RUHR-UNIVERSITÄT BOCHUM

INSTITUT FÜR NEUROINFORMATIK

RUB

Predictable Feature Analysis (PFA)

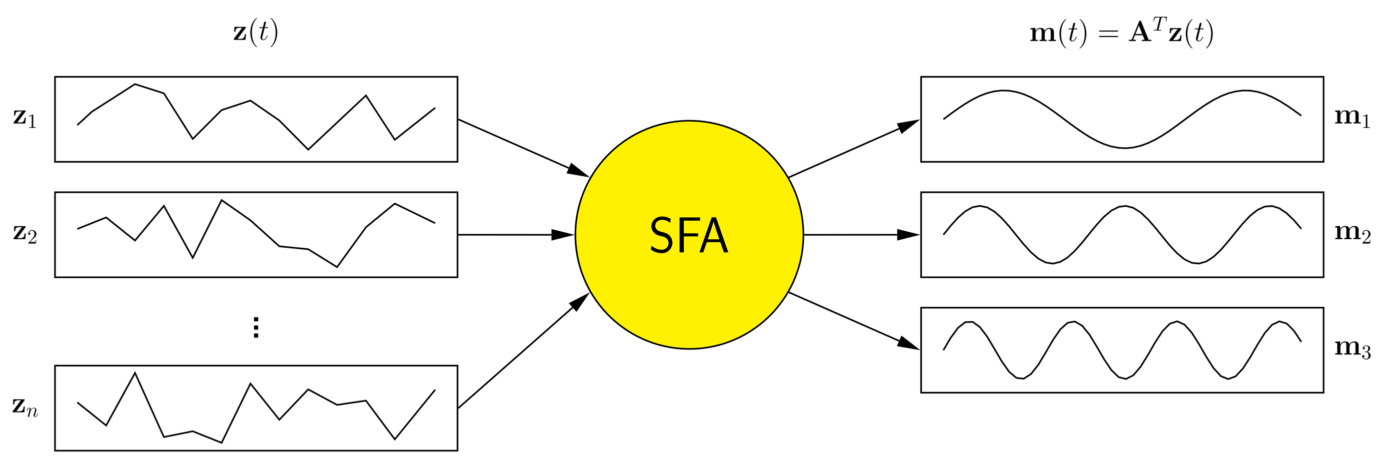
Stefan Richthofer, Björn Weghenkel, Laurenz Wiskott

stefan.richthofer@ini.rub.de bjoern.weghenkel@ini.rub.de laurenz.wiskott@ini.rub.de

Inspired by Slow Feature Analysis (SFA), PFA aims to extract features by predictability rather than slowness.
This way, we prepare to handle interactive scenarios, which involve notions of control, planning and decision-making. In order to perform any kind of planning or intelligent control, it is crucial to have a model that is capable of estimating the consequences of possible actions. Since predictability is a desired property of such a model by definition, the PFA-setup is well suited for this purpose. While in control theory, the involved models are usually formulated as a set of (partial) differential equations in a problem specific manner, with PFA we hope to preserve the main advantages of SFA - namely its unsupervised nature and the ability to build a model in a self-organizing fashion.

Slow Feature Analysis

Slow Feature Analysis is an algorithm that has proven valuable in several fields and problems concerning signal- and data-analysis. The idea is that a drastic, yet reasonable dimensionality reduction can be obtained by focusing on slowly varying sub-signals, the so-called “slow features”. Typical data-analysis and recognition tasks, such as regression and classification, become much more feasible on the reduced signal and can be applied afterwards.
SFA has been successfully applied to perform tasks like the self-organization of complex-cell receptive fields, the recognition of whole objects invariant to spatial transformations, the self-organization of place-cells, extraction of driving forces, or nonlinear blind source separation.

