Estimation de réseaux biologiques modélisés par des équations différentielles

N. Brunel

ENSIIE / AMIS-Bio, Laboratoire IBISC, Université d'Evry
Joint work with F. d'Alché-Buc

21 octobre 2009

Outline

Outline

Motivation : dynamical systems with differential equations

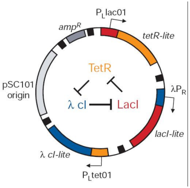
- Deterministic Differential equations (ODEs,DDEs,...): reference models in Systems Biology for dynamical behavior of gene regulation or metabolic networks,...

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

  with $x(t)$ concentrations in the cell, $u(t)$ is an input (control) variable (e.g. u is an extracellular signal).
- Parameter θ describes the network/system (possibly structure network)
- Justification of these models and mathematical expression : chemical kinetics, mechanistic description (mass action laws, Hill kinetics, S-systems).

Example of Gene Regulatory Network : the Repressilator

- Negative loop with 3 genes and 3 proteins with oscillatory behavior, proposed by Elowitz and Leibler (Nature, 2000)



- Nonlinear ODE in 6 dimensions with R_i, P_i, $i = 1, \ldots, 3$ representing the concentration of mRNAs and proteins:

$$\begin{cases} \dot{R}_i(t) & = & V_i^{max} \frac{K_{i,(i+1)}}{K_{i,(i+1)} + P_{(i+1)}^n(t)} - k_{i,r} R_i(t) \\ \dot{P}_i(t) & = & \gamma_i R_i(t) - k_{i,p} P_i(t) \end{cases} \quad \text{for } i = 1, \ldots, 3$$

Outline

1. Motivations from Systems Biology
   - Estimation of differential equations in systems biology
   - Computational challenges for inference of Differential Equations
2. Statistical estimation
   - Two-step estimators
   - Example on Repressilator
   - Asymptotics
3. Ameliorations in step 1: Shape constraints
   - Meaningful decomposition
4. Conclusion

The model and classical estimation

- General model is

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

with initial condition $x(0) = x_0$. f smooth enough for existence and uniqueness of the solution $\phi(\cdot, (x_0, \theta))$ to IVP.

- There exists a true parameter (x_0^*, θ^*) such that the observations are

$$y_i = \phi(t_i, (x_0^*, \theta^*)) + \varepsilon_i$$

where ε_i is a white noise (i.e. $x(t_i), y_i \in \mathbb{R}^d$), i.e. **ALL** the concentration profiles are observed.

- If f is C^m in (x, θ) then ϕ is also C^m $(m \geq 1)$ and if $E(\varepsilon^2) < \infty$: a basic nonlinear regression problem we have to estimate the big parameter $\theta^* = (x_0^*, \theta^*)$ from $(t_i, y_i)_{i=1,\ldots,n}$.

The model and classical estimation

- General model is
$$\dot{x}(t) = f(t, x(t), \theta)$$
with initial condition $x(0) = x_0$. f smooth enough for existence and uniqueness of the solution $\phi(\cdot, (x_0, \theta))$ to IVP.
- There exists a true parameter (x_0^*, θ^*) such that the observations are
$$y_i = \phi\left(t_i, (x_0^*, \theta^*)\right) + \varepsilon_i$$
where ε_i is a white noise (i.e. $x(t_i), y_i \in \mathbb{R}^d$), i.e. **ALL** the concentration profiles are observed.
- If f is C^m in (x, θ) then ϕ is also C^m $(m \geq 1)$ and if $E(\varepsilon^2) < \infty$: a basic nonlinear regression problem we have to estimate the big parameter $\theta^* = (x_0^*, \theta^*)$ from $(t_i, y_i)_{i=1,\ldots,n}$.

Direct Least Squares estimation

$$Q_n(\theta) = \sum_{i=1}^{n} \|y_i - \phi(t_i, \theta)\|_2^2$$
$$\hat{\theta}^{LS} = \arg\min_{\theta \in \Theta \times \mathbb{R}^d} Q_n(\theta)$$

Problems:

1. Numerical integration of ODE is needed (no closed-form for ϕ)
2. No closed-form expression for the derivatives (need to compute the "first variational equations")
3. Difficult reliable optimization: (intrinsicly numerous local minima) and θ is often high-dimensional
4. Difficulty to integrate prior knowledge on the qualitative behavior of the system
5. Difficult problem of identifiability (we can have $\phi(\cdot, \theta) = \phi(\cdot, \theta')$ with $\theta \neq \theta'$)

Well-described in Ramsay et al's paper (JRSS(B), 2007).

