Parameter estimation of ODEs using Support Vector Regression and qualitative constraints

Paola Bouchet

Joint work with Nicolas Brunel and Florence d'Alché-Buc
AMIS-Bio Team, IBISC Laboratory - FRE CNRS 2873
Université d'Evry Val d'Essonne, France

September 18 2007 - PIMMS'07, Glasgow

Ordinary Differential Equations in Systems Biology

Motivation

- Numerous quantitative models for the description of biological systems (metabolic and signaling pathway, regulatory networks) based on ODEs e.g. Repressilator, JAK/STAT pathway, cell cycle models ...
- Dynamical systems theory gives tools for understanding the behavior of biological networks.

Ordinary Differential Equations in Systems Biology

Motivation

- Numerous quantitative models for the description of biological systems (metabolic and signaling pathway, regulatory networks) based on ODEs e.g. Repressilator, JAK/STAT pathway, cell cycle models ...
- Dynamical systems theory gives tools for understanding the behavior of biological networks.

Objective

- Network of interactions is known as the functional form of the equations.
- Our contribution concerns the final step in reverse engineering: **the estimation of phenomenological parameters**.
  Difficult task because too few observations and hard optimization problem
  $\Rightarrow$ Need to incorporate additional knowledge = qualitative information about the system.

Model and assumptions

Model and assumptions

$$\begin{cases} y_i = x_\theta(t_i) + \epsilon(t_i) \\ \dot{x}(t) = f(x, t, \theta) \end{cases} \tag{1}$$

where $y(t_i)$ are observed data points, $x_\theta(\cdot)$ solution of the ODE for parameter θ and $\epsilon(t_i)$ the measurement noise (i.i.d.), $i = 1 \dots n$.

- Functional expression of f is known, derived from biochemical kinetics.
- Initial conditions $x_\theta(0) = x_0$ are unknown.
- All the system is observed.
- True parameter is θ^*, and $x_{\theta^*} = x^*$.

General principle of the two-step estimator

The two steps

1. Estimation of the solution x^* by $\hat{x}_n$ (nonparametric regression of y_i on time t_i).
2. Differentiation of $\hat{x}_n$ and minimization over θ of the distance $\|\dot{\hat{x}}_n - f(\hat{x}_n(t), \theta)\|_{L^2}$.

Consistency of the method relies on the consistency of $\hat{x}_n$.

This work is about the influence of estimator $\hat{x}_n$ in the first step

- Use of **Support Vector Regression** instead of LS Splines (N. Brunel, PESB'07).
- Construction of better estimates $\hat{x}_n$ with qualitative knowledge.

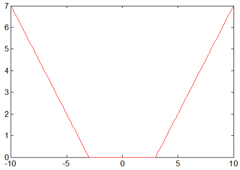
Support Vector Regression

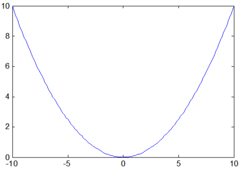
Definition via optimization problem

$$\hat{x}_n = \arg\min_{x \in \mathcal{H}} \|x\|_{\mathcal{H}} + C \sum_{i=1}^{n} L(y_i - x(t_i))$$

where L is a loss function, $\|\|_{\mathcal{H}}$ is a RKHS norm (with kernel k) and C a regularization parameter.

- Huge family of nonparametric estimators (choices of splines, gaussian kernels ...), and of their robust counterpart with adapted loss function L.





The ϵ–SVR (Vapnik, 1995)

Attractive Optimization Program for ϵ–SVR

- L is the ϵ–insensitive loss function.
- The form of the function $\hat{x}_n$ is

$$\hat{x}_n(t) = \langle w, \Phi(t) \rangle + b \quad (2)$$

$$= \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(t_i, t) + b \quad (3)$$

where $w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \Phi(t_i)$.

- The accuracy parameter ϵ, the regularization constant C and kernel parameters are chosen by Generalized Cross-Validation GCV (Craven and Wahba, 1978).

