Maximum Entropy Property of Discrete-time Stable Spline Kernel

Tohid Ardeshiri and Tianshi Chen (tohid@isy.liu.se)

Contributions

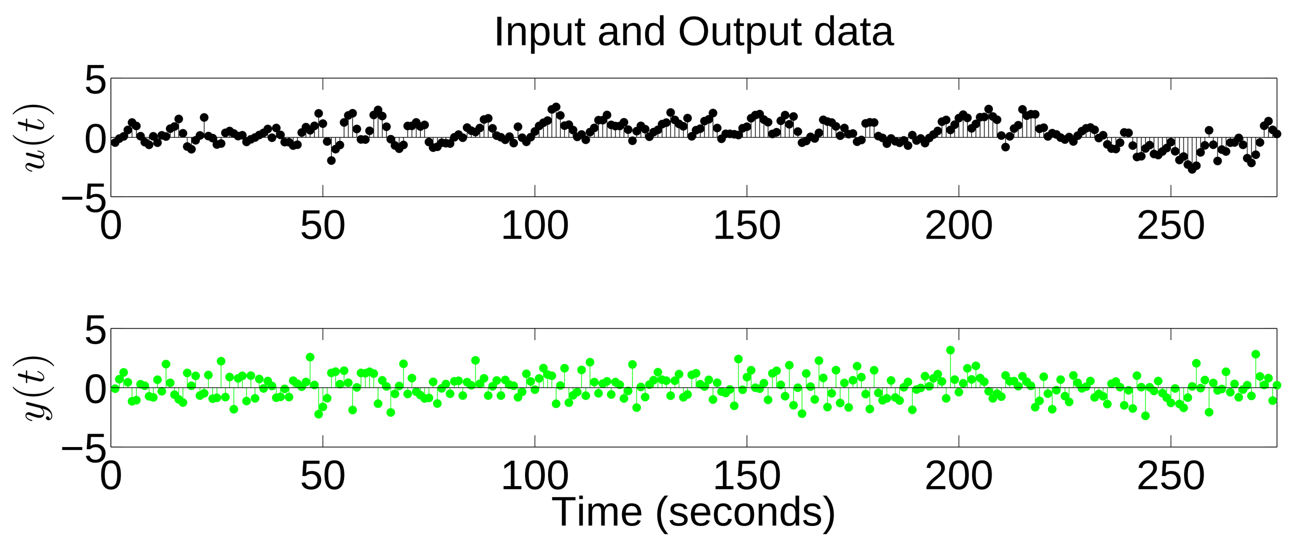
We provide a simple and self-contained proof to show the maximum entropy property of the *Discrete-time First-Order Stable Spline Kernel*. The advantages of working in discrete-time domain include

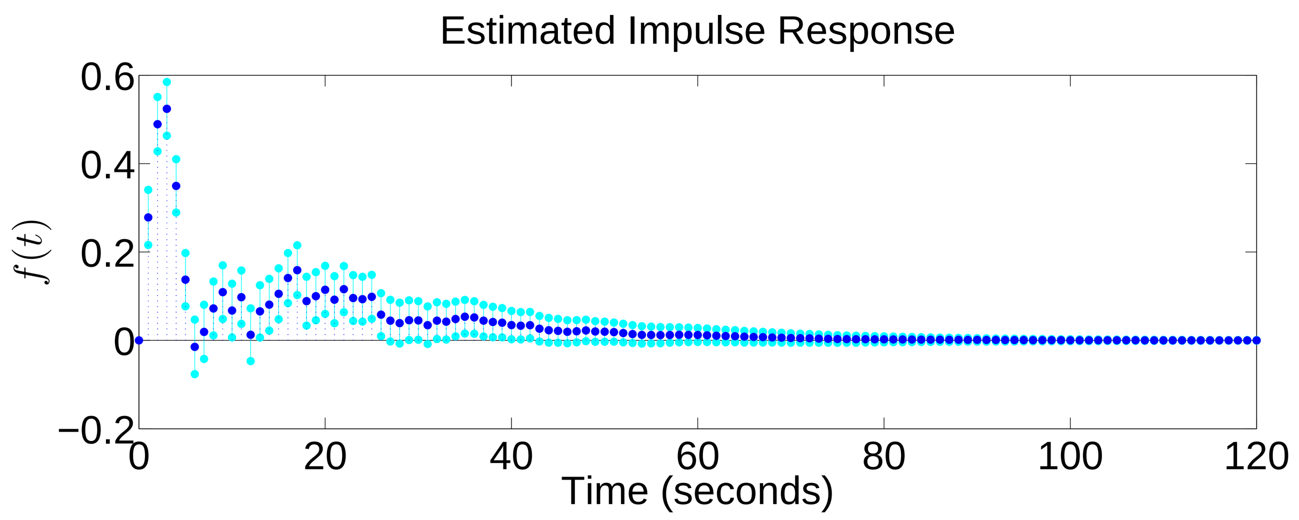
1. The differential entropy rate is well-defined for discrete-time stochastic process.
2. Given a stochastic process, its finite difference process can be well-defined in discrete-time domain.
3. It is possible to show what maximum entropy property a zero-mean discrete-time Gaussian process with following covariance function has.