Direct Least Squares estimation

$$Q_n(\theta) = \sum_{i=1}^{n} \| y_i - \phi(t_i, \theta) \|_2^2$$
$$\hat{\theta}^{LS} = \arg\min_{\theta \in \Theta \times \mathbb{R}^d} Q_n(\theta)$$

Problems:

1. Numerical integration of ODE is needed (no closed-form for ϕ)
2. No closed-form expression for the derivatives (need to compute the "first variational equations")
3. Difficult reliable optimization: (intrinsicly numerous local minima) and θ is often high-dimensional
4. Difficulty to integrate prior knowledge on the qualitative behavior of the system
5. Difficult problem of identifiability (we can have $\phi(\cdot, \theta) = \phi(\cdot, \theta')$ with $\theta \neq \theta'$)

Well-described in Ramsay et al's paper (JRSS(B), 2007).

Direct Least Squares estimation

$$Q_n(\theta) = \sum_{i=1}^{n} \|y_i - \phi(t_i, \theta)\|_2^2$$
$$\hat{\theta}^{LS} = \arg\min_{\theta \in \Theta \times \mathbb{R}^d} Q_n(\theta)$$

Problems:

1. Numerical integration of ODE is needed (no closed-form for ϕ)
2. No closed-form expression for the derivatives (need to compute the "first variational equations")
3. Difficult reliable optimization: (intrinsicly numerous local minima) and θ is often high-dimensional
4. Difficulty to integrate prior knowledge on the qualitative behavior of the system
5. Difficult problem of identifiability (we can have $\phi(\cdot, \theta) = \phi(\cdot, \theta')$ with $\theta \neq \theta'$)

Well-described in Ramsay et al's paper (JRSS(B), 2007).

Strategies and alternatives

- Computational improvements : better (faster) numerical integrators (e.g sundials software),
- Better global optimization algorithms (Banga et al., 2006),
- Integration scheme based on local linearization + SAEM (Donnet and Samson, 2007)
- Nonparametric Smoothing Technics (Ramsay et al., 2007)

Strategy is to get "simple" estimators:

- Avoid numerous integrations: *replace* ϕ *by a direct estimation of the solution*
- Deal with simpler optimization problems: *control model fitting directly by checking derivatives*

Outline

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\ldots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, Neural Networks, Nadaraya-Watson,...)
   2. estimate the derivative $\dot{\phi}$ with $\hat{\dot{\phi}}_n$ (typically $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\| \hat{\dot{\phi}}_n - f(t, \hat{\phi}_n, \theta) \right\|_{L^2}^2 = \int_0^1 \left\| \hat{\dot{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\dots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, Neural Networks, Nadaraya-Watson,...)
   2. estimate the derivative $\dot{\phi}$ with $\dot{\hat{\phi}}_n$ (typically $\dot{\hat{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\| \dot{\hat{\phi}}_n - f(t, \hat{\phi}_n, \theta) \right\|_{L^2}^2 = \int_0^1 \left\| \dot{\hat{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Two-step estimators

1. Functional estimation from $(t_i, y_i)_{i=1,\ldots,n}$:
   1. estimate $\hat{\phi}_n$ with nonparametric estimators (Splines, Support Vector Regression, Neural Networks, Nadaraya-Watson,…)
   2. estimate the derivative $\dot{\phi}$ with $\hat{\dot{\phi}}_n$ (typically $\hat{\dot{\phi}}_n = \dot{\hat{\phi}}_n$)
2. Minimize the discrepancy between the two estimators of the derivatives measured

$$R_n^2(\theta) = \left\| \hat{\dot{\phi}}_n - f(t, \hat{\phi}_n, \theta) \right\|_{L^2}^2 = \int_0^1 \left\| \hat{\dot{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|_2^2 dt$$

The two step estimator is

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} R_n^2(\theta)$$

Initially proposed by Varah 1982 (with LS splines), commonly used since then.

