Reproducing kernel Hilbert space based estimation of systems of ordinary differential equations

Javier González, Ivan Vujačić and Ernst Wit

University of Groningen
Bernoulli Institute for Mathematics and Computer Science

PEDSII, June, 2012

rijksuniversiteit groningen

Notation

Consider a dynamical system modelled by a set of ODEs

$$P_{\boldsymbol{\theta}_j} x_j = f_{\beta_j}(x_1, \ldots, x_m, u_j), \;\; j = 1, \ldots, m$$

that describes the time evolution of m interacting elements, *e. g.*,

- gene regulatory networks in system biology,
- prey-predators systems in ecology or bussines.

Elements

- x_j, u_j: state variables and external forces defined on a time interval T.
- $P_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} D^k$ with $D^k = d^k/dt$, $k \in \mathbb{N}$ and $\theta_j = \{\theta_{j1}, \ldots, \theta_{jd}\}$.
- f_{β_j} known parametric function where $\beta_j = \{\beta_{j1}, \ldots, \beta_{jq}\}$.
- $\Theta = \{\theta_1, \ldots, \theta_m\}$ and $B = \{\beta_1, \ldots, \beta_1\}$, parameters.

Notation

Consider a dynamical system modelled by a set of ODEs

$$P_{\boldsymbol{\theta}_j} x_j = f_{\beta_j}(x_1, \ldots, x_m, u_j), \;\; j = 1, \ldots, m$$

that describes the time evolution of m interacting elements, *e. g.*,

- gene regulatory networks in system biology,
- prey-predators systems in ecology or bussines.

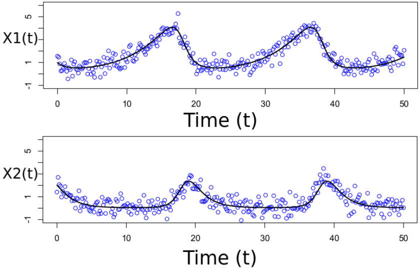
Elements

- x_j, u_j: state variables and external forces defined on a time interval T.
- $P_{\boldsymbol{\theta}_j} = \sum_{k=0}^{d} \theta_{jk} D^k$ with $D^k = d^k/dt$, $k \in \mathbb{N}$ and $\theta_j = \{\theta_{j1}, \ldots, \theta_{jd}\}$.
- f_{β_j} known parametric function where $\beta_j = \{\beta_{j1}, \ldots, \beta_{jq}\}$.
- $\Theta = \{\theta_1, \ldots, \theta_m\}$ and $B = \{\beta_1, \ldots, \beta_1\}$, parameters.

Problem statement

Noisy measurements y_{ij} of the state variables $x_1, \ldots, x_m$ at n time points.

$$y_{ji} \sim \mathcal{N}(x_j(t_i), \sigma_j^2)$$



$$\frac{d}{dt}x_1 = x_1(\theta_1 - \beta_1 x_2),$$
$$\frac{d}{dt}x_2 = -x_2(\theta_2 - \beta_2 x_1),$$

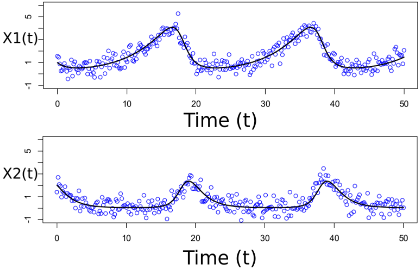
Problem to solve

Use the sample $S = \{(y_{ji}, t_i) \in \mathbb{R} \times T\}_{i,j=1}^{n,m}$ to provide estimators of $\Theta = \{\theta_1, \theta_2\}, B = \{\beta_1, \beta_2\}, \Sigma = \{\sigma_1, \sigma_2\}$.

