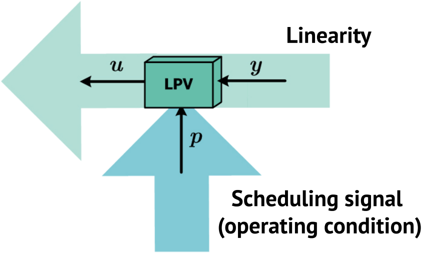
Canonical Correlation Analysis Based Identification of LPV Systems

Roland Tóth[1]

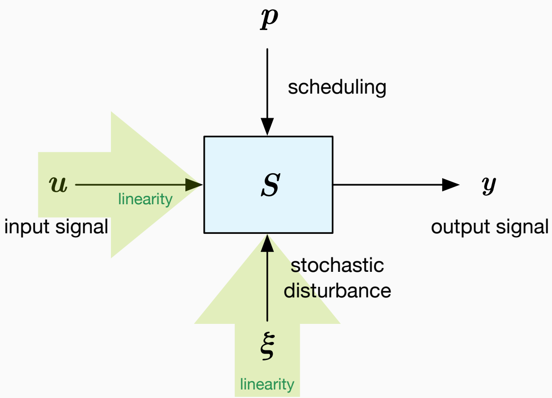
Based on work with S.Z. Rizvi and P.B. Cox

ERNSI Workshop, Lyon, France
September 25, 2017

[1]Eindhoven University of Technology, The Netherlands



Concept of a Linear Parameter-Varying (LPV) system



Data-generating LPV system (innovation noise)

$$x_{k+1} = A(p_k)x_k + B(p_k)u_k + K(p_k)e_k,$$
$$y_k = C(p_k)x_k + D(p_k)u_k + e_k,$$

(jointly minimal State-Space (SS) representation)

$$M(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) & K(\cdot) \\ C(\cdot) & D(\cdot) & I \end{bmatrix} : \mathbb{P} \to \mathbb{R}^{(n_\mathrm{x}+n_\mathrm{y})\times(n_\mathrm{x}+n_\mathrm{u}+n_\mathrm{y})}$$ smooth matrix function

Input	Output	State	Scheduling
$u : \mathbb{N} \to \mathcal{U} \subseteq \mathbb{R}^{n_\mathrm{u}}$	$y : \mathbb{N} \to \mathcal{Y} \subseteq \mathbb{R}^{n_\mathrm{y}}$	$x : \mathbb{N} \to \mathcal{X} \subseteq \mathbb{R}^{n_\mathrm{x}}$	$p : \mathbb{N} \to \mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$

$\mathbb{P}$ compact set; e ind. white noise with $e_k \sim \mathcal{N}(0, \Sigma_\mathrm{e})$.

Data-generating LPV system (innovation noise)

$$x_{k+1} = A(p_k)x_k + B(p_k)u_k + K(p_k)e_k,$$
$$y_k = C(p_k)x_k + D(p_k)u_k + e_k,$$

(jointly minimal State-Space (SS) representation)

$$M(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) & K(\cdot) \\ C(\cdot) & D(\cdot) & I \end{bmatrix} : \mathbb{P} \to \mathbb{R}^{(n_\mathrm{x}+n_\mathrm{y})\times(n_\mathrm{x}+n_\mathrm{u}+n_\mathrm{y})}$$ smooth matrix function

Input	Output	State	Scheduling
$u : \mathbb{N} \to \mathcal{U} \subseteq \mathbb{R}^{n_\mathrm{u}}$	$y : \mathbb{N} \to \mathcal{Y} \subseteq \mathbb{R}^{n_\mathrm{y}}$	$x : \mathbb{N} \to \mathcal{X} \subseteq \mathbb{R}^{n_\mathrm{x}}$	$p : \mathbb{N} \to \mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$

$\mathbb{P}$ compact set; e ind. white noise with $e_k \sim \mathcal{N}(0, \Sigma_\mathrm{e})$.

