Latent Forces Models using Gaussian Processes

Mauricio A. Álvarez

(joint with David Luengo and Neil D. Lawrence)

Latent Force Models Workshop, Sheffield, June 2013

Contenido

Data driven paradigm

- Traditionally, the main focus in machine learning has been model generation through a *data driven paradigm*.

- Combine a data set with a flexible class of models and, through regularization, make predictions on unseen data.

- Problems
  - Data is scarce relative to the complexity of the system.
  - Model is forced to extrapolate.

Mechanistic models

- Models inspired by the underlying knowledge of a physical system are common in many areas.
- Description of a well characterized physical process that underpins the system, typically represented with a set of differential equations.
- Identifying and specifying all the interactions might not be feasible.
- A mechanistic model can enable accurate prediction in regions where there may be no available training data

Hybrid systems

- We suggest a *hybrid approach* involving a mechanistic model of the system augmented through machine learning techniques.
- Dynamical systems (e.g. incorporating first order and second order differential equations).
- Partial differential equations for systems with multiple inputs.

Contenido

Latent variable model: definition

- Our approach can be seen as a type of latent variable model.

$$\mathbf{Y} = \mathbf{U}\mathbf{W}^\top + \mathbf{E},$$

where $\mathbf{Y} \in \mathbb{R}^{N\times D}$, $\mathbf{U} \in \mathbb{R}^{N\times Q}$, $\mathbf{W} \in \mathbb{R}^{D\times Q}$ ($Q < D$) and $\mathbf{E}$ is a matrix variate white Gaussian noise with columns $\mathbf{e}_{:,d} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$.

- In PCA and FA the common approach to deal with the unknowns is to integrate out $\mathbf{U}$ under a Gaussian prior and optimize with respect to $\mathbf{W}$.

Latent variable model: alternative view

- Data with temporal nature and Gaussian (Markov) prior for rows of $\mathbf{U}$ leads to the Kalman filter/smoother.

- Consider a joint distribution for $p(\mathbf{U}|\mathbf{t})$, $\mathbf{t} = [t_1 \dots t_N]^\top$, with the form of a Gaussian process (GP),

$$p(\mathbf{U}|\mathbf{t}) = \prod_{q=1}^{Q} \mathcal{N}\left(\mathbf{u}_{:,q}|\mathbf{0}, \mathbf{K}_{u_{:,q},u_{:,q}}\right).$$

  The latent variables are random functions, $\{u_q(t)\}_{q=1}^{Q}$ with associated covariance $\mathbf{K}_{u_{:,q},u_{:,q}}$.

- The GP for $\mathbf{Y}$ can be readily implemented.

Latent force model: mechanistic interpretation (1)

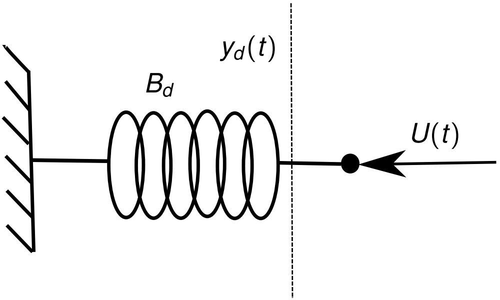
- We include a further dynamical system with a *mechanistic* inspiration.

- Reinterpret equation $\mathbf{Y} = \mathbf{U}\mathbf{W}^\top + \mathbf{E}$, as a force balance equation

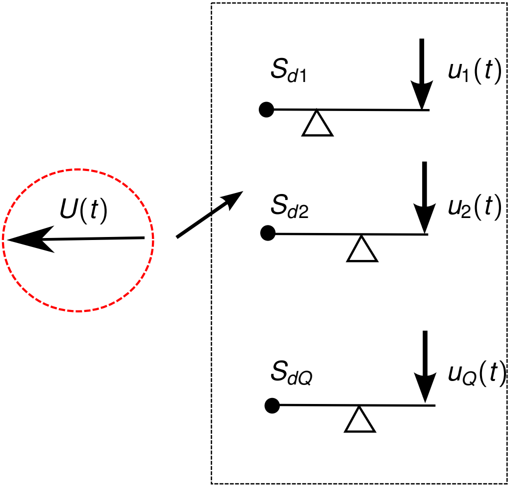
$$\mathbf{Y}\mathbf{B} = \mathbf{U}\mathbf{S}^\top + \widetilde{\mathbf{E}},$$

  where $\mathbf{S} \in \mathbb{R}^{D\times Q}$ is a matrix of sensitivities, $\mathbf{B} \in \mathbb{R}^{D\times D}$ is diagonal matrix of spring constants, $\mathbf{W}^\top = \mathbf{S}^\top\mathbf{B}^{-1}$ and $\widetilde{\mathbf{e}}_{:,d} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{B}^\top\mathbf{\Sigma}\mathbf{B}\right)$.

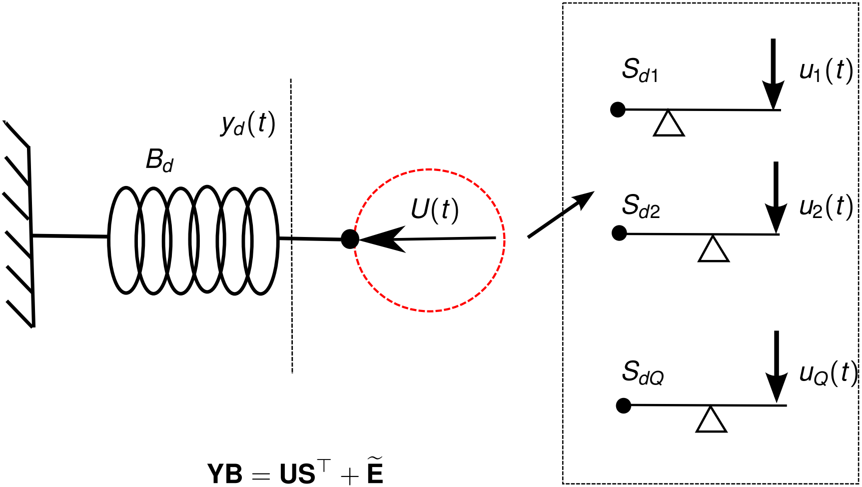
Latent force model: mechanistic interpretation (2)



Latent force model: mechanistic interpretation (2)



Latent force model: mechanistic interpretation (2)



$$\mathbf{YB} = \mathbf{US}^\top + \widetilde{\mathbf{E}}$$

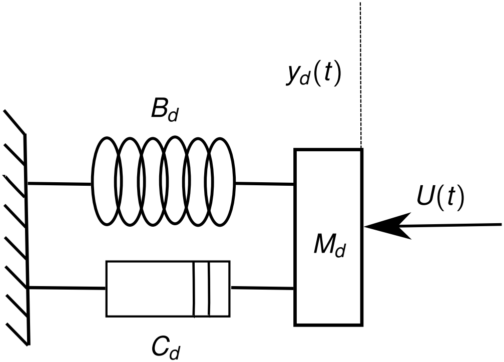
Latent force model: extension (1)

- The model can be extended including dampers and masses.