Problem statement

Noisy measurements y_{ij} of the state variables $x_1, \ldots, x_m$ at n time points.

$$y_{ji} \sim \mathcal{N}(x_j(t_i), \sigma_j^2)$$



$$\frac{d}{dt}x_1 = x_1(\theta_1 - \beta_1 x_2),$$
$$\frac{d}{dt}x_2 = -x_2(\theta_2 - \beta_2 x_1),$$

Problem to solve

Use the sample $S = \{(y_{ji}, t_i) \in \mathbb{R} \times T\}_{i,j=1}^{n,m}$ to provide estimators of $\Theta = \{\theta_1, \theta_2\}, B = \{\beta_1, \beta_2\}, \Sigma = \{\sigma_1, \sigma_2\}$.

Likelihood approach

Likelihood approach

$$l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) = -\frac{n}{2}\log(\sigma_j^2) - \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(y_{ji} - x(t_i))^2$$

for x_j satisfying that $P_{\theta_j} x_j = f_{\beta_j}$.

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta,B,\Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j)$$

- To solve the ODE is needed.
- Parameter identification might be non-stable with noisy data.

Likelihood approach

Likelihood approach

$$l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) = -\frac{n}{2}\log(\sigma_j^2) - \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(y_{ji} - x(t_i))^2$$

for x_j satisfying that $P_{\theta_j} x_j = f_{\beta_j}$.

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j)$$

- To solve the ODE is needed.
- Parameter identification might be non-stable with noisy data.

Likelihood approach

Likelihood approach

$$l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) = -\frac{n}{2}\log(\sigma_j^2) - \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(y_{ji} - x(t_i))^2$$

for x_j satisfying that $P_{\theta_j} x_j = f_{\beta_j}$.

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j)$$

- To solve the ODE is needed.
- Parameter identification might be non-stable with noisy data.

Likelihood approach

Likelihood approach

$$l_j(\theta_j, \beta_j, \sigma_j, x_j|S_j) = -\frac{n}{2}\log(\sigma_j^2) - \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(y_{ji} - x(t_i))^2$$

for x_j satisfying that $P_{\theta_j} x_j = f_{\beta_j}$.

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n|S) = \arg max_{\Theta,B,\Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j|S_j)$$

- To solve the ODE is needed.
- Parameter identification might be non-stable with noisy data.

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta,B,\Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta,B,\Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

Regularization approach

- Regularization methods are appropiate in this context.
- Replace the original problem by a family of problems where the ODE is used to penalize the likelihood.

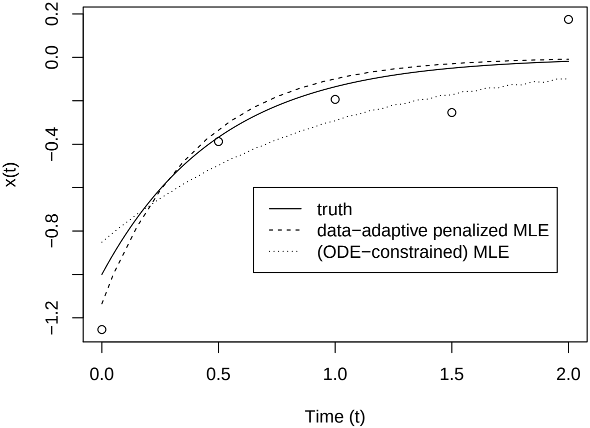
Penalized Likelihood approach

$$(\hat{\Theta}, \hat{B}, \hat{\Sigma}, \hat{x}_1, \ldots, \hat{x}_n | S) = \arg max_{\Theta, B, \Sigma} \sum_{j=1}^{m} l_j(\theta_j, \beta_j, \sigma_j, x_j | S_j) - \lambda \Omega_j(x_j)$$

for $\lambda > 0$ and $\Omega(x_j)$ a convex functional.

How to define $\Omega(x_j)$?

MLE vs PMLE: $dx/dt = \theta x$.



$$\theta_{true} = -2,\ \theta_{MLE} = -1{,}12,\ \theta_{PMLE} = -1{,}99$$

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the $x_j's$ by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the x_j's by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the $x_j's$ by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the $x_j's$ by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the $x_j's$ by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Regularization approaches in the literature

Likelihood based approaches [Ramsay et al., 2007, Bouchet, 2007]

- Estimation of the $x_j's$ by nonparametric regression (splines, SVR).
- Differentiation of $\hat{x}_j$ and minimization over the parameters using $\Omega(\hat{x}_j) = \|P_{\theta_j}\hat{x}_j - f_{\beta_j}(\hat{x}_1, \ldots, \hat{x}_m, u_j))\|_{L_2}$.

