How to choose the covariance for Gaussian process regression independently of the basis

Workshop *Gaussian Processes in Practice*

Matthias O. Franz and Peter V. Gehler

June 12th, 2006
Motivation: Nonlinear system identification using Volterra series

Characterisation of a nonlinear system $y(t) = T[x(t)]$ by a series expansion $y(t) = \sum_n H_n[x(t)]$ (Volterra, 1887):

$$
y(t) = h^{(0)} + \int_{\mathbb{R}} h^{(1)}(\tau_1)x(t - \tau_1) \, d\tau_1 
+ \int_{\mathbb{R}^2} h^{(2)}(\tau_1, \tau_2)x(t - \tau_1)x(t - \tau_2) \, d\tau_1 d\tau_2 
+ \int_{\mathbb{R}^3} h^{(3)}(\tau_1, \tau_2, \tau_3)x(t - \tau_1)x(t - \tau_2)x(t - \tau_3) \, d\tau_1 d\tau_2 d\tau_2 
+ \cdots
$$

Discretised form for $x = (x_1, \ldots, x_m)^\top \in \mathbb{R}^m$

$$
H_n[x] = \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} h^{(n)}_{i_1 \ldots i_n} x_{i_1} \cdots x_{i_n}.
$$
Polynomial regression and Volterra systems

Volterra expansions can be efficiently estimated by a regression in polynomial kernel functions (Franz & Schölkopf, 2006)

\[ k_{ihp}(x, x') = (1 + x^\top x')^p \]

⇒ GP framework is applicable for the estimation of Volterra systems.

Problems:

- Polynomial covariance implies strong correlation of distant inputs. In real-world problems, the reverse situation is more common.
- Typically, polynomial regression shows inferior performance than localized covariance functions.

⇒ Independent choice of covariance and basis
Decoupling of basis and covariance

- Basic idea: approximate a desired covariance function $k_{GP}(x_i, x_j)$ on a finite set $S = \{x_1, \ldots, x_p\}$ of input points.
- Weight-space view of a GP: $k(x_i, x_j) = \phi(x_i)^\top \Sigma_w \phi(x_j)$.
- Choose basis $\phi(x)$ and prior $\Sigma_w$ such that
  \[
  k_{GP}(x_i, x_j) = \phi(x_i)^\top \Sigma_w \phi(x_j) \quad \forall x_i, x_j \in S.
  \]
- Basis: Kernel PCA map $\phi(x) = K^{-\frac{1}{2}}(k(x, x_1), \ldots, k(x, x_n))^\top$, solve system of linear equations in $\Sigma_w$.

⇒

- Arbitrary covariances can be approximated.
- Performance of polynomial regression can be significantly improved.