Data-generating LPV system (innovation noise)

$$x_{k+1} = A(p_k)x_k + B(p_k)u_k + K(p_k)e_k,$$
$$y_k = C(p_k)x_k + D(p_k)u_k + e_k,$$

(jointly minimal State-Space (SS) representation)

$$M(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) & K(\cdot) \\ C(\cdot) & D(\cdot) & I \end{bmatrix} : \mathbb{P} \to \mathbb{R}^{(n_\mathrm{x}+n_\mathrm{y})\times(n_\mathrm{x}+n_\mathrm{u}+n_\mathrm{y})}$$ smooth matrix function

Input	Output	State	Scheduling
$u : \mathbb{N} \to \mathcal{U} \subseteq \mathbb{R}^{n_\mathrm{u}}$	$y : \mathbb{N} \to \mathcal{Y} \subseteq \mathbb{R}^{n_\mathrm{y}}$	$x : \mathbb{N} \to \mathcal{X} \subseteq \mathbb{R}^{n_\mathrm{x}}$	$p : \mathbb{N} \to \mathbb{P} \subseteq \mathbb{R}^{n_\mathrm{p}}$

$\mathbb{P}$ compact set; e ind. white noise with $e_k \sim \mathcal{N}(0, \Sigma_\mathrm{e})$.

The LPV predictor form

$$x_{k+1} = \tilde{A}(p_k)x_k + \tilde{B}(p_k)u_k + K(p_k)y_k,$$
$$y_k = C(p_k)x_k + D(p_k)u_k + e_k,$$

(Required to be asymptotically stable)

$$e_k = y_k - \underbrace{(C(p_k)x_k + D(p_k)u_k)}_{\mathbb{E}\{y_k|M, x_k, p_k, u_k\}} \quad \text{(prediction error)}$$

Predictor state matrix
$\tilde{A}(p_k) = A(p_k) - K(p_k)C(p_k)$

Predictor input matrix
$\tilde{B}(p_k) = B(p_k) - K(p_k)D(p_k)$

Measurements: $\mathcal{D}_N = \{u_k, y_k, p_k\}_{k=1}^N$

The LPV predictor form

$$x_{k+1} = \tilde{A}(p_k)x_k + \tilde{B}(p_k)u_k + K(p_k)y_k,$$
$$y_k = C(p_k)x_k + D(p_k)u_k + e_k,$$

(Required to be asymptotically stable)

$$e_k = y_k - \underbrace{(C(p_k)x_k + D(p_k)u_k)}_{\mathbb{E}\{y_k|M,x_k,p_k,u_k\}} \quad \text{(prediction error)}$$

Predictor state matrix
$\tilde{A}(p_k) = A(p_k) - K(p_k)C(p_k)$

Predictor input matrix
$\tilde{B}(p_k) = B(p_k) - K(p_k)D(p_k)$

Measurements: $\mathcal{D}_N = \{u_k, y_k, p_k\}_{k=1}^{N}$

The LPV predictor form

$$
\begin{aligned}
x_{k+1} &= \tilde{A}(p_k)x_k + \tilde{B}(p_k)u_k + K(p_k)y_k, \\
y_k &= C(p_k)x_k + D(p_k)u_k + e_k,
\end{aligned}
$$

(Required to be asymptotically stable)

$$
e_k = y_k - \underbrace{(C(p_k)x_k + D(p_k)u_k)}_{\mathbb{E}\{y_k | M, x_k, p_k, u_k\}} \quad \text{(prediction error)}
$$

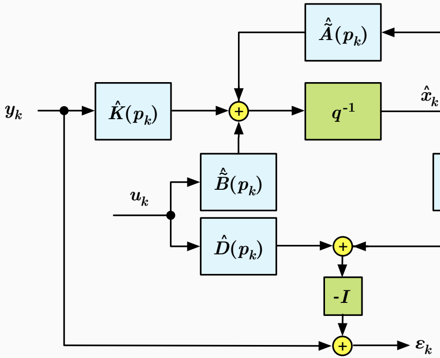
Predictor state matrix	Predictor input matrix
$\tilde{A}(p_k) = A(p_k) - K(p_k)C(p_k)$	$\tilde{B}(p_k) = B(p_k) - K(p_k)D(p_k)$

Measurements: $\mathcal{D}_N = \{u_k, y_k, p_k\}_{k=1}^{N}$

Identification setting

For a given choice of $\hat{M}$ and $\hat{x}_0$, let the prediction error be ε_k w.r.t $\mathcal{D}_N$.