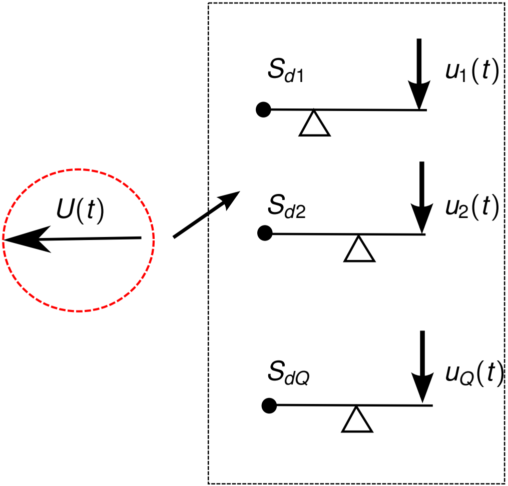
- We can write
$$\ddot{\mathbf{Y}}\mathbf{M} + \dot{\mathbf{Y}}\mathbf{C} + \mathbf{Y}\mathbf{B} = \mathbf{U}\mathbf{S}^{\top} + \widehat{\boldsymbol{E}},$$
  where

  $\dot{\mathbf{Y}}$ is the first derivative of $\mathbf{Y}$ w.r.t. time
  $\ddot{\mathbf{Y}}$ is the second derivative of $\mathbf{Y}$ w.r.t. time
  $\mathbf{C}$ is a diagonal matrix of damping coefficients
  $\mathbf{M}$ is a diagonal matrix of masses
  $\widehat{\boldsymbol{E}}$ is a matrix variate white Gaussian noise.

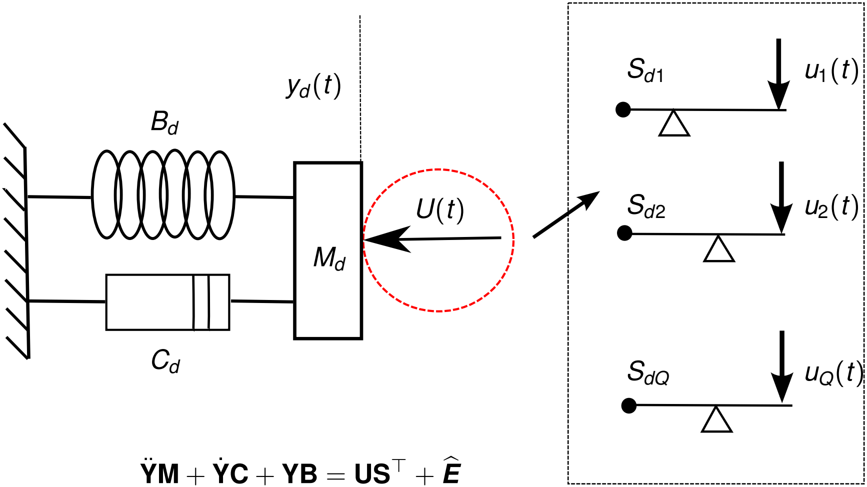
Latent force model: extension (2)



Latent force model: extension (2)



Latent force model: extension (2)



$$\ddot{\mathbf{Y}}\mathbf{M} + \dot{\mathbf{Y}}\mathbf{C} + \mathbf{Y}\mathbf{B} = \mathbf{U}\mathbf{S}^\top + \widehat{\boldsymbol{E}}$$

Latent force model: properties

- This model allows to include behaviors like inertia and resonance.
- We refer to these systems as *latent force models* (LFMs).
- One way of thinking of our model is to consider puppetry.

Contenido

General Dynamical LFM

- Dynamical latent force model of order M

$$\sum_{m=0}^{M} \mathcal{D}^m[\mathbf{Y}]\mathbf{A}_m = \mathbf{U}\mathbf{S}^\top + \widehat{\mathbf{E}},$$

where $\mathcal{D}^m[\mathbf{Y}]$ has elements $\mathcal{D}^m y_d(t) = \frac{\mathrm{d}^m y_d(t)}{\mathrm{d}t^m}$, and $\mathbf{A}_m$ is a diagonal matrix with elements $A_{m,d}$ that weight the contribution of $\mathcal{D}^m y_d(t)$.

- Each element in the expression above can be written as

$$\mathcal{D}_0^M y_d = \sum_{m=0}^{M} A_{m,d}\mathcal{D}^m y_d(t) = \sum_{q=1}^{Q} S_{d,q} u_q(t) + \hat{e}_d(t),$$

where we have introduced an operator $\mathcal{D}_0^M$ that is equivalent to applying the weighted sum of operators $\mathcal{D}^m$.

Green's functions

- The operator $\mathcal{D}_0^M$ is related to a so called *Green's function* $G_d(t,s)$ by

$$\mathcal{D}_0^M[G_d(t,s)] = \delta(t-s),$$

with s fixed.

- The solution for $y_d(t)$ can be written in terms of the Green's function like

$$y_d(t) = \sum_{q=1}^{Q} S_{d,q} f_d(t, u_q(t)) + w_d(t),$$

with

$$f_d(t, u_q(t)) = \int_{\mathcal{T}} G_d(t,\tau) u_q(\tau) \mathrm{d}\tau,$$

and $w_d(t)$ is a general stochastic process.

Covariance for the outputs

- We assume that the latent functions $\{u_q(t)\}_{q=1}^{Q}$ are independent.

- We also assume that each $u_q(t)$ follows a Gaussian process prior, this is, $u_q(t) \sim \mathcal{GP}(0, k_{u_q,u_q}(t,t'))$.

- Furthermore, the processes $\{w_d\}_{d=1}^{D}$ are also assumed independent.