Difference between SVR and LS Splines Regression

Numerical comparison

- Use of gaussian kernel for SVR and B–Splines.
- Monte-Carlo study (1000 independent drawings) based on Lotka Volterra equations (prey-predator interactions)

$$\frac{dx}{dt} = ax - bxy \qquad \text{(prey)}$$
$$\frac{dy}{dt} = -cy + dxy \qquad \text{(predator)}$$

	Mean $(\hat{a},\hat{b},\hat{c},\hat{d})$		Standard Deviation	
Obs.	Spline	SVR	Spline	SVR
50	(0.7, 1.22, 1.35, 1.81)	(0.98, 1.32, 1.20, 1.45)	(0.29, 0.31, 0.39, 0.48)	(0.34, 0.47, 0.36, 0.53)
100	(0.73, 1.28, 1.50, 1.99)	(1.06, 1.44, 1.35, 1.69)	(0.20, 0.21, 0.27, 0.35)	(0.29, 0.38, 0.30, 0.47)
200	(0.93, 1.46, 1.41, 1.92)	(1.11, 1.51, 1.42, 1.79)	(0.17, 0.18, 0.20, 0.26)	(0.24, 0.32, 0.23, 0.37)
1000	(0.91, 1.41, 1.43, 1.91)	(1.13, 1.54, 1.50, 1.89)	(0.08, 0.08, 0.10, 0.14)	(0.15, 0.17, 0.12, 0.20)

True values are $(a, b, c, d) = (1, 1.5, 1.5, 2)$.

Use of L^2 or L^1–norm for parameter estimation (2^{nd} step)

Two minimization procedures

$$\hat{\theta}_n^2 = \arg\min_\theta \|\dot{\hat{x}}_n - f(\hat{x}_n(t), \theta)\|_{L^2}$$

vs

$$\hat{\theta}_n^1 = \arg\min_\theta \|\dot{\hat{x}}_n - f(\hat{x}_n(t), \theta)\|_{L^1}$$

Numerical Comparison

	Mean $(\hat{a}, \hat{b}, \hat{c}, \hat{d})$		Standard Deviation	
Obs.	$\hat{\theta}_n^1$	$\hat{\theta}_n^2$	$\hat{\theta}_n^1$	$\hat{\theta}_n^2$
100	(1.03, 1.42, 1.38, 1.76)	(1.05, 1.44, 1.36, 1.70)	(0.29, 0.38, 0.33, 0.47)	(0.28, 0.38, 0.30, 0.47)
200	(1.07, 1.48, 1.44, 1.85)	(1.11, 1.52, 1.43, 1.81)	(0.23, 0.32, 0.24, 0.33)	(0.26, 0.35, 0.24, 0.36)
1000	(1.07, 1.54, 1.46, 1.92)	(1.06, 1.53, 1.48, 1.94)	(0.10, 0.14, 0.13, 0.18)	(0.12, 0.12, 0.11, 0.16)

True values are $(a, b, c, d) = (1, 1.5, 1.5, 2)$.

Incorporation of qualitative knowledge on the nature of dynamics

Incorporation of qualitative knowledge on the dynamics

Motivation

- High number of parameters to estimate in comparison with the relatively small number of noisy observations.
- Dynamical systems tools allow us to know something else about the system.
  $\Rightarrow$ Incorporation of prior qualitative knowledge on the dynamics.

Incorporation of qualitative knowledge on the dynamics

Motivation

- High number of parameters to estimate in comparison with the relatively small number of noisy observations.
- Dynamical systems tools allow us to know something else about the system.
  $\Rightarrow$ Incorporation of prior qualitative knowledge on the dynamics.

Two main approaches

1. Dynamical constraints obtained from bifurcation diagram or stability analysis $\Rightarrow$ Complex and difficult task with general nonlinear systems: hard to incorporate into the learning algorithm.

Incorporation of qualitative knowledge on the dynamics

Motivation

- High number of parameters to estimate in comparison with the relatively small number of noisy observations.
- Dynamical systems tools allow us to know something else about the system.
  $\Rightarrow$ Incorporation of prior qualitative knowledge on the dynamics.

Two main approaches

1. Dynamical constraints obtained from bifurcation diagram or stability analysis $\Rightarrow$ Complex and difficult task with general nonlinear systems: hard to incorporate into the learning algorithm.
2. **Shape constraints** $\Rightarrow$ Decomposition of the solution $x_\theta = \text{shape} + \text{noise}$, where shape is the asymptotic, stable or ideal response and noise is a transition or unmodeled part.

Incorporation of qualitative knowledge

- The decomposition 'shape + noise' can be incorporated in the SVR estimation step.
- Introduction of a vector space spanned by functions $\left(\psi_j(t,\eta)\right)_{j=1..m}$ representing the shape.
- The functional to be minimized stays the same

$$\arg\min_{x\in\mathcal{H}'} \|x\|_{\mathcal{H}'} + C\sum_{i=1}^{n} L(y_i - x(t_i))$$

- The new expression of $\hat{x}_n$ is

$$\hat{x}_n(t) = \underbrace{\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(t_i,t)}_{noise} + \underbrace{\sum_{j=1}^{m}\beta_j\psi_j(t,\eta)}_{shape}$$

- Additionally, the shape parameters η have to be estimated.