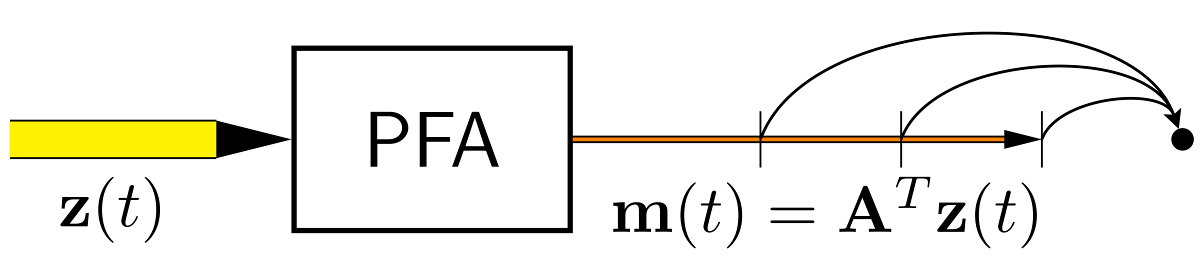


$$\begin{aligned} &\underset{\mathbf{a}_i \in \mathbb{R}^{d_z}}{\text{minimize}} && \mathbf{a}_i^T \left\langle \dot{\mathbf{z}}\dot{\mathbf{z}}^T \right\rangle_t \mathbf{a}_i && \\ &\text{subject to} && \mathbf{a}_i^T \left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t \mathbf{a}_i = 1 && \text{(unit variance)} \\ & && \mathbf{a}_i^T \left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t \mathbf{a}_j = 0 \quad \forall\, j < i && \text{(uncorrelated)} \end{aligned}$$

Predictable Feature Analysis

While there exist model-independent notions from information theory (cf. information bottleneck approach), we consider predictability with respect to a certain prediction model. In the current setup, we consider aspects as predictable if they can be predicted by a linear autoregressive model after an optional, non-linear preprocessing. So the desired property of the extracted model-function $\mathbf{m}$ can be expressed as follows:

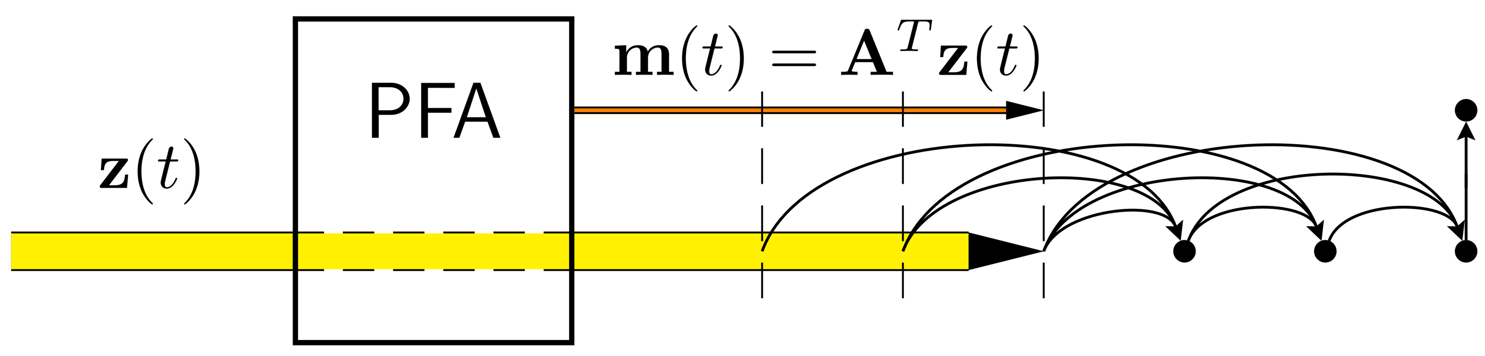
$$\mathbf{m}_i(t) \quad := \quad \mathbf{a}_i^T \mathbf{z}(t) \quad \overset{!}{\approx} \quad b_1 \mathbf{a}_i^T \mathbf{z}(t-1) + \ldots + b_p \mathbf{a}_i^T \mathbf{z}(t-p)$$