- The covariance $\text{cov}[y_d(t), y_{d'}(t')]$ is the given as

$$\text{cov}[f_d(t), f_{d'}(t')] + \text{cov}[w_d(t), w_{d'}(t')]\delta_{d,d'},$$

  with $\text{cov}[f_d(t), f_{d'}(t')]$ equals to

$$\sum_{q=1}^{Q} S_{d,q} S_{d',q} \int_{\mathcal{T}} \int_{\mathcal{T}'} G_d(t-\tau) G_{d'}(t'-\tau') k_{u_q,u_q}(\tau,\tau') \mathrm{d}\tau' \mathrm{d}\tau.$$

Multidimensional inputs

- In dynamical latent force models the input variable is one-dimensional (time).
- For higher-dimensional inputs, $\mathbf{x} \in \mathbb{R}^p$, partial differential equations are used.
- Once the Green's function associated to the linear partial differential operator has been established, we employ similar equations to the ones shown before to compute covariances.
- The input t is replaced by a high-dimensional vector $\mathbf{x}$.

Contenido

Hyperparameter Learning

- Let $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}$ represents a set of inputs, and $\boldsymbol{\theta}$ represents the hyperparameters of the covariance function.

- The marginal likelihood for the outputs can be written as

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{f}} + \boldsymbol{\Sigma}),$$

  where $\mathbf{y} = \operatorname{vec} \mathbf{Y}$, $\mathbf{K}_{\mathbf{f},\mathbf{f}} \in \mathbb{R}^{ND \times ND}$ with each element given by $\operatorname{cov}[f_d(\mathbf{x}_n), f_{d'}(\mathbf{x}_{n'})]$ (Neil's talk on Tuesday and today).

- The matrix $\boldsymbol{\Sigma}$ represents the covariance associated with the independent processes $w_d(\mathbf{x})$.

- Hyperparameters are estimated by maximizing the logarithm of the marginal likelihood.

Predictive distribution

- Prediction for a set of test inputs $\mathbf{X}_*$ is done using standard Gaussian process regression techniques.

- The predictive distribution is given by

$$p(\mathbf{y}_*|\mathbf{y}, \mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}_*|\boldsymbol{\mu}_*, \mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*}),$$

  with

$$\begin{aligned} \boldsymbol{\mu}_* &= \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \Sigma\right)^{-1} \mathbf{y}, \\ \mathbf{K}_{\mathbf{y}_*,\mathbf{y}_*} &= \mathbf{K}_{\mathbf{f}_*,\mathbf{f}_*} - \mathbf{K}_{\mathbf{f}_*,\mathbf{f}} \left(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \Sigma\right)^{-1} \mathbf{K}_{\mathbf{f}_*,\mathbf{f}}^\top + \Sigma_*. \end{aligned}$$

Efficient approximations (I)

- Learning θ through marginal likelihood maximization involves the inversion of the matrix $\mathbf{K}_{\mathbf{f},\mathbf{f}} + \Sigma$.
- The inversion of this matrix scales as $\mathcal{O}(D^3N^3)$.
- Single output case ($D = 1$) (James' talk on Tuesday).
- Recently, Álvarez and Lawrence (2009) introduced an efficient approximation for the case $D > 1$.

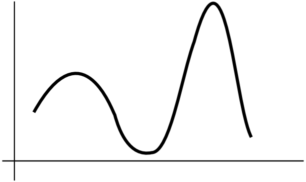
Efficient approximations (II)

- If only a few number $K < N$ of values of $u(\mathbf{x})$ are known, then the set of outputs $f_d(\mathbf{x}, u(\mathbf{x}))$ are uniquely determined.

- Similar to Partially Independent Training Conditional (PITC) approximation.

Efficient approximations (III)

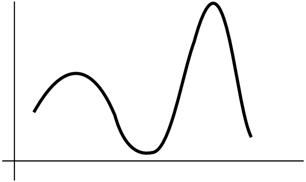
Sample from $p(u)$



$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x} - \mathbf{z}) u(\mathbf{z}) \mathrm{d}\mathbf{z}$$

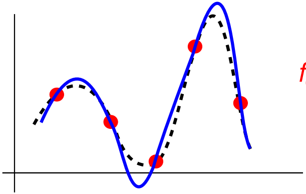
Efficient approximations (III)

Sample from $p(u)$



$$f_d(\mathbf{x}) = \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})u(\mathbf{z})\mathrm{d}\mathbf{z}$$

Sample from $p(u|\mathbf{u})$



$$f_d(\mathbf{x}) \approx \int_{\mathcal{X}} G_d(\mathbf{x}-\mathbf{z})\,\mathrm{E}\left[u(\mathbf{z})|\mathbf{u}\right]\mathrm{d}\mathbf{z}$$

Efficient approximations (IV)

- Another approximation (Álvarez et al., 2010) establishes a lower bound on the marginal likelihood and reduces computational complexity to $\mathcal{O}(DNK^2)$.

- Maximizing the lower bound with respect to $\phi(\mathbf{u})$

$$\mathcal{L}(\mathbf{Z}, \boldsymbol{\theta}) = \log \mathcal{N}\left(\mathbf{y}|\mathbf{0}, \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{K}_{\mathbf{u},\mathbf{f}} + \boldsymbol{\Sigma}\right) - \frac{1}{2}\operatorname{trace}\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{K}_{\mathbf{u},\mathbf{f}}\right)\right].$$

- Deterministic Training Conditional Variational (DTCVAR) approximation for multiple-output GP regression.

Contenido

Contenido

Second Order Dynamical System

Using the system of second order differential equations

$$M_d \frac{\mathrm{d}^2 y_d(t)}{\mathrm{d}t^2} + C_d \frac{\mathrm{d}y_d(t)}{\mathrm{d}t} + B_d y_d(t) = \sum_{q=1}^{Q} S_{d,q} u_q(t) + \widehat{e}_d(t),$$

where

$u_q(t)$	latent forces
$y_d(t)$	displacements over time
C_d	damper constant for the d-th output
B_d	spring constant for the d-th output
M_d	mass constant for the d-th output
$S_{d,q}$	sensitivity of the d-th output to the q-th input.

Second Order Dynamical System: solution

Solving for $y_d(t)$, we obtain

$$y_d(t) = \sum_{q=1}^{Q} S_{d,q} f_d(t, u_q(t)) + w_d(t),$$

where the linear operator is given by a convolution:

$$f_d(t, u_q(t)) = \int_0^t \underbrace{\frac{1}{\omega_d} \exp(-\alpha_d(t-\tau)) \sin(\omega_d(t-\tau))}_{G_d(t-\tau)} u_q(\tau) \mathrm{d}\tau,$$

with $\omega_d = \sqrt{4B_d - C_d^2}/2$ and $\alpha_d = C_d/2$.