$$\min_{\hat{M},\hat{x}_0} V_N := \frac{1}{N}\sum_{k=0}^{N-1} \|\varepsilon_k\|_2^2$$



Available methods

Assuming that M is an affine function $\Rightarrow$ many methods:

- PEM-SS (gradient based) [Cox et. al. 2017, Verdult et al. 2002]
- Expectation maximization (EM) [Wills and Ninnes 2011]
- Ho-Kalman realization based [Cox et al. 2016]
- PBSID and its variants e.g., [van Wingerden and Verhaegen 2009]
- Early attempts for subspace schemes e.g., [Verdult and Verhaegen 2002]
- Kalman filter based [dos Santos et al. 2008]
- Set membership (SM) e.g., [Sznaier et al. 2000]

Only a few attempts for non-parametric estimation of M under measured $\{x_k\}_{k=1}^{N}$.

Available methods

Assuming that M is an affine function $\Rightarrow$ many methods:

- PEM-SS (gradient based) [Cox et. al. 2017, Verdult et al. 2002]
- Expectation maximization (EM) [Wills and Ninnes 2011]
- Ho-Kalman realization based [Cox et al. 2016]
- PBSID and its variants e.g., [van Wingerden and Verhaegen 2009]
- Early attempts for subspace schemes e.g., [Verdult and Verhaegen 2002]
- Kalman filter based [dos Santos et al. 2008]
- Set membership (SM) e.g., [Sznaier et al. 2000]

Only a few attempts for non-parametric estimation of M under measured $\{x_k\}_{k=1}^{N}$.

Problems

- serious curse of dimensionality (research on LPV subspace schemes)
- parametrization of dependency (huge assumptions/simplifications)
- choice of the innovation structure (limited noise scenario)

Main question

How to mitigate the curse of dimensionality (reach efficiency of the LTI case) and eliminate the need for parametrization (functional estimate of M) in LPV-SS identification?

Consider a feedback-free scenario and let's think out of the box!

Kernelized Canonical Correlation Analysis

Canonical Correlation Analysis (CCA)

(originally studied by Hotelling 1936)

Given two random vectors $u \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ with

$$\Sigma = \mathbb{E}\left\{ \begin{bmatrix} u \\ y \end{bmatrix} \begin{bmatrix} u^\top & y^\top \end{bmatrix} \right\} = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uy} \\ \Sigma_{yu} & \Sigma_{yy} \end{bmatrix}$$

Find vectors v_i and w_i such that all correlation moments

$$\rho_i = \frac{\mathrm{cov}\{v_i^\top u, w_i^\top y\}}{\sqrt{\mathrm{var}\{v_i^\top u\}}\sqrt{\mathrm{var}\{w_i^\top y\}}} = \frac{v_i^\top \Sigma_{uy} w_i}{\sqrt{v_i^\top \Sigma_{uu} v_i}\sqrt{w_i^\top \Sigma_{yy} w_i}}$$

are maximized. For normalization (canonical) impose that $v_i^\top \Sigma_{uu} v_i = 1$ and $w_i^\top \Sigma_{yy} w_i = 1$. Define also $\Pi = \mathrm{diag}(\rho_1, \ldots \rho_n)$ and $z_\mathrm{u} = V^\top u$, $z_\mathrm{y} = W^\top y$ as canonical correlates.

Why? We want to find a latent variable $x = \Pi^{-1/2} V^\top u$ such that $\hat{y} = W \Pi^{1/2} x$

(core idea of state realization / estimation)

Canonical Correlation Analysis (CCA)

(originally studied by Hotelling 1936)

Given two random vectors $u \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ with

$$\Sigma = \mathbb{E}\left\{\begin{bmatrix} u \\ y \end{bmatrix}\begin{bmatrix} u^\top & y^\top \end{bmatrix}\right\} = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uy} \\ \Sigma_{yu} & \Sigma_{yy} \end{bmatrix}$$