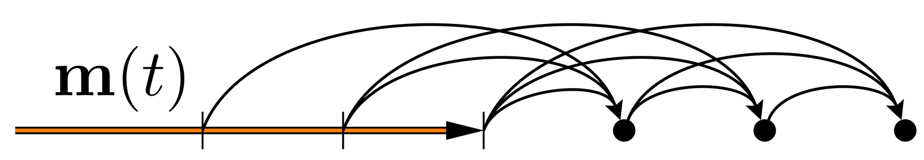
So for each i, we face the problem of finding optimal vectors $\mathbf{a}_i$ and $\mathbf{b}$. We formalize this in a least squares sense and include the constrains from SFA to avoid trivial or repeated solutions:

$$\begin{aligned} &\underset{\mathbf{a}_i \in \mathbb{R}^{d_z},\, \mathbf{b} \in \mathbb{R}^p}{\text{minimize}} && \left\langle \left(\mathbf{a}_i^T \mathbf{z}(t) - \sum_{i=1}^{p} b_i \mathbf{a}_i^T \mathbf{z}(t-i) \right)^2 \right\rangle_t && \\ &\text{subject to} && \mathbf{a}_i^T \left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t \mathbf{a}_i = 1 && \text{(unit variance)} \\ & && \mathbf{a}_i^T \left\langle \mathbf{z}\mathbf{z}^T \right\rangle_t \mathbf{a}_j = 0 \quad \forall\, j < i && \text{(uncorrelated)} \end{aligned}$$

The extracted features must be optimized for predictability, but judging their predictability is an optimization problem by itself: If $\mathbf{a}$ or $\mathbf{b}$ is fixed, the other one can be obtained efficiently, but with a different algorithm.
However, optimizing in turns runs into local optima, which depend on the starting point. The following sketch illustrates a relaxation which allows efficient computation of globally optimal $\mathbf{a}$ and $\mathbf{b}$:

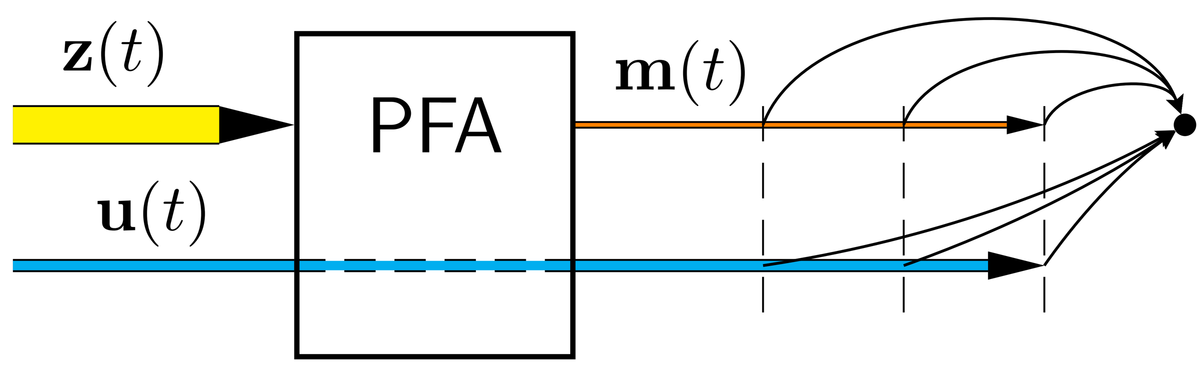


We extract those signals that have the lowest error after an iterated linear extrapolation of the entire input signal. As a side effect, $\mathbf{m}$ comes out as a signal which can be predicted independently from $\mathbf{z}$ with a low error:



PFA using additional information

A variant of PFA allows to optimize predictability regarding to an additional stream of information.



This way, we attempt to model problems that involve analysis of the relationship between two signals - typically representing control and sensor components - action and outcome.

Future work

The next step is to apply PFA to more realistic data. For this, we have several scenarios in mind, including:

- let an agent explore an environment, learning the relation between his movement commands and the resulting position
- use PFA in a car simulator, learning how steering wheel movements influence the position in the lane
- analyze the relation between control commands and end effector movement of a robot arm

The authors acknowledge support from the German Federal Ministry of Education and Research within the National Network Computational Neuroscience - Bernstein Fokus: "Learning behavioral models: From human experiment to technical assistance", grant FKZ 01GQ0951.