Second Order Dynamical System: covariance matrix

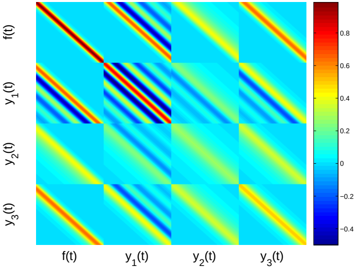
Behaviour of the system summarized by the damping ratio:

$$\zeta_d = \frac{1}{2} C_d / \sqrt{B_d}$$

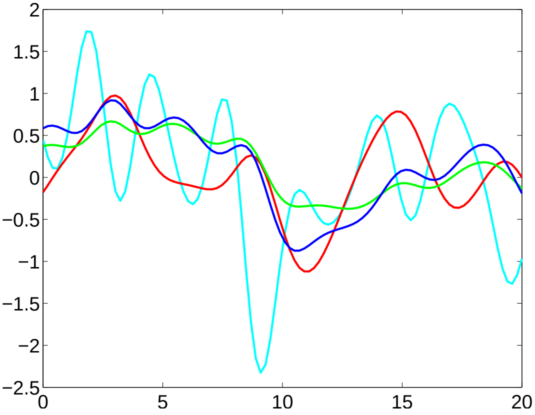
$\zeta_d > 1$ overdamped system
$\zeta_d = 1$ critically damped system
$\zeta_d < 1$ underdamped system
$\zeta_d = 0$ undamped system (no friction)

Example covariance matrix:

$\zeta_1 = 0.125$ underdamped
$\zeta_2 = 2$ overdamped
$\zeta_3 = 1$ critically damped

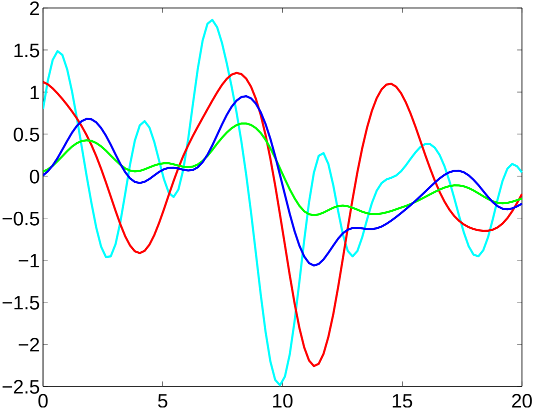


Second Order Dynamical System: samples from GP



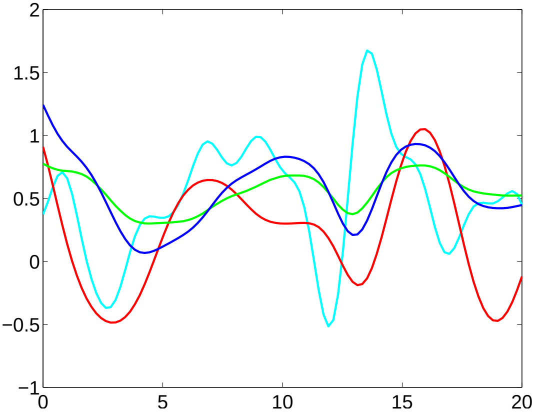
Joint samples from the ODE covariance, *cyan*: $u(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP



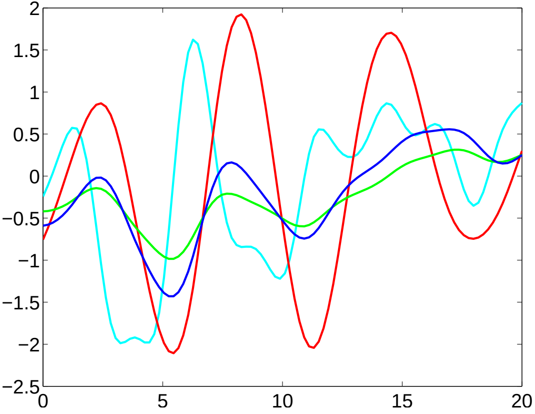
Joint samples from the ODE covariance, *cyan*: $u(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP



Joint samples from the ODE covariance, *cyan*: $u(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Second Order Dynamical System: samples from GP



Joint samples from the ODE covariance, *cyan*: $u(t)$, *red*: $y_1(t)$(underdamped) and *green*: $y_2(t)$ (overdamped) and *blue*: $y_3(t)$ (critically damped).

Motion Capture Data (1)

- CMU motion capture data, motions 18, 19 and 20 from subject 49.
- Motions 18 and 19 for training and 20 for testing.

Motion Capture Data (2)

- The data down-sampled by 32 (from 120 frames per second to 3.75).

- We focused on the subject's left arm.

- For testing, we condition only on the observations of the shoulder's orientation (motion 20) to make predictions for the rest of the arm's angles.

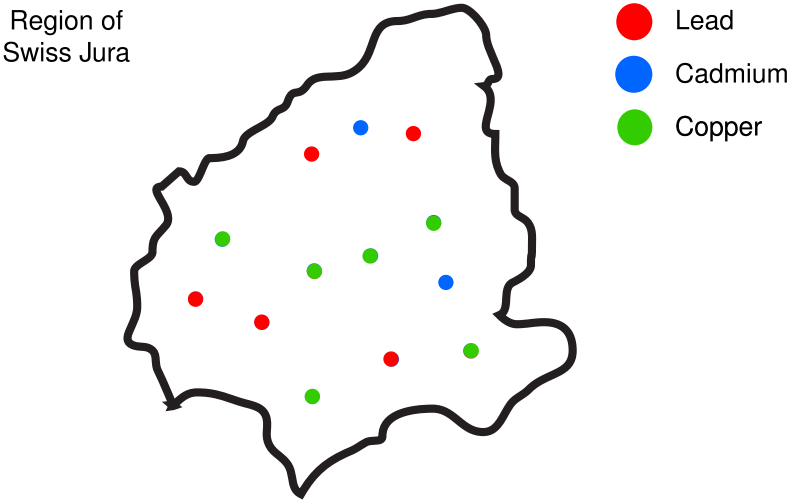
Motion Capture Results

Root mean squared (RMS) angle error for prediction of the left arm's configuration in the motion capture data. Prediction with the latent force model outperforms the prediction with regression for all apart from the radius's angle.