Advantages

- No numerical integration : $R_n^2(\theta)$ is easier to compute and minimize than $Q_n(\theta)$
- R_n^2 does not depend on the type of smoothers ϕ_n
- Componentwise optimization (decoupled equations): for $j = 1, \ldots d$,

$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

- Intuitive interpretation and practical implementation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

Advantages

- No numerical integration : $R_n^2(\theta)$ is easier to compute and minimize than $Q_n(\theta)$
- R_n^2 does not depend on the type of smoothers ϕ_n
- Componentwise optimization (decoupled equations): for $j = 1, \ldots d$,

$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

- Intuitive interpretation and practical implementation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

Advantages

- No numerical integration : $R_n^2(\theta)$ is easier to compute and minimize than $Q_n(\theta)$
- R_n^2 does not depend on the type of smoothers ϕ_n
- Componentwise optimization (decoupled equations): for $j = 1, \ldots d$,

$$\hat{\theta}_n^{[j]} = \arg\min_{\theta^{[j]}} \left\| \hat{\dot{\phi}}_n^j - f_j(t, \hat{\phi}_n, \theta^{[j]}) \right\|_{L^2}^2$$

- Intuitive interpretation and practical implementation: Riemann discretization of the integral turns the optimization problem in a classical nonlinear regression problem

$$R_n^2(\theta) \approx \sum_{t_k} \left(\hat{\dot{\phi}}_n^j(t_k) - f_j(t_k, \hat{\phi}_n(t_k), \theta^{[j]}) \right)^2 \Delta t_k$$

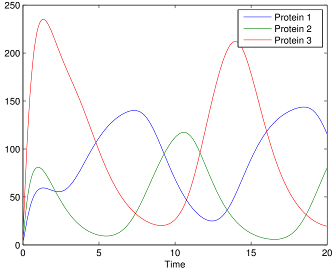
Outline

Example : Repressilator

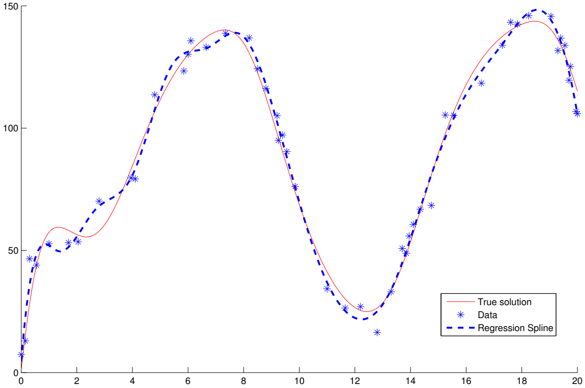
- Estimation of 16 parameters from 50 noisy observations of all the system ($V(\varepsilon) = 1$).

$$\begin{cases} \dot{r}_1(t) &= v_1^{max} \frac{k_{12}^n}{k_{12}^n + p_2(t)^n} - k_1^r r_1(t) \\ \dot{r}_2(t) &= v_2^{max} \frac{k_{23}^n}{k_{23}^n + p_3(t)^n} - k_2^r r_2(t) \\ \dot{r}_3(t) &= v_3^{max} \frac{k_{31}^n}{k_{31}^n + p_1(t)^n} - k_3^r r_3(t) \\ \dot{p}_1(t) &= \gamma_1 r_1(t) - k_1^p p_1(t) \\ \dot{p}_2(t) &= \gamma_2 r_2(t) - k_2^p p_2(t) \\ \dot{p}_3(t) &= \gamma_3 r_3(t) - k_3^p p_3(t) \end{cases}$$