Incorporation of qualitative knowledge

The semiparametric SVR (Smola *et al.*, 1999)

- For fixed η, the optimization problem is solved analogously to SVR.
- The corresponding dual optimization problem is

$$\begin{array}{l} \max \; -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)k(t_i,t_j) \\ -\,\epsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{n}y_i(\alpha_i-\alpha_i^*) \\ \text{such that} \begin{cases} \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\psi_j(t_i,\eta)=0, \forall 1\le j\le m \\ \alpha_i,\alpha_i^*\in[0,C] \end{cases} \end{array} \tag{4}$$

Example of qualitative knowledge: oscillating behavior

- Many oscillating biological systems that can be explained by convergence to a limit cycle (periodical solution).
- Fourier decomposition of a periodical signal f

$$f(t) = C_0 + \sum_{k=1}^{\infty} C_k cos(k\omega t + \phi_k)$$

Example of qualitative knowledge: oscillating behavior

- Many oscillating biological systems that can be explained by convergence to a limit cycle (periodical solution).
- Fourier decomposition of a periodical signal f

$$f(t) = C_0 + \sum_{k=1}^{\infty} C_k cos(k\omega t + \phi_k)$$

- Truncation of Fourier series $\Rightarrow$ our estimator $\hat{x}_n$ for semiparametric SVR becomes

$$\hat{x}_n(t) = C_0 + \sum_{k=1}^{3} C_k \cos(kat + b_k) + E(t)$$

  Hence $\eta = (a, b_1, b_2, b_3)$ and $E = \textit{approximation error} + \textit{transition behavior}$.
- Interesting information : kind of natural period of the signal $T = \frac{2\pi}{a}$.
  $\rightarrow$ Either we use it, either we obtain it.

Example of qualitative knowledge: oscillating behavior

- Many oscillating biological systems that can be explained by convergence to a limit cycle (periodical solution).
- Fourier decomposition of a periodical signal f

$$f(t) = C_0 + \sum_{k=1}^{\infty} C_k cos(k\omega t + \phi_k)$$

- Truncation of Fourier series $\Rightarrow$ our estimator $\hat{x}_n$ for semiparametric SVR becomes

$$\hat{x}_n(t) = C_0 + \sum_{k=1}^{3} C_k \cos(kat + b_k) + E(t)$$

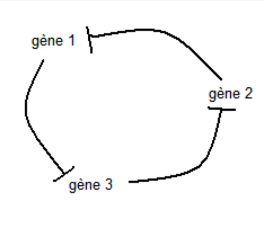
  Hence $\eta = (a, b_1, b_2, b_3)$ and $E = approximation\, error + transition\, behavior$.
- Interesting information : kind of natural period of the signal $T = \frac{2\pi}{a}$.
  $\rightarrow$ Either we use it, either we obtain it.
- Finally, shape parameters η are chosen by GCV and grid search (similar to Ansley and Kohn, 1996 or Steinke and Schölkopf, 2005).

Example: first results on the Repressilator

The Repressilator (Elowitz and Leibler, 2000)

Gene regulatory network

- Synthetic genetic regulatory networks of 3 mutually inhibited genes (via their encoded proteins), designed from scratch to exhibit a stable oscillation.
- Implemented in *Escherichia Coli*, the desired behavior was verified.



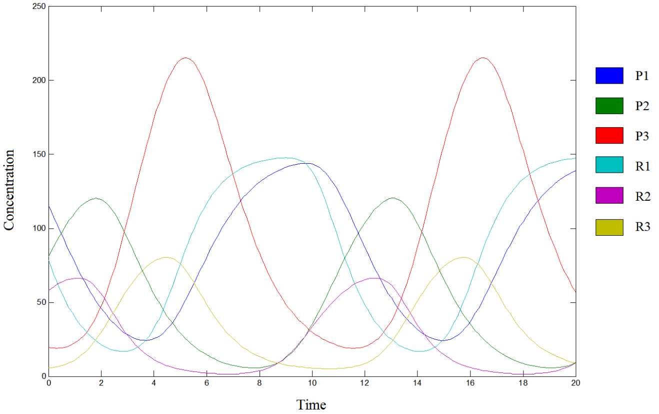
The Repressilator (Elowitz and Leibler, 2000)