$$k(t,s) = \min\{e^{-\beta t}, e^{-\beta s}\}$$

Also, we define the discrete-time Wiener process and prove its maximum entropy property.

Impulse Response Identification





$$y(t_i) = f * u(t_i) + v(t_i), \quad i = 0, 1, \cdots, N \tag{1}$$

$$f(t) \sim \mathrm{GP}(m(t), k(t,s)), \tag{2}$$

where $m(t)$ is the mean function and is often set to be zero, and $k(t,s)$ is the covariance function, also called the kernel function.

Continuous-time Approach [2]

BIBO stability: For a *continuous time linear time invariant (LTI) system*, the condition for *BIBO stability* is that the impulse response be absolutely integrable, i.e., its L^1 norm exists.

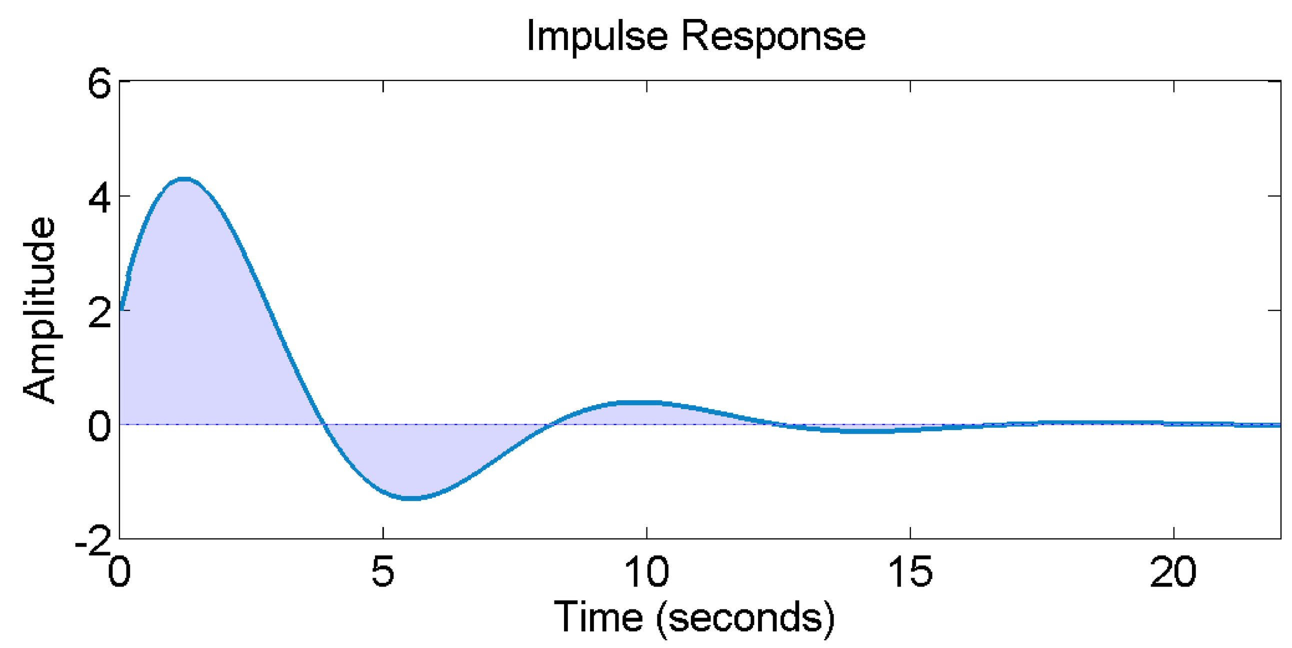
$$\int_{-\infty}^{\infty} |f(t)| \, \mathrm{d}t = \|f\|_1 < \infty \tag{3}$$