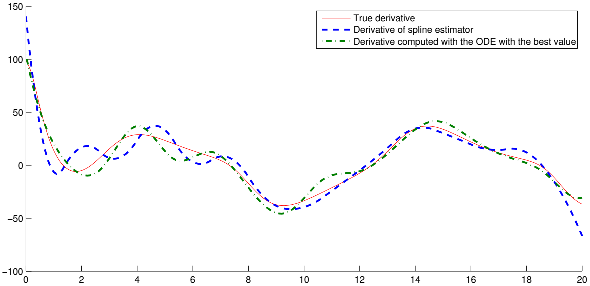
Evolution of protein concentrations



Nonparametric estimation of the solution by splines



Estimates of the derivative of $\hat{\phi}_n$

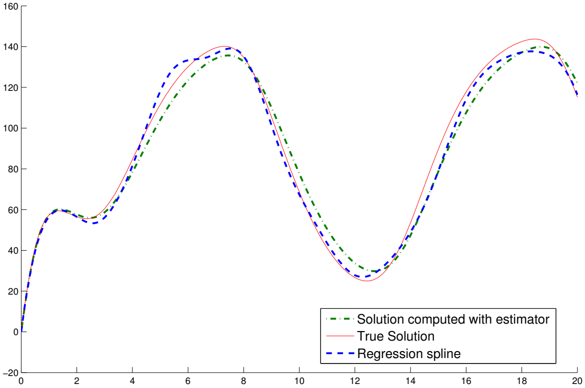


Estimated parameters

Component i	v_i^{max}	k_{ij}	k_i	γ_i	k_i^p
$i = 1$	134.3 (150)	50.5 (50)	0.9 (1)	0.96 (1)	0.97 (1)
$i = 2$	69 (80)	43 (40)	1 (1)	1.9 (2)	0.94 (1)
$i = 3$	125 (100)	47.4 (50)	1.1 (1)	2.9 (3)	0.97 (1)

and $n = 2.95(3)$.

The reconstructed curves



Outline

Statistical properties of two-step estimators: Consistency

- Introduction of a more general fitting criterion

$$R_{n,w}^2(\theta) = \int_0^1 \left\| \hat{\dot{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|^2 w(t)dt$$

with asymptotic version
$R_w^2(\theta) = \int_0^1 \left\| \dot{\phi}^*(t) - f(t, \phi^*(t), \theta) \right\|^2 w(t)dt$ and $w \geq 0$ is a weight function (ϕ^* is the true fonction with θ^*).

- Under assumptions:
  1. $\forall \theta \in \Theta, \forall t \in [0,1]$, $f(t, \cdot, \theta)$ is uniformly Lipschitz in (t, θ) on a compact $\mathscr{K} \subset \mathbb{R}^d$,
  2. $\forall \theta \in \Theta, \forall x_0, \forall t \in [0,1]$, $\phi(t, (x_0, \theta)) \in \mathscr{K}$,
  3. If $\hat{\phi}_n, \hat{\dot{\phi}}_n$ are consistent L^2 estimators, and $\hat{\phi}_n \in \mathscr{K}$ a.s.
  4. Identifiability condition: $\forall \varepsilon > 0$, $\inf_{\|\theta - \theta^*\| > \varepsilon} R_w^2(\theta) > R_w^2(\theta^*)$

then

$$\hat{\theta}_n - \theta^* = o_P(1).$$

Statistical properties of two-step estimators: Consistency

- Introduction of a more general fitting criterion

$$R^2_{n,w}(\theta) = \int_0^1 \left\| \hat{\dot{\phi}}_n(t) - f(t, \hat{\phi}_n(t), \theta) \right\|^2 w(t)dt$$

 with asymptotic version
$R^2_w(\theta) = \int_0^1 \left\|\dot{\phi}^*(t) - f(t, \phi^*(t), \theta)\right\|^2 w(t)dt$ and $w \geq 0$ is a weight function (ϕ^* is the true fonction with θ^*).

- Under assumptions:
  1. $\forall \theta \in \Theta, \forall t \in [0,1]$, $f(t,\cdot,\theta)$ is uniformly Lipschitz in (t,θ) on a compact $\mathscr{K} \subset \mathbb{R}^d$,
  2. $\forall \theta \in \Theta, \forall x_0, \forall t \in [0,1]$, $\phi(t,(x_0,\theta)) \in \mathscr{K}$,
  3. If $\hat{\phi}_n, \hat{\dot{\phi}}_n$ are consistent L^2 estimators, and $\hat{\phi}_n \in \mathscr{K}$ a.s.
  4. Identifiability condition: $\forall \varepsilon > 0$, $\inf_{\|\theta - \theta^*\| > \varepsilon} R^2_w(\theta) > R^2_w(\theta^*)$

then

$$\hat{\theta}_n - \theta^* = o_P(1).$$

Asymptotics: Is root-n rate preserved ?

