) &= 0\,,\\
Induced\_BRET_i(t) &= kf \cdot FSHFSHR\beta arrestin_i(t) + \varepsilon_t\,, && t \geq 0
\end{aligned}
\tag{7}
$$

where ε_t is a collection of independent Gaussian random variable of zero mean and variance σ^2, and the initial quantities for FSH are given by

$$
(Init\_FSH_i)_{i=1,\cdots,7} = (0.0128, 0.064, 0.32, 1.6, 8, 40, 200)\, 10^{-9} M\,. \tag{8}
$$

5.3. **Parameterization.** To improve parameter optimization, we use an adimentionalized system and parameters, namely we define

$$
\begin{aligned}
x_1^i(t) &= \frac{FSH_i(t)}{Init\_FSHR}, && t \geq 0 \\
x_2^i(t) &= \frac{FSHR_i(t)}{Init\_FSHR}, && t \geq 0 \\
x_3^i(t) &= \frac{FSHFSHR_i(t)}{Init\_FSHR}, && t \geq 0 \\
x_4^i(t) &= \frac{\beta arrestin_i(t)}{Init\_FSHR}, && t \geq 0 \\
x_5^i(t) &= \frac{FSHFSHR\beta arrestin_i(t)}{Init\_FSHR}, && t \geq 0 \\
\overline{k}_{on} &= \frac{k_{on} Init\_FSHR}{k^-}, \\
\overline{k}_{off} &= \frac{k_{off}}{k^-}, \\
\overline{k}^+ &= \frac{k^+ Init\_FSHR}{k^-}, \\
Init\_x_1^i &= \frac{Init\_FSH_i}{Init\_FSHR}, \\
Init\_x_4 &= \frac{Init\_\beta arrestin}{Init\_FSHR}, \\
\overline{kf} &= kf \cdot Init\_FSHR, && t \geq 0
\end{aligned}
\tag{9}
$$

and we replace the system (7) by

$$
\begin{aligned}
\frac{d}{dt}x_1^i &= -\overline{k}_{on}k^- x_1^i \cdot x_2^i + \overline{k}_{off}k^- x_3^i\,, && t \geq 0\\
\frac{d}{dt}x_2^i &= -\overline{k}_{on}k^- x_1^i \cdot x_2^i + \overline{k}_{off}k^- x_3^i\,, && t \geq 0\\
\frac{d}{dt}x_3^i &= \overline{k}_{on}k^- x_1^i \cdot x_2^i - \overline{k}_{off}k^- x_3^i - \overline{k}^+ k^- x_4^i \cdot x_3^i\,, && t \geq 0\\
\frac{d}{dt}x_4^i &= -\overline{k}^+ k^- x_4^i \cdot x_3^i\,, && t \geq 0\\
\frac{d}{dt}x_5^i &= \overline{k}^+ k^- x_4^i \cdot x_3^i - k^- x_5^i\,, && t \geq 0\\
x_1^i(t=0) &= Init\_x_1^i\,, && (10)\\
x_2^i(t=0) &= 1\,,\\
x_3^i(t=0) &= 0\,,\\
x_4^i(t=0) &= Init\_x_4\,,\\
x_5^i(t=0) &= 0\,,\\
Induced\_BRET_i(t) &= \overline{kf} \cdot x_5^i(t) + \varepsilon_t\,, && t \geq 0
\end{aligned}
$$

Thus, to perform data fitting, we actually simulate the system (10). For the optimization algorithm, we further transform the parameter in log scale, so that we search for the parameter set

$$\theta = \left(\log_{10}(\overline{k}_{on}), \log_{10}(\overline{k}_{off}), \log_{10}(\overline{k}^+), \log_{10}(k^-), \log_{10}(Init\_FSHR), \log_{10}(Init\_x_4), \log_{10}(\overline{kf}), \log_{10}(\sigma)\right) \tag{11}$$

5.4. **Parameter ranges for numerical optimization.** To start multi-run deterministic optimizations (using D2D), we used uniform distributions with following ranges to search for the

optimal parameter sets (**in** $\log_{10}$ **scale**):

Parameter	Min	Max
$\overline{k}_{on}$	-10	10
$\overline{k}_{off}$	-10	10
$\overline{k}^{+}$	-10	10
k^{-}	-10	10
$Init\_FSHR$	-20	0
$Init\_x_4$	-10	10
$\overline{kf}$	-10	10
σ	-4	1

(12)

5.5. **Objective function.** To search for best parameter sets, we minimize the $-2log(L)$, where L is the likelihood function according to the error measurement model. Thus, we are lead to minimize

$$J(\theta) = \sum_{k=1}^{n_t} \sum_{i=1}^{7} \log(2\pi\sigma^2) + \left(\frac{d_{i,k} - Induced\_BRET_i(t_k)}{\sigma} \right)^2 , \tag{13}$$

where we denoted by $d_{i,k}$ the measured induced BRET signal for experiment i at time t_k.

6. Example of over parameterization

We show in a slightly different model a concrete case of over parameterization. The following model is an example of catalysis signaling cascade, where the complex Ligand-Receptor catalyse the activation of a downstream molecule (like cAMP for instance). We take as species list

$$\mathcal{S} = \{x_1, x_2, x_3, x_4, x_5\} . \tag{14}$$

The reaction list together with its rate constants are given by

- $x_1 + x_2 \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} x_3$
- $x_3 + x_4 \xrightarrow{k^+} x_3 + x_5$
- $x_5 \xrightarrow{k^-} \emptyset$

Suppose as in the BRET measurement technique that one observes the activated molecule x_5, up to an unknown constant kf. Thus, the dynamic model together with the observation model

is given by (note the close similarity with the system (7)):

$$
\begin{aligned}
\frac{d}{dt}x_1 &= -k_{on}x_1 \cdot x_2 + k_{off}x_3\,, \quad t \geq 0 \\
\frac{d}{dt}x_2 &= -k_{on}x_1 \cdot x_2 + k_{off}x_3\,, \quad t \geq 0 \\
\frac{d}{dt}x_3 &= k_{on}x_1 \cdot x_2 - k_{off}x_3\,, \quad t \geq 0 \\
\frac{d}{dt}x_4 &= -k^{+}x_4 \cdot x_3\,, \quad t \geq 0 \\
\frac{d}{dt}x_5 &= k^{+}x_4 \cdot x_3 - k^{-}x_5\,, \quad t \geq 0 \\
x_1(t=0) &= Init\_x_1\,, \\
x_2(t=0) &= Init\_x_2\,, \\
x_3(t=0) &= 0\,, \\
x_4(t=0) &= Init\_x_4\,, \\
x_5(t=0) &= 0\,, \\
Induced\_BRET(t) &= kf \cdot x_5(t) + \varepsilon_t\,, \quad t \geq 0\,.
\end{aligned}
\tag{15}
$$

In the system (15), a structural identifiability analysis reveals that the parameters $(Init\_x_4, kf)$ are not identifiable, and actually only the product $Init\_x_4 \cdot kf$ can be identified. In this example, this fact can actually be shown doing an adimentionalization step, normalizing the variables x_4 and x_5 by $Init\_x_4$.