Other approaches

- Bayesian method similar in spirit to Ramsay et al. (2007). Solution of the ODE given as a Gaussian process [Calderhead et al., 2008].
- Kernel Method for estimating 1-dimensional, periodic differential equations [Steinke et al. 2008].

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\theta_j} x_j = 0$, generalization to non-homogeneous is feasible.
- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}} = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|^2_{\mathcal{H}} = 0$, x_j is a solution of $P_{\theta_j} x_j = 0$.
- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|^2_{\mathcal{H}}$ to $\|P_{\theta_j} \tilde{x}_j\|^2_{\mathcal{H}}$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\theta_j} x_j = 0$, generalization to non-homogeneous is feasible.
- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}} = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|^2_{\mathcal{H}} = 0$, x_j is a solution of $P_{\theta_j} x_j = 0$.
- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|^2_{\mathcal{H}}$ to $\|P_{\theta_j} \tilde{x}_j\|^2_{\mathcal{H}}$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\theta_j} x_j = 0$, generalization to non-homogeneous is feasible.
- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}} = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|^2_{\mathcal{H}} = 0$, x_j is a solution of $P_{\theta_j} x_j = 0$.
- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|^2_{\mathcal{H}}$ to $\|P_{\theta_j} \tilde{x}_j\|^2_{\mathcal{H}}$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\boldsymbol{\theta}_j} x_j = 0$, generalization to non-homogeneous is feasible.
- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}} = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|^2_{\mathcal{H}} = 0$, x_j is a solution of $P_{\theta_j} x_j = 0$.
- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|^2_{\mathcal{H}}$ to $\|P_{\theta_j} \tilde{x}_j\|^2_{\mathcal{H}}$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\boldsymbol{\theta}_j} x_j = 0$, generalization to non-homogeneous is feasible.

- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|_{\mathcal{H}}^2 = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|_{\mathcal{H}}^2 = 0$, x_j is a solution of $P_{\boldsymbol{\theta}_j} x_j = 0$.

- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|_{\mathcal{H}}^2$ to $\|P_{\theta_j} \tilde{x}_j\|_{\mathcal{H}}^2$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Idea and approach

Idea

1. Combine the frequentist set-up with the kernel approach.
2. Parameter estimation problem as the maximization of a likelihood with a Reproducing kernel Hilbert space (RKHS) based penalty.

- $P_{\boldsymbol{\theta}_j} x_j = 0$, generalization to non-homogeneous is feasible.
- Penalty, P_{θ_j} is a differential operator on some space of functions $\mathcal{H}$

$$\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}} = \int_T (P_{\theta_j} x_j(t))^2 dt.$$

- When $\|x_j\|^2_{\mathcal{H}} = 0$, x_j is a solution of $P_{\boldsymbol{\theta}_j} x_j = 0$.
- Non homogeneous: Transform $\|P_{\theta_j} x_j - f_{\beta_j}\|^2_{\mathcal{H}}$ to $\|P_{\theta_j} \tilde{x}_j\|^2_{\mathcal{H}}$ where the $\tilde{x}_j$ depends on β_j. Transform the y_{ij} (details next talk).

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P^*_\theta P_\theta$.
- $P^*_\theta P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$.
- $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P^*_\theta P_\theta$.
- $P^*_\theta P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$.
- $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \rightarrow \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g\rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$.
- $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g\rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$.
- $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Properties of $\Omega_j(x_j) = \|x_j\|^2_{\mathcal{H}}$

RKHS in a nutshell

- Mercer kernel: continous, symmetric and positive definite function $K : T \times T \to \mathbb{R}$.
- RKHS: completed space spanned by $x(t) = \sum_{i=1}^{n} \alpha_i K(t_i, t)$, where $n \in \mathbb{N}$, $t_i \in T$ and $\alpha_i \in \mathbb{R}$ and $\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j K(t_i, t_j)$.