Smoothness: The smoothness constraint on the *continuous time* impulse responses is addressed by using [1, Theorem 1] which suggests that the smoothness of a signal can be imposed by assuming that the variances of these derivatives are finite.

$$\mathrm{Var}\left[\frac{\mathrm{d}f}{\mathrm{d}t}\right] < \infty \tag{4}$$

Entropy rate: The differential entropy rate of a real-valued *continuous-time* stochastic process $f(\cdot)$ is defined in [1] as

$$\overline{H}(f) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\big(S(\omega)\big)\, \mathrm{d}\omega. \tag{5}$$



Maximum Entropy Rate Prior

Let Λ_B be the class of the zero-mean stationary and differentiable Gaussian processes on $[0,1]$ with bandlimited spectrum, i.e. $S(\omega) = 0$ for $|\omega| \le B$

Proposition 1. *[2, Proposition 2] Let f be a stochastic process on $\mathbb{R}^+$ such that $f(-\log(t)/\beta) = g(t)$, where $g \in \Lambda_B$ with the variance of $g^{(1)}$ finite. Then, as the bandwidth B goes to ∞, the kernel of f induced by the maximum entropy prior for g, conditional on $\lim_{t\to\infty} f(t) = 0$, is $k(s,t) := \mathbb{E}[f(s).f(t)] = \min\{e^{-\beta t}, e^{-\beta s}\}$*

Discrete-time Approach

BIBO stability: For a *discrete-time LTI system*, the condition for *BIBO stability* is that the impulse response be absolutely summable, i.e., its ℓ^1 norm exists.

$$\sum_{n=-\infty}^{\infty} |f[n]| = \|f\|_1 < \infty \tag{6}$$

Smoothness: The smoothness constraint on the *discrete-time* impulse responses can be imposed by assuming that the variances of finite difference is finite.

$$\mathrm{Var}\left[f(t_{i+1}) - f(t_i)\right] = \lambda(t_{i+1} - t_i), \quad \infty > \lambda > 0 \tag{7}$$

Entropy rate: The differential entropy rate of a real-valued *discrete-time* stochastic process $\{f(t_i) : \ f(t_i) \in \mathbb{R}, \ t_i \in \mathcal{T}\}$ is defined as

$$\overline{H}(f) = \lim_{n\to\infty} \frac{1}{n} H(f(t_1), f(t_2), ..., f(t_n)) \tag{8}$$

if the limit exists and where the differential entropy of a continuous *random variable* X with density $p(x)$ is defined as

$$H(X) = -\int_{\mathcal{S}} p(x) \log p(x) \, \mathrm{d}x, \tag{9}$$

where, $\mathcal{S}$ is the support set of the random variable.

The Main result

Proposition 2. *Let $g(\tau)$ denote a zero-mean discrete-time stochastic process defined on an ordered index set $\{\tau_i | \tau_0 = 0, \tau_\infty = 1, 0 < \tau_i < \tau_j < 1, \ 0 < i < j < \infty\}$. Now consider a finite segment of g with index set $\mathcal{T}_g = \{\tau_i | \tau_0 = 0, \tau_n < 1, 0 < \tau_i < \tau_j < t_n, \ 0 < i < j < n\}$. Then for any $n \in \mathbb{N}$, the zero-mean Gaussian process with covariance function $k(t,s) = \min\{e^{-\beta t}, e^{-\beta s}\}$ is the solution to the maximum differential entropy problem:*

$$\begin{aligned} \underset{f}{\text{maximize}} \quad & H(f(t_0), \cdots, f(t_{n-1})) \\ \text{subject to} \quad & f(t) = g(e^{-\beta t}), \ \ \beta > 0, t \in \overline{\mathcal{T}_g}, \\ & g(\tau_0) = 0, \\ & \mathbb{E}\left[g(\tau)\right] = 0, \\ & \mathrm{Var}\left[g(\tau_{i+1}) - g(\tau_i)\right] = \lambda(\tau_{i+1} - \tau_i), i = 0, 1, \cdots, n-1 \end{aligned}$$

References

[1] G. De Nicolao, G. Ferrari-Trecate, and A. Lecchini. MAXENT priors for stochastic filtering problems. In *Mathematical Theory of Networks and Systems*, Padova, Italy, July 1998.

[2] G. Pillonetto and G. De Nicolao. Kernel selection in linear system identification part i: A Gaussian process perspective. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 4318–4325, Dec 2011.