- If f smooth enough, linearization of $R^2_{n,w}$ around θ^* enables to write

$$\hat{\theta}_n - \theta^* = \Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) + \Gamma_{b,w}\left(\hat{\phi}_n - \phi^*\right) + o_P(1)$$

  with smooth $\Gamma_{s,w}$ and nonsmooth $\Gamma_{b,w}$ functionals defined as

$$\begin{aligned}\Gamma_{s,w}(x) &= \int_0^1 \left(\partial_3 f(t,\phi^*,\theta^*)^\top \partial_2 f(t,\phi^*,\theta^*)w(t) - \partial_t\left(\partial_3 f(t,\phi^*,\theta^*)w(t)\right)\right)x(t) \\ \Gamma_{b,w}(x) &= w(1)\partial_3 f(t,\phi^*,\theta^*)^\top x(1) - w(0)\partial_3 f(t,\phi^*,\theta^*)^\top x(0)\end{aligned}$$

- For Splines, NW, local polynomials,...: Linear functionals are asymptotically normal, but $\Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) = O_P(n^{-1/2})$ and $\Gamma_{b,w}\left(\hat{\phi}_n - \phi^*\right) = O_P\left(n^{-m/(2m+1)}\right)$.

Asymptotics: Is root-n rate preserved ?

- If f smooth enough, linearization of $R^2_{n,w}$ around θ^* enables to write

$$\hat{\theta}_n - \theta^* = \Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) + \Gamma_{b,w}\left(\hat{\phi}_n - \phi^*\right) + o_P(1)$$

with smooth $\Gamma_{s,w}$ and nonsmooth $\Gamma_{b,w}$ functionals defined as

$$\begin{aligned} \Gamma_{s,w}(x) &= \int_0^1 \left(\partial_3 f(t,\phi^*,\theta^*)^\top \partial_2 f(t,\phi^*,\theta^*)w(t) - \partial_t\left(\partial_3 f(t,\phi^*,\theta^*)w(t)\right)\right)x(t) \\ \Gamma_{b,w}(x) &= w(1)\partial_3 f(t,\phi^*,\theta^*)^\top x(1) - w(0)\partial_3 f(t,\phi^*,\theta^*)^\top x(0) \end{aligned}$$

- For Splines, NW, local polynomials,...: Linear functionals are asymptotically normal, but $\Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) = O_P(n^{-1/2})$ *and* $\Gamma_{b,w}\left(\hat{\phi}_n - \phi^*\right) = O_P\left(n^{-m/(2m+1)}\right)$.

Optimal rate and boundary effects

- Asymptotics show the fundamental role of weigth function w:
  - Can enhance the discriminating power of criterion R_w^2
  - Get a root-n rate estimators with boundary vanishing weight function: $w(0) = w(1) = 0$.
- Consequence on asymptotics :

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \sqrt{n}\Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) \rightsquigarrow N(0, \Sigma)$$

  - good computational properties
  - classical parametric rate of convergence

Reference : *Parameter estimation of ODEs via nonparametric estimators*, Electronic Journal of Statistics, 2008, pp. 1242-1267.

Optimal rate and boundary effects

- Asymptotics show the fundamental role of weigth function w:
  - Can enhance the discriminating power of criterion R_w^2
  - Get a root-n rate estimators with boundary vanishing weight function: $w(0) = w(1) = 0$.

- Consequence on asymptotics :

$$\sqrt{n}(\hat{\theta}_n - \theta^*) = \sqrt{n}\Gamma_{s,w}\left(\hat{\phi}_n - \phi^*\right) \rightsquigarrow N(0, \Sigma)$$

  - good computational properties
  - classical parametric rate of convergence

Reference : *Parameter estimation of ODEs via nonparametric estimators*, Electronic Journal of Statistics, 2008, pp. 1242-1267.