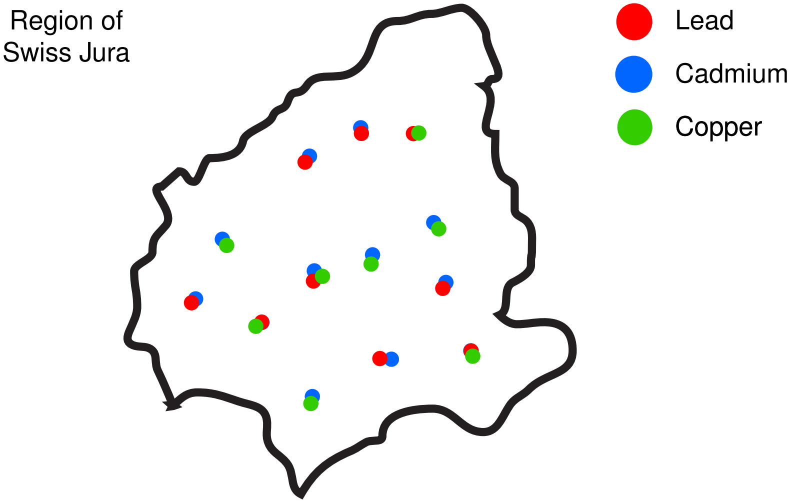
Angle	Latent Force Error	Regression Error
Radius	4.11	**4.02**
Wrist	**6.55**	6.65
Hand X rotation	**1.82**	3.21
Hand Z rotation	**2.76**	6.14
Thumb X rotation	**1.77**	3.10
Thumb Z rotation	**2.73**	6.09

Contenido

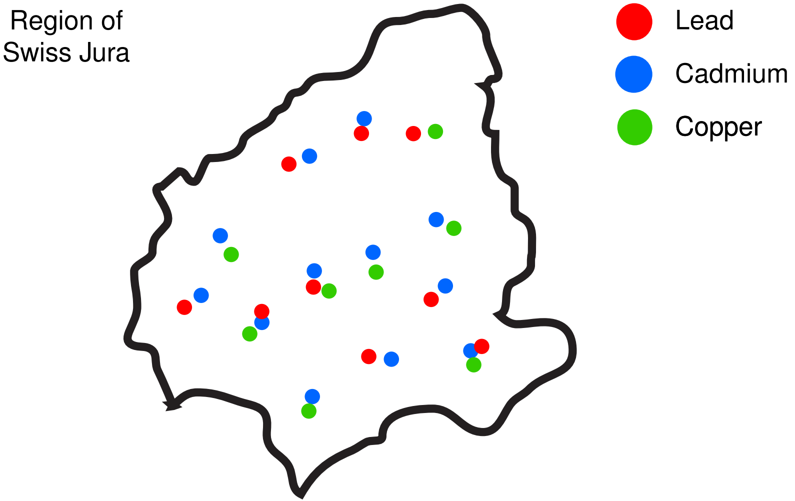
Diffussion in the Swiss Jura



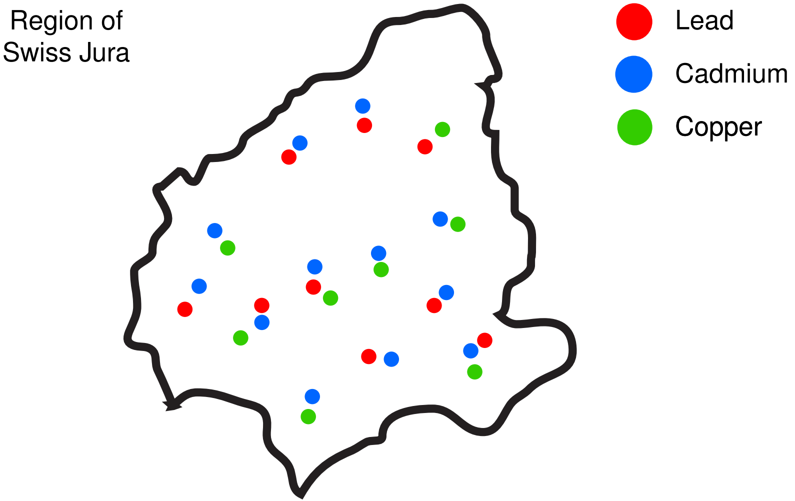
Diffussion in the Swiss Jura



Diffussion in the Swiss Jura



Diffussion in the Swiss Jura



Diffusion equation

- A simplified version of the diffusion equation is

$$\frac{\partial f_d(\mathbf{x},t)}{\partial t} = \sum_{j=1}^{p} \kappa_d \frac{\partial^2 f_d(\mathbf{x},t)}{\partial x_j^2},$$

  where $f_d(\mathbf{x},t)$ are the concentrations of each pollutant.

- The solution to the system is then given by

$$f_d(\mathbf{x},t) = \sum_{q=1}^{Q} S_{d,q} \int_{\mathbb{R}^p} G_d(\mathbf{x},\mathbf{x}',t) u_q(\mathbf{x}')\mathrm{d}\mathbf{x}',$$

  where $u_q(\mathbf{x})$ represents the concentration of pollutants at time zero and $G_d(\mathbf{x},\mathbf{x}',t)$ is the Green's function given as

$$G_d(\mathbf{x},\mathbf{x}',t) = \frac{1}{2^p \pi^{p/2} T_d^{p/2}} \exp\left[-\sum_{j=1}^{p} \frac{(x_j - x_j')^2}{4T_d}\right],$$

  with $T_d = \kappa_d t$.

Prediction of Metal Concentrations

- Prediction of a *primary variable* by conditioning on the values of some *secondary variables*.

Primary variable	**Secondary Variables**
Cd	Ni, Zn
Cu	Pb, Ni, Zn
Pb	Cu, Ni, Zn
Co	Ni, Zn

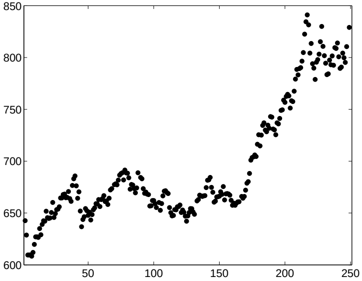
- Comparison bewteen diffusion kernel, independent GPs and "ordinary co-kriging".