- $\mathcal{H}$ is a RKHS whose reproducing kernel is a Green's function of $P_\theta^* P_\theta$.
- $P_\theta^* P_\theta K(t, z) = \delta(t - z)$.
- Functions in $\mathcal{H}$ are characterized by vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$.

$$\sum_{j=1}^{m} \left[-\frac{n}{2} \log(\sigma_j^2) - \frac{1}{2\sigma_j^2} \|\mathbf{y}_j - \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j\|^2 - \lambda \boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j \right]$$

where $(\mathbf{K}_{\theta_j})_{is} = K_{\theta_j}(t_i, t_s)$ and $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P^*_{\theta_j} P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

  where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.
- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P^*_{\theta_j} P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.

- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P^*_{\theta_j} P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.

- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P_{\theta_j}^* P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.
- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P^*_{\theta_j} P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

  where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.
- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Computation of $\mathbf{K}_{\theta_j}$

- A Green's function for $P^*_{\theta_j} P_{\theta_j}$ might be hard or impossible to compute.
- Replace $\boldsymbol{\alpha}_j^T \mathbf{K}_{\theta_j} \boldsymbol{\alpha}_j$ by an approximation $\boldsymbol{\alpha}_j^T \tilde{\mathbf{K}}_{\theta_j} \boldsymbol{\alpha}_j$.
- $\mathbf{P}_{\theta_j} = \sum_{k=0}^{d} \theta_{jk} \mathbf{D}^k$: difference operator defined on $t_1, \ldots, t_n$ and

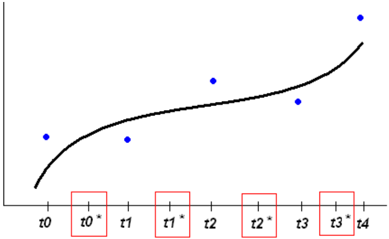
$$\mathbf{D} = \Delta^{-1} \cdot \begin{pmatrix} -1 & 1 & & & & \\ -1 & 0 & 1 & & & \\ & & & \ddots & & \\ & & & -1 & 0 & 1 \\ & & & & -1 & 1 \end{pmatrix}.$$

  where $\Delta = diag(t_2 - t_1, t_4 - t_2, \ldots, t_n - t_{n-2}, t_n - t_{n-1})$.
- Focus on the difference equation $\mathbf{P}_{\theta_j} \mathbf{x}_j = 0$.
- $\tilde{\mathbf{K}}_{\theta_j} = (\mathbf{P}_{\theta_j}^T \mathbf{P}_{\theta_j})^{-1}$.

Finite dimensional approximation, EM algorithm

Idea

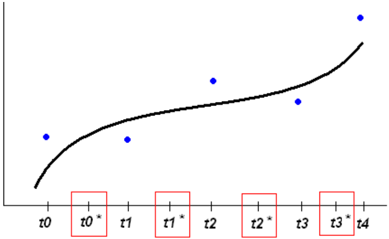
- To reduce the error of the finite dimensional approximation.
- To include a number of hidden data points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- K_{θ_j} only depends on the $t_i's$.
- More points -> better approximations of the derivatives.
- Iterate EM algorithm.
  1. **E-step** Expectation of the likelihood over $\mathbf{y}_H$.
  2. **M-step** Maximize the likelihood over the parameters using $(\mathbf{y}_O, \hat{\mathbf{y}}_H)$.