Illustration on Lotka-Volterra system

Bias and standard deviation of two-step estimators with a weight function ($\hat{\theta}_{n,w}$) and with uniform weight ($\hat{\theta}_n$).

$$\begin{cases} \dot{x} = x(a_1 x + a_2 y + a_3) \\ \dot{y} = y(b_1 x + b_2 y + b_3) \end{cases} \quad \theta^* = (0, \mathbf{-1.5}, \mathbf{1}; 2, \mathbf{1}, \mathbf{-1.5})$$

n	mean$\hat{\theta}_{2,w}$	mean$\hat{\theta}_2$
20	$(-0.88, 0.56, 0.48, 0.03)$	$(-1.18, 0.74, 0.22, 0.38)$
50	$(-1.17, 0.78, 1.15, -0.94)$	$(-1.58, 1.03, 0.99, -0.69)$
100	$(-1.39, 0.92, 1.12, -0.93)$	$(-1.45, 0.95, 1.01, -0.78)$
200	$(-1.43, 0.95, 1.29, -1.19)$	$(-1.54, 1.0, 1.17, -1.00)$
500	$(-1.45, 0.97, 1.43, -1.40)$	$(-1.58, 1.05, 1.34, -1.26)$
1000	$(-1.47, 0.98, 1.47, -1.45)$	$(-1.58, 1.05, 1.41, -1.37)$

n	std$(\hat{\theta}_{2,w})$	std$(\hat{\theta}_2)$
20	$(0.27, 0.18, 0.36, 0.52)$	$(0.28, 0.20, 0.35, 0.49)$
50	$(0.18, 0.13, 0.33, 0.49)$	$(0.29, 0.20, 0.44, 0.64)$
100	$(0.18, 0.12, 0.33, 0.48)$	$(0.18, 0.19, 0.53, 0.77)$
200	$(0.13, 0.09, 0.25, 0.37$	$(0.24, 0.16, 0.43, 0.63)$
500	$(0.08, 0.05, 0.16, 0.24)$	$(0.19, 0.12, 0.31, 0.45)$
1000	$(0.05, 0.04, 0.10, 0.14)$	$(0.14, 0.09, 0.20, 0.29)$

Lotka-Volterra continued

- Evolution of RMSE: Mean Squared Errors of the weighted and unweighted parametric estimators versus the Mean Squared Error of the nonparametric estimators of the 2 curves.

n	MSE($\hat{\theta}_{2,w}$)	MSE($\hat{\theta}_2$)	MSE($\hat{\phi}$)
20	2.0	2.34	(0.93, 0.97)
30	1.48	1.88	(0.68, 0.73)
50	0.91	1.14	(0.57, 0.59)
100	0.83	1.18	(0.44, 0.40)
200	0.52	0.87	(0.53, 0.34)
500	0.28	0.56	(0.20, 0.20)
1000	0.17	0.37	(0.15, 0.14)

Amelioration of two-step estimators

$\hat{\theta}_n - \theta^*$ is controlled by $\hat{\phi}_n - \phi^*$and $\Gamma_{s,w} \implies$ 2 ways for ameliorating $\hat{\theta}_n$

1. Construction of better estimator $\hat{\phi}_n$
2. Better choice of w (more generally of $\Gamma_{s,w}$)

- Focus on *way 1*: "Good" nonparametric estimators $\hat{\phi}_n$ of ϕ^* can be obtained with prior information, such as
  - positivity, monotony, convexity i.e. shape-constrained inference,
  - known initial (boundary) values,
  - "semiparametric" estimation
- Motivation: constrained (nonparametric) estimation can give simpler estimation procedure (constrained parametric)

Amelioration of two-step estimators

$\hat{\theta}_n - \theta^*$ is controlled by $\hat{\phi}_n - \phi^*$and $\Gamma_{s,w} \implies$ 2 ways for ameliorating $\hat{\theta}_n$

1. Construction of better estimator $\hat{\phi}_n$
2. Better choice of w (more generally of $\Gamma_{s,w}$)

- Focus on *way 1*: "Good" nonparametric estimators $\hat{\phi}_n$ of ϕ^* can be obtained with prior information, such as
  - positivity, monotony, convexity i.e. shape-constrained inference,
  - known initial (boundary) values,
  - "semiparametric" estimation
- Motivation: constrained (nonparametric) estimation can give simpler estimation procedure (constrained parametric)

Outline

Meaningful decomposition

- Ameliorate the estimation of ϕ^* by writing the decomposition

$$\phi^*(t) = S(t) + N(t)$$

S : Main shape, trend
N : Transient behavior, perturbation w.r.t a reference situation.