Corresponding system of ODEs

$$
\begin{aligned}
\frac{dR_1}{dt} &= \frac{V_{1max} K_{12}^n}{K_{12}^n + P_2^n} - k_{1m} R_1 \\
\frac{dP_1}{dt} &= K_1 R_1 - k_{1p} P_1 \\
\frac{dR_2}{dt} &= \frac{V_{2max} K_{23}^n}{K_{23}^n + P_3^n} - k_{2m} R_2 \\
\frac{dP_2}{dt} &= K_2 R_2 - k_{2p} P_2 \\
\frac{dR_3}{dt} &= \frac{V_{3max} K_{31}^n}{K_{31}^n + P_1^n} - k_{3m} R_3 \\
\frac{dP_3}{dt} &= K_3 R_3 - k_{3p} P_3
\end{aligned}
$$

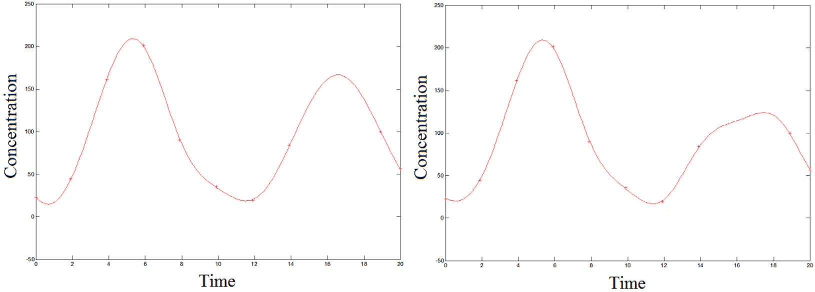
where R_i is the concentration of $mRNA$ transcripted from gene i and P_i is the concentration of proteins translated from R_i.

Theoretical curve



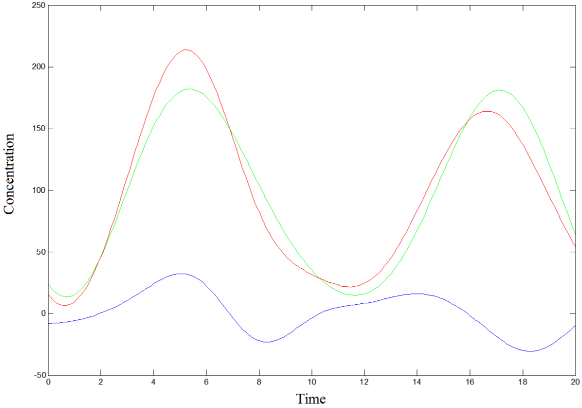
Influence of observations position

Focus on P_3: SVR (continuous line) of 10 noisy observations, with (left) and without (right) qualitative constraints.



Decomposition of $\hat{P}_3$ (Semiparametric SVR)

$$\hat{P}_3(t) = \text{shape} + \text{noise}$$



Nonparametric SVR vs Semiparametric SVR

Parameter	Real value	Estimated value (Nonparametric SVR)	Estimated value (Semiparametric SVR)
V_{1max}	150	139.23	141.05
V_{2max}	80	53.85	62.77
V_{3max}	100	67.80	79.28
K_{12}	50	57.85	54.55
K_{23}	40	49.90	48.46
K_{31}	50	66.87	66.69
K_1	1	0.97	1.04
K_2	2	1.79	1.44
K_3	3	1.88	2.21
k_{1m}	1	0.93	0.93
k_{2m}	1	0.73	0.88
k_{3m}	1	0.88	1.09
k_{1p}	1	0.96	1.03
k_{2p}	1	0.88	0.67
k_{3p}	1	0.64	0.75
n	3	3.25	3.47
Mean Squared Error (curve)	0	1577.80	788.45
Mean Squared Error (parameters)	0	142.65	73.64

Conclusion and perspectives

Conclusion

- Original approach to parameter estimation : use of two-step estimator with SVR and constraints on dynamics.
- First results on a single system: very promising.

Conclusion and perspectives

Conclusion

- Original approach to parameter estimation : use of two-step estimator with SVR and constraints on dynamics.
- First results on a single system: very promising.

Perspectives

- Step to be improved : the choice of hyperparameters (GCV).
  $\longrightarrow$ Model selection via bilevel optimization (Bennett *et al.*, 2006).
- Application to other shape constraint.
  $\longrightarrow$ Convergence to a steady state.