Metals	IGPs	GPDK	OCK
Cd	0.5823±0.0133	**0.4505±0.0126**	0.5
Cu	15.9357±0.0907	**7.1677±0.2266**	7.8
Pb	22.9141±0.6076	**10.1097±0.2842**	10.7
Co	2.0735±0.1070	1.7546±0.0895	**1.5**

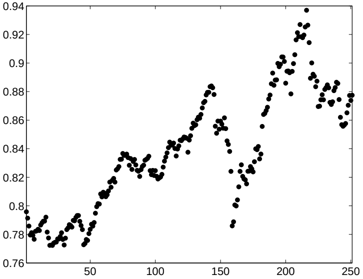
Contenido

A dynamic model for financial data (I)

Multivariate financial data set: the dollar prices of the 3 precious metals and top 10 currencies.



Gold



AUD

A dynamic model for financial data (I)

- Our model: a set of coupled differential equations, driven by either a Gaussian process, a white noise process or both,

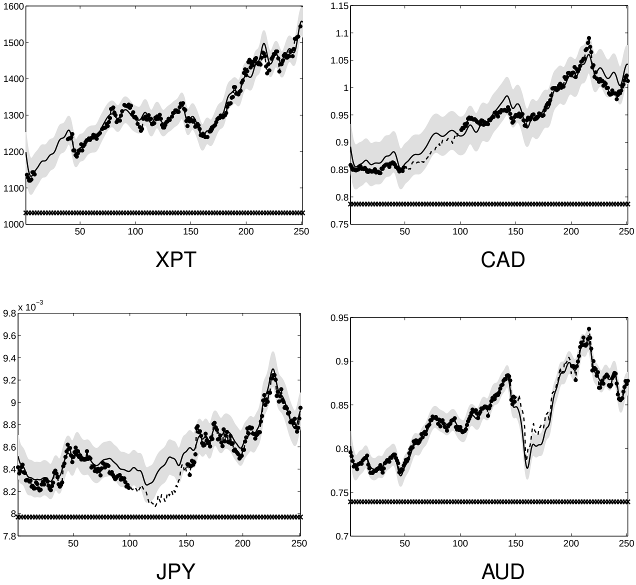
$$\frac{\mathrm{d}f_d(t)}{\mathrm{d}t} = \lambda_d f_d(t) + S_d u(t),$$

  where λ_d is a decay coefficient and S_d quantifies the influence of the process $u(t)$.

- If $u(t)$ is a white noise process $\rightarrow$ Langevin equation $\rightarrow$ a linear stochastic differential equation.

- Solution for $f_d(t)$ has the form of convolutions. For a single output and white noise process, $f_d(t) \rightarrow$ Ornstein-Uhlenbeck (OU) process.

A dynamic model for financial data (III)



Contenido

Contenido

Non-linear and Cascaded LFMs

- Non-linear LFMs
  - If the likelihood function is not Gaussian or the differential equation is nonlinear, approximations are needed.
  - Approximations used before include the Laplace's approximation (Lawrence et al., 2007) or sampling (Titsias et al., 2009).
- Cascaded Latent Force Models
  - Latent forces $u_q(t)$ could be the outputs of another latent force model.
  - For example, in Honkela et al. (2010), the authors use a cascaded system to describe gene expression data

Contenido

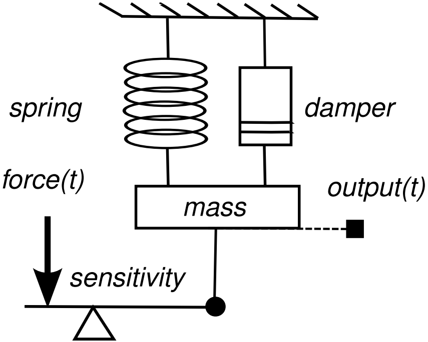
Recap

- Latent force models encode the interaction between multiple related dynamical systems in the form of a covariance function.
- Each variable to be modeled is represented as the output of a differential equation.
- Differential equations are driven by a weighted sum of latent functions with uncertainty given by a Gaussian process priors.

Discontinuous latent forces

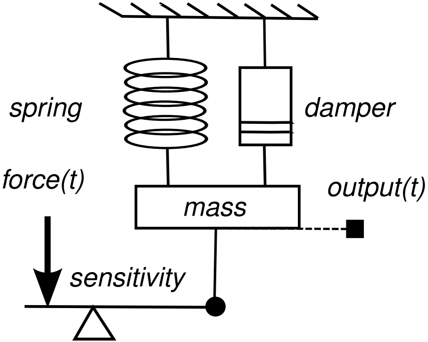
- If a single Gaussian process prior is used to represent each latent function then the models we consider are limited to smooth driving functions.
- However, discontinuities and segmented latent forces are omnipresent in real-world data.
- Impact forces due to contacts in a mechanical dynamical system or a switch in an electrical circuit.
- Motor primitives: most non-rhythmic natural motor skills consist of a sequence of segmented, discrete movements.

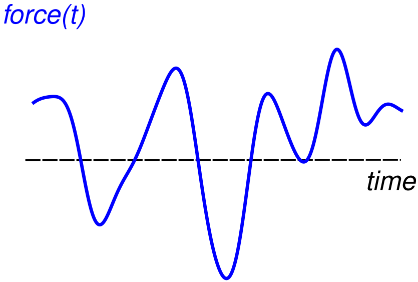
LFM



force(t) $\sim \mathcal{GP}$(0, k(t,t'))