"Refined" examples of possible shapes for S:

- Periodic solution (limit cycle) of nonlinear ODE

$$S(t) = \sum_{k=0}^{\infty} b_k \cos(2\pi k \omega t + \phi_k)$$

- Likely ("normal") parameters values $\theta_1, \ldots, \theta_\ell$

$$S(t) = \phi\left(t, (x_0, \theta_1)\right) \text{ or } \sum_{k=1}^{\ell} b_k \phi(t, (x_0, \theta_k))$$

Meaningful decomposition

- Ameliorate the estimation of ϕ^* by writing the decomposition

$$\phi^*(t) = S(t) + N(t)$$

S : Main shape, trend
N : Transient behavior, perturbation w.r.t a reference situation.

"Refined" examples of possible shapes for S:

- Periodic solution (limit cycle) of nonlinear ODE

$$S(t) = \sum_{k=0}^{\infty} b_k \cos(2\pi k\omega t + \phi_k)$$

- Likely ("normal") parameters values $\theta_1, \ldots, \theta_\ell$

$$S(t) = \phi\left(t, (x_0, \theta_1)\right) \text{ or } \sum_{k=1}^{\ell} b_k \phi(t, (x_0, \theta_k))$$

Constraints and "semiparametric" SVR

- Classical SVR (with RKHS $\mathscr{H}$ and associated kernel $k(\cdot,\cdot)$):

$$\hat{\phi}_n = \arg\min_{f\in\mathscr{H}} \sum_{i=1}^{n} L_\varepsilon(y_i - f(t_i)) + C\|f\|^2_{\mathscr{H}} \implies \hat{\phi}_n(t) = b + \sum_{i\in SV} c_i k(t_i,t)$$

with $L_\varepsilon(x) = \max(|x|-\varepsilon, 0)$, C=trade-off constant, SV = set of Support Vectors (k is typically a Gaussian kernel).

- Semiparametric SVR: $S \in \text{span}\{\psi_1,\ldots,\psi_\ell\}$:

$$\hat{\phi}_n = \arg\min_{N\in\mathscr{H}} \sum_{i=1}^{n} L_\varepsilon(y_i - (S(t_i)+N(t_i))) + C\|N\|^2_{\mathscr{H}}$$

$$\implies \hat{\phi}_n(t) = \underbrace{\sum_{k=1}^{\ell} b_k\psi_k(t)}_{\hat{S}(t)} + \underbrace{\sum_{i\in SV} c_i k(t_i,t)}_{\hat{N}(t)}$$

Coefficients b_k, c_i are computed as solution of constrained convex (quadratic) problem. If $\psi_k \in \mathscr{H}$, then $\hat{S}\perp\hat{N}$ in $\mathscr{H}$.

Constraints and "semiparametric" SVR

- Classical SVR (with RKHS $\mathscr{H}$ and associated kernel $k(\cdot,\cdot)$):

$$\hat{\phi}_n = \arg\min_{f\in\mathscr{H}} \sum_{i=1}^{n} L_\varepsilon(y_i - f(t_i)) + C\,\|f\|^2_{\mathscr{H}} \implies \hat{\phi}_n(t) = b + \sum_{i\in SV} c_i k(t_i,t)$$

with $L_\varepsilon(x) = \max(|x|-\varepsilon, 0)$, C=trade-off constant, SV = set of Support Vectors (k is typically a Gaussian kernel).