Finite dimensional approximation, EM algorithm

Idea

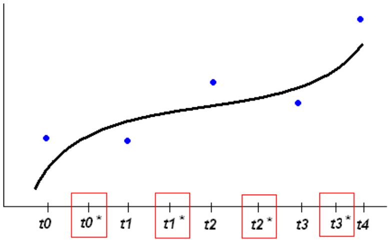
- To reduce the error of the finite dimensional approximation.
- To include a number of hidden data points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- K_{θ_j} only depends on the $t_i's$.
- More points -> better approximations of the derivatives.
- Iterate EM algorithm.
  1. **E-step** Expectation of the likelihood over $\mathbf{y}_H$.
  2. **M-step** Maximize the likelihood over the parameters using $(\mathbf{y}_O, \hat{\mathbf{y}}_H)$.



Finite dimensional approximation, EM algorithm

Idea

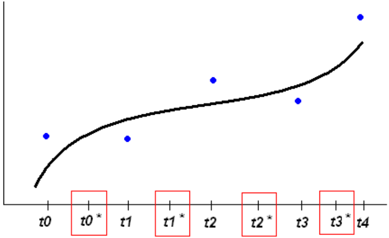
- To reduce the error of the finite dimensional approximation.
- To include a number of hidden data points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- K_{θ_j} only depends on the $t_i's$.
- More points -> better approximations of the derivatives.
- Iterate EM algorithm.
  1. **E-step** Expectation of the likelihood over $\mathbf{y}_H$.
  2. **M-step** Maximize the likelihood over the parameters using $(\mathbf{y}_O, \hat{\mathbf{y}}_H)$.



Finite dimensional approximation, EM algorithm

Idea

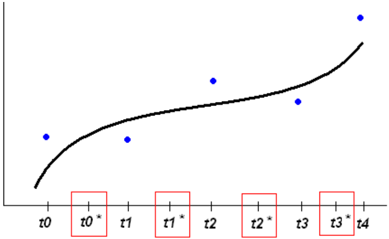
- To reduce the error of the finite dimensional approximation.
- To include a number of hidden data points $(\mathbf{t}^*_H, \mathbf{y}^*_H)$.
- K_{θ_j} only depends on the $t_i's$.
- More points -> better approximations of the derivatives.
- Iterate EM algorithm.
  1. **E-step** Expectation of the likelihood over $\mathbf{y}_H$.
  2. **M-step** Maximize the likelihood over the parameters using $(\mathbf{y}_O, \hat{\mathbf{y}}_H)$.



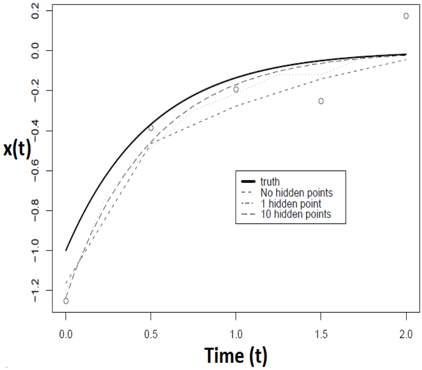
Finite dimensional approximation, EM algorithm

Idea

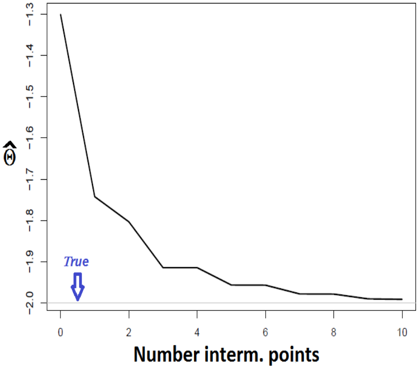
- To reduce the error of the finite dimensional approximation.
- To include a number of hidden data points $(\mathbf{t}_H^*, \mathbf{y}_H^*)$.
- K_{θ_j} only depends on the $t_i's$.
- More points -> better approximations of the derivatives.
- Iterate EM algorithm.
  1. **E-step** Expectation of the likelihood over $\mathbf{y}_H$.
  2. **M-step** Maximize the likelihood over the parameters using $(\mathbf{y}_O, \hat{\mathbf{y}}_H)$.



Finite dimensional approximation, EM algorithm



(a) ODE model, $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$. True function and obtained solution for 0, 1, and 10 intermediate points.



(b) Estimation of θ for different number of intermediate points.