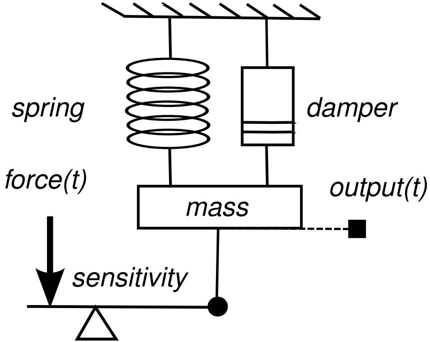
LFM



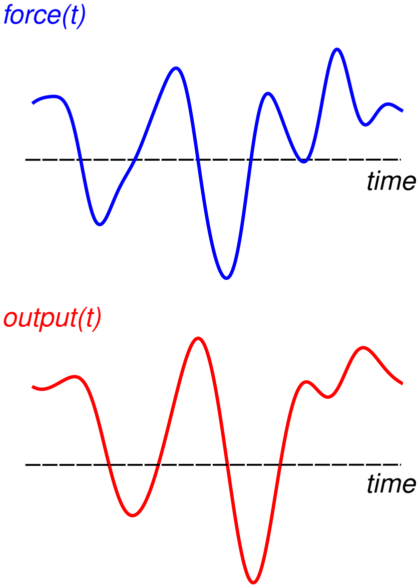


force(t) $\sim \mathcal{GP}(0, k(t,t'))$

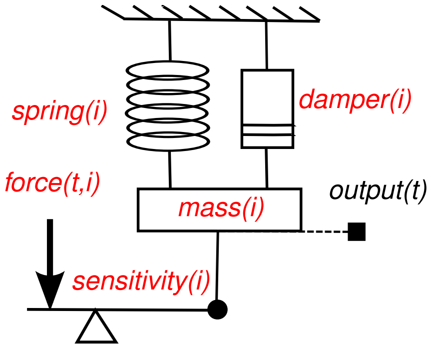
LFM



force(t) $\sim \mathcal{GP}(0, k(t,t'))$

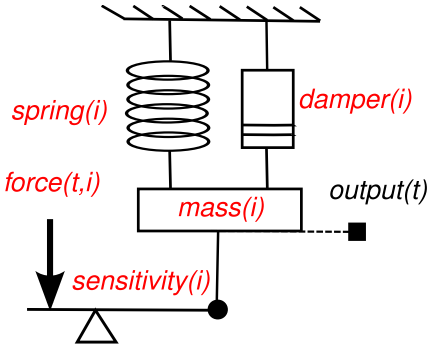


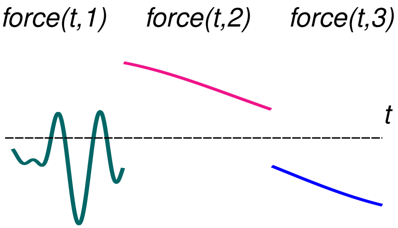
Switched LFM



$force(t,i) \sim \mathcal{GP}(0, k(t,t',i))$

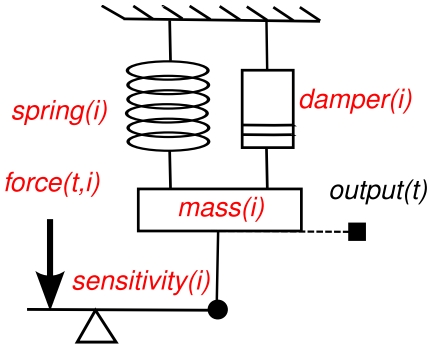
Switched LFM

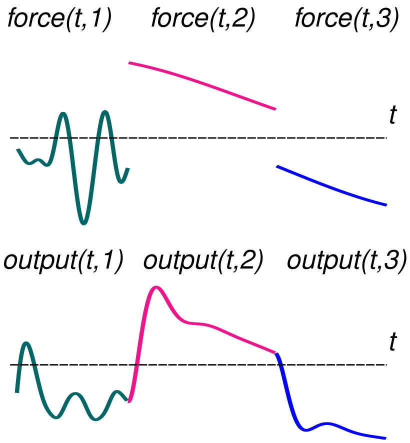




force(t,i) $\sim \mathcal{GP}$(0, k(t,t',i))

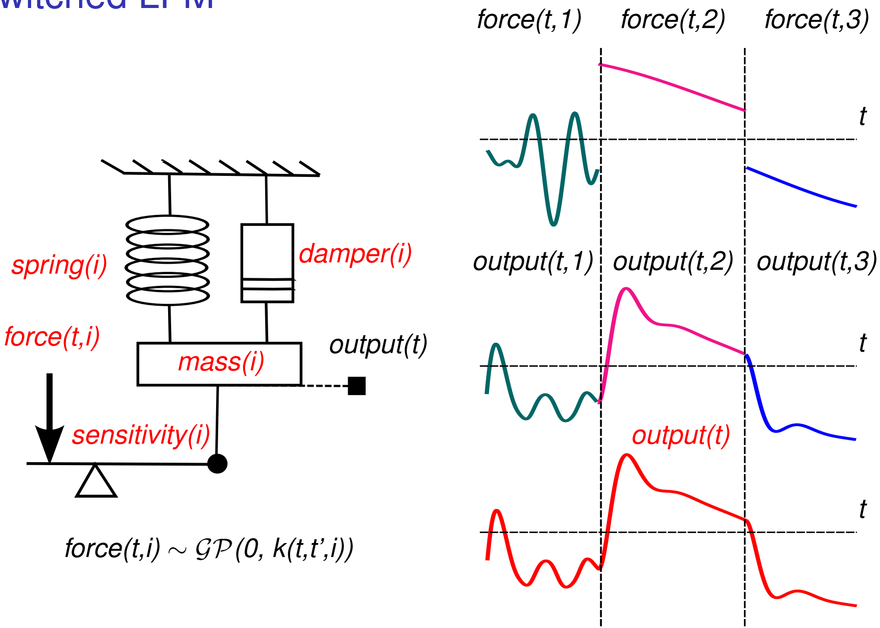
Switched LFM





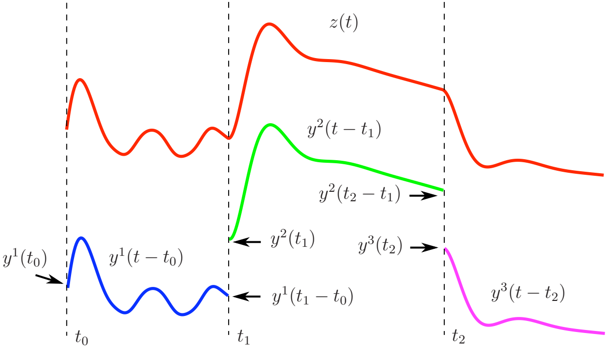
force(t,i) $\sim \mathcal{GP}$(0, k(t,t',i))

Switched LFM



$force(t,i) \sim \mathcal{GP}(0, k(t,t',i))$

Continuous in the outputs

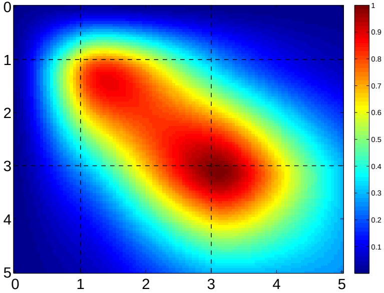


$$z_d(t) = c_d^i(t-t_{i-1})y_d^i(t_{i-1}) + e_d^i(t-t_{i-1})\dot{y}_d^i(t_{i-1}) + S_{d,i}f_d^i(t-t_{i-1}, u_{i-1}),$$