- Semiparametric SVR: $S \in \text{span}\{\psi_1,\ldots,\psi_\ell\}$:

$$\hat{\phi}_n = \arg\min_{N\in\mathscr{H}} \sum_{i=1}^{n} L_\varepsilon\left(y_i - (S(t_i)+N(t_i))\right) + C\,\|N\|^2_{\mathscr{H}}$$

$$\implies \hat{\phi}_n(t) = \underbrace{\sum_{k=1}^{\ell} b_k\psi_k(t)}_{\hat{S}(t)} + \underbrace{\sum_{i\in SV} c_i k(t_i,t)}_{\hat{N}(t)}$$

Coefficients b_k, c_i are computed as solution of constrained convex (quadratic) problem. If $\psi_k \in \mathscr{H}$, then $\hat{S}\perp\hat{N}$ in $\mathscr{H}$.

Examples of constraints

- More prior information via constraints on values of $\hat{\phi}_n$:

$$\hat{\phi}_n = \arg\min_{f \in \mathscr{H}} \sum_{i=1}^{n} L_\varepsilon(y_i - f(t_i)) + C \|f\|_{\mathscr{H}}^2$$
$$s.t. f(0) = x_0, f(T) = x_T$$

Still convex problem to solve for $b, c_i's$.

- Shape constraints:

Truncated Fourier Series $\hat{S} = \sum_{k \leq \ell} b_\ell \cos(2\pi k \omega t + \phi_k)$
(with known ω, ϕ_k $k = 1 \ldots \ell$)

Mixtures of Solutions $\hat{S} = \sum_{k=1}^{\ell} b_k \phi(t, (x_{0,k}, \theta_k))$
(with known $(x_{0,k}, \theta_k)$ $k = 1 \ldots \ell$)

Examples of constraints

- More prior information via constraints on values of $\hat{\phi}_n$:

$$\hat{\phi}_n = \arg\min_{f \in \mathscr{H}} \sum_{i=1}^{n} L_\varepsilon(y_i - f(t_i)) + C \|f\|^2_{\mathscr{H}}$$
$$s.t. f(0) = x_0, f(T) = x_T$$

Still convex problem to solve for b, $c_i's$.

- Shape constraints:

Truncated Fourier Series $\hat{S} = \sum_{k \leq \ell} b_\ell \cos(2\pi k \omega t + \phi_k)$
(with known ω, ϕ_k $k = 1 \ldots \ell$)
Mixtures of Solutions $\hat{S} = \sum_{k=1}^{\ell} b_k \phi(t, (x_{0,k}, \theta_k))$
(with known $(x_{0,k}, \theta_k)$ $k = 1 \ldots \ell$)

Comparisons of constraints in Repressilator

Use of $\hat{\phi}_n^{period}$ with $\ell = 2, \omega = 0.55$;
$\hat{\phi}_n^{ode}$, with $\theta_1 = \theta^* + 0.2$, $\theta_2 = \theta^* - 0.1$; $\hat{\phi}_n^{cons}$ with $\hat{\phi}_n^{cons}(0) = \phi^*(0)$ and $\hat{\phi}_n^{cons}(T) = \phi^*(T)$.

	True Parameter	$\hat{\phi}_n$	$\hat{\phi}_n^{period}$	$\hat{\phi}_n^{ode}$	$\hat{\phi}_n^{cons}$
v_1	150	150.0	113.2	149.1	149.17
v_2	80	79.99	87.6	78.2	78.4
v_3	100	101.98	81.2	101.8	100.6
$k_{1,2}$	50	50.5	66.6	50.4	50.5
$k_{2,3}$	40	40.4	53.2	40.1	40.5
$k_{3,1}$	50	49.65	39.0	48.9	50.2
k_1	1	0.98	1.24	1.23	0.99
k_2	2	1.96	1.9	1.91	1.95
k_3	3	2.85	3.21	3.18	2.8

Table: Mean of the two-step estimator computed with different estimators of the true solution of the data when $T = 40$ observations, computed with 100 Monte Carlo runs

Conclusion : ameliorations and future works

- Amelioration of the nonparametric estimation $\hat{\phi}_n$:
  - adaptive estimation,
  - better shape constrained estimation procedures with true semiparametric procedures, and shape constrained inference (e.g. monotonicity).
- Adaptive optimal choice for weight function w in $R^2_{n,w}$
- Develop efficient algorithms for partially observed differential equations
- Develop Bayesian version