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	**Proposed method**
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

Comparisons

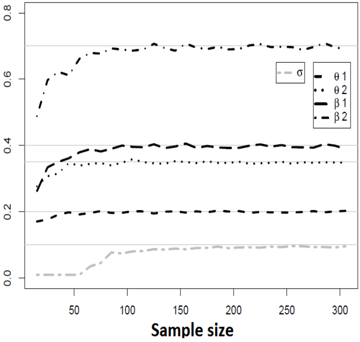
- PMLE vs. TS-Ramsay approaches in small-sample-size cases.
- Model $dx/dt = \theta x$ with x(0)=-1 and $\theta = -2$.
- 100 independent data sets of size 5.
- PMLE method with 10 equally spaced points between each pair of observed data.
- Penalization λ selected using the GCV criteria.
- 100 iterations of the EM algorithm.

TS-Ramsay	Proposed method
-1.6708 (0.2809)	-1.9448 (0.1594)

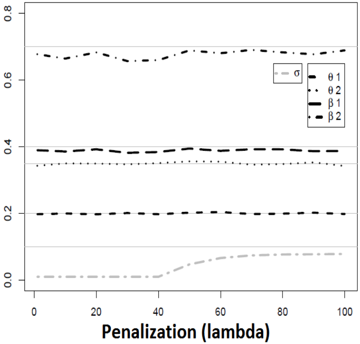
Simulated result, the Lotka-Volterra system

$$\frac{dx_1}{dt} = x_1(\theta_1 - \beta_1 x_2), \quad \frac{dx_2}{dt} = -x_2(\theta_2 - \beta_2 x_2)$$

$\theta_1 = 0{,}2$, $\beta_1 = 0{,}35$ $\theta_2 = 0{,}7$ and $\beta_2 = 0{,}40$, $x_{1,0} = 1$, $x_{2,0} = 2$.



(a) $\lambda = 100$ and a level noise of $\sigma = 0{,}1$.



(b) $n = 100$ and a level noise of $\sigma = 0{,}1$.

Conclusions and final remarks

- General methodology to estimate the parameters of system of ordinary differential equations in presence of noisy data.
- The system of equations is directly used as regularizer in the likelihood. A RKHS framework is used for this task. No need to solve the ODE to estimate the parameters.
- Method specially useful in problems with small samples. EM algorithm allows to incorporate into the system missing (or hidden) observations.
- Performance in real applications, next talk!

Conclusions and final remarks

- General methodology to estimate the parameters of system of ordinary differential equations in presence of noisy data.
- The system of equations is directly used as regularizer in the likelihood. A RKHS framework is used for this task. No need to solve the ODE to estimate the parameters.
- Method specially useful in problems with small samples. EM algorithm allows to incorporate into the system missing (or hidden) observations.
- Performance in real applications, next talk!

Conclusions and final remarks

- General methodology to estimate the parameters of system of ordinary differential equations in presence of noisy data.
- The system of equations is directly used as regularizer in the likelihood. A RKHS framework is used for this task. No need to solve the ODE to estimate the parameters.
- Method specially useful in problems with small samples. EM algorithm allows to incorporate into the system missing (or hidden) observations.
- Performance in real applications, next talk!

Conclusions and final remarks

- General methodology to estimate the parameters of system of ordinary differential equations in presence of noisy data.
- The system of equations is directly used as regularizer in the likelihood. A RKHS framework is used for this task. No need to solve the ODE to estimate the parameters.
- Method specially useful in problems with small samples. EM algorithm allows to incorporate into the system missing (or hidden) observations.
- Performance in real applications, next talk!

Conclusions and final remarks

- General methodology to estimate the parameters of system of ordinary differential equations in presence of noisy data.
- The system of equations is directly used as regularizer in the likelihood. A RKHS framework is used for this task. No need to solve the ODE to estimate the parameters.
- Method specially useful in problems with small samples. EM algorithm allows to incorporate into the system missing (or hidden) observations.
- Performance in real applications, next talk!