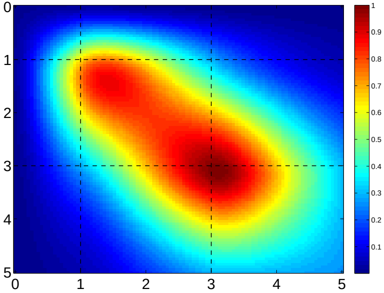
where

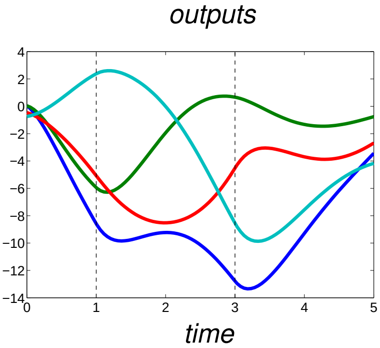
$$f_d^i(t, u_{i-1}) = \int_0^t \frac{1}{\omega_d} e^{-\alpha_d(t-\tau)} \sin[(t-\tau)\omega_d] u_{i-1}(\tau) \mathrm{d}\tau.$$

Covariance and Samples

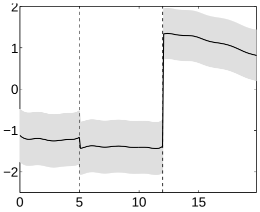


Covariance and Samples

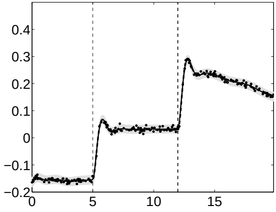




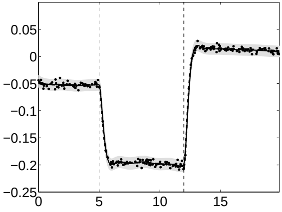
Toy examples



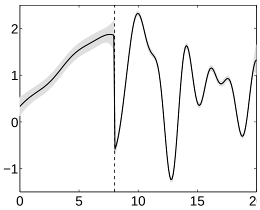
LF toy example 1.



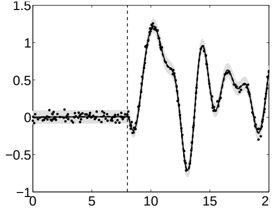
Output 1 toy example 1.



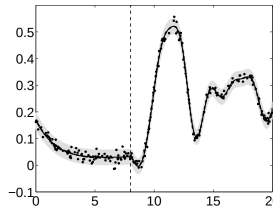
Output 2 toy example 1.



LF toy example 2.



Output 1 toy example 2.

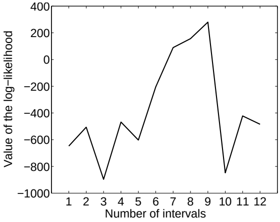


Output 2 toy example 2.

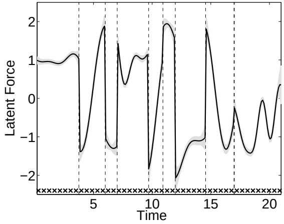
Segmentation of human movement (I)

- The task is to segment discrete movements related to motor primitives.
- Data collection was performed using a Barrett WAM robot as haptic input device, with 7 DOF.

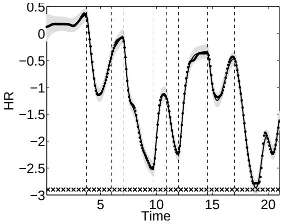
Segmentation of human movement (II)



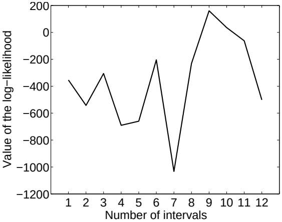
Log-Likelihood Try 1.



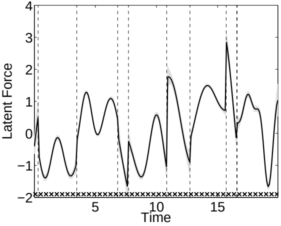
Latent force Try 1.



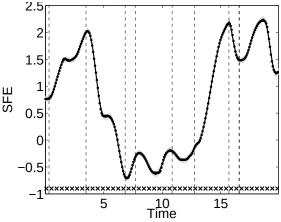
HR Output Try 1.



Log-Likelihood Try 2.



Latent force Try 2.



SFE Output Try 2.

References I

Álvarez, M. A., Lawrence, N. D., 2009. Sparse convolved Gaussian processes for multi-output regression. In: Koller et al. (2009), Vol. 21, pp. 57–64.

Álvarez, M. A., Luengo, D., Titsias, M. K., Lawrence, N. D., 13-15 May 2010. Efficient multioutput Gaussian processes through variational inducing kernels. In: Teh, Y. W., Titterington, M. (Eds.), Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR W&CP 9, Chia Laguna, Sardinia, Italy, pp. 25–32.

Honkela, A., Girardot, C., Gustafson, E. H., Liu, Y.-H., Furlong, E. E. M., Lawrence, N. D., Rattray, M., 2010. Model-based method for transcription factor target identification with limited data. Proc. Natl. Acad. Sci. 107 (17), 7793–7798.

Koller, D., Schuurmans, D., Bengio, Y., Bottou, L. (Eds.), 2009. NIPS. Vol. 21. MIT Press, Cambridge, MA.

Lawrence, N. D., Sanguinetti, G., Rattray, M., 2007. Modelling transcriptional regulation using Gaussian processes. In: Schölkopf, B., Platt, J. C., Hofmann, T. (Eds.), NIPS. Vol. 19. MIT Press, Cambridge, MA, pp. 785–792.

Titsias, M., Lawrence, N. D., Rattray, M., 2009. Efficient sampling for Gaussian process inference using control variables. In: Koller et al. (2009), Vol. 21, pp. 1681–1688.

Titsias, M. K., 16-18 April 2009. Variational learning of inducing variables in sparse Gaussian processes. In: van Dyk, D., Welling, M. (Eds.), Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics. JMLR W&CP 5, Clearwater Beach, Florida, pp. 567–574.