The Green's function of a differential operator P

Definition

Let $T \in [a, b] \in \mathbb{R}$ and let $P : \mathcal{H} \longrightarrow L^2(T)$ be a differential operator on a class of functions $\mathcal{H}$ then the Green's function of P is a function such that

$$PG(s,t) = \delta(s-t)$$

where $s, t \in T$

Remark

Notice that this equality holds in the distributional sense. This means that for $f \in L^2(T)$ then

$$\langle PG(s,t), f\rangle = \langle \delta(s-t), f\rangle = f(t)$$

Connection between Differential Operators, Green's functions and Kernels

Theorem

Let $T = \mathbb{R}^d$ and P a differential operator on a class of functions $\mathcal{H}$ such that, endowed with the inner product:

$$\langle f, g \rangle_{\mathcal{H}} = \langle Pf, Pg \rangle_{L^2(T)}$$

where $(f, g) \in \mathcal{H}^2$ it is a Hilbert space. Then $\mathcal{H}$ is a RKHS that admits as reproducing kernel the Green function of the operator P^*P, where P^* denotes the adjoint operator of P.

Idea of the proof

Let H be a Hilbert space endowed with the inner product

$$\langle f, g\rangle_{\mathcal{H}} = \langle Pf, Pg\rangle_{L^2(T)}$$

and K be the Green function of the operator P^*P , that is

$$P^*PK(s,t) = \delta(s-t)$$

Then, for all $s \in T$, (the evaluation functionals) $K_t = K(t, \cdot) \in \mathcal{H}$ because:

- The evaluation functional K_t are bounded.
- K_t has the reproducing property: for all $f \in \mathcal{H}$ and $x \in X$, we have that

$$\langle K_t, f\rangle_{\mathcal{H}} = \langle PK_t, Pf\rangle_{L^2(T)} = \langle P^*PK_t, f\rangle_{L^2(T)} = \langle \delta(s-t), f\rangle_{L^2(T)} =$$

$$= f(t)$$

Non homogeneous equation I

- $\|P_{\boldsymbol{\theta}}x - f_{\beta}\|^2$ cannot be used as a norm in an RKHS.
- If $x = 0$ then $\|P_{\boldsymbol{\theta}}x - f_{\beta}\|^2$ is not necessarily zero.
- Let G be a Green's function of P_{θ} and take

$$\tilde{x}(t) = x(t) - x^*(t), \tag{1}$$

where $x^*(t) = \int_T G(z,t)f_{\beta}(z)dz$ is effectively a collection of solutions of the differential equation.

Non homogeneous equation II

- $\tilde{x}$ can be calculated independent from the sample S.
- Since P_θ is a linear operator we have that for all $\tilde{x}$

$$P_{\boldsymbol{\theta}}\tilde{x}(t) = P_\theta x(t) - P_\theta x^*(t) = P_\theta x(t) - f_\beta(t),$$

  including for the trivial solution $\tilde{x} = 0$.
- Then $\|P_{\boldsymbol{\theta}}\tilde{x}\|^2 = \|P_{\boldsymbol{\theta}}x - f_\beta\|^2$ and we can use $\|P_{\boldsymbol{\theta}}\tilde{x}\|$ as a penalty
- This requires the transformation of the original observations,

$$\tilde{y}_i = y_i - x^*(t_i)$$

  for $j = 1, \ldots, n$.
- In the discrete case G is $\mathbf{P}_\theta^{-1}$

Transformation, the Lotka-Volterra system

$$\frac{dx_1}{dt} = x_1(\theta_1 - \beta_1 x_2), \quad \frac{dx_2}{dt} = -x_2(\theta_1 - \beta_2 x_2)$$

- $\tilde{\mathbf{y}}_1 = \mathbf{y}_1 - (\mathbf{D} - \theta_1 \mathbf{I})^{-1} \beta_1 (\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2)$
- $\tilde{\mathbf{y}}_2 = \mathbf{y}_2 - (\mathbf{D} - \theta_2 \mathbf{I})^{-1} \beta_2 (\hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2)$

where $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ are spline smoothers of the original data.