Find vectors v_i and w_i such that all correlation moments

$$\rho_i = \frac{\mathrm{cov}\{v_i^\top u, w_i^\top y\}}{\sqrt{\mathrm{var}\{v_i^\top u\}}\sqrt{\mathrm{var}\{w_i^\top y\}}} = \frac{v_i^\top \Sigma_{uy} w_i}{\sqrt{v_i^\top \Sigma_{uu} v_i}\sqrt{w_i^\top \Sigma_{yy} w_i}}$$

are maximized. For normalization (canonical) impose that $v_i^\top \Sigma_{uu} v_i = 1$ and $w_i^\top \Sigma_{yy} w_i = 1$.
Define also $\Pi = \mathrm{diag}(\rho_1, \ldots \rho_n)$ and $z_{\mathrm{u}} = V^\top u$, $z_{\mathrm{y}} = W^\top y$ as canonical correlates.

Why? We want to find a latent variable $x = \Pi^{-1/2} V^\top u$ such that $\hat{y} = W \Pi^{1/2} x$
(core idea of state realization / estimation)

Canonical Correlation Analysis (CCA)

(originally studied by Hotelling 1936)

Given two random vectors $u \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ with

$$\Sigma = \mathbb{E}\left\{\begin{bmatrix} u \\ y \end{bmatrix}\begin{bmatrix} u^\top & y^\top \end{bmatrix}\right\} = \begin{bmatrix} \Sigma_{uu} & \Sigma_{uy} \\ \Sigma_{yu} & \Sigma_{yy} \end{bmatrix}$$

Find vectors v_i and w_i such that all correlation moments

$$\rho_i = \frac{\mathrm{cov}\{v_i^\top u, w_i^\top y\}}{\sqrt{\mathrm{var}\{v_i^\top u\}}\sqrt{\mathrm{var}\{w_i^\top y\}}} = \frac{v_i^\top \Sigma_{uy} w_i}{\sqrt{v_i^\top \Sigma_{uu} v_i}\sqrt{w_i^\top \Sigma_{yy} w_i}}$$

are maximized. For normalization (canonical) impose that $v_i^\top \Sigma_{uu} v_i = 1$ and $w_i^\top \Sigma_{yy} w_i = 1$. Define also $\Pi = \mathrm{diag}(\rho_1, \dots \rho_n)$ and $z_{\mathrm{u}} = V^\top u$, $z_{\mathrm{y}} = W^\top y$ as canonical correlates.

Why? We want to find a latent variable $x = \Pi^{-1/2} V^\top u$ such that $\hat{y} = W \Pi^{1/2} x$ (core idea of state realization / estimation)

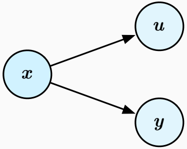
By the Lagrangian method, the solutions correspond to the GEP problem:

$$\begin{bmatrix} 0 & \Sigma_{uy} \\ \Sigma_{yu} & 0 \end{bmatrix} \begin{bmatrix} v_i \\ w_i \end{bmatrix} = \rho_i \begin{bmatrix} \Sigma_{uu} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix} \begin{bmatrix} v_i \\ w_i \end{bmatrix}$$

giving that $i \in \{1, \ldots n_{\mathrm{x}}\}$ with $0 < n_{\mathrm{x}} \leq n$.

Solved via sample based approximation of Σ and SVD.

Stochastic interpretation, e.g., [Bach and Jordan 2006]:



Optimal conditional entropy realization:

$$\begin{aligned} x &\sim \mathcal{N}(0, I_{n_{\mathrm{x}}}) \\ u \mid x &\sim \mathcal{N}(\tilde{V} I_{n_{\mathrm{x}}}, \Psi_{\mathrm{u}}) & \tilde{V} &= V\Pi^{1/2}; \\ y \mid x &\sim \mathcal{N}(\tilde{W} I_{n_{\mathrm{x}}}, \Psi_{\mathrm{y}}) & \tilde{W} &= W\Pi^{1/2}; \end{aligned}$$

Estimation of $(\tilde{V}, \tilde{W}, \Psi_{\mathrm{u}}, \Psi_{\mathrm{y}})$ via sampled Σ is an ML solution under Gaussian noise.

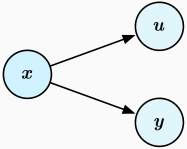
By the Lagrangian method, the solutions correspond to the GEP problem:

$$\begin{bmatrix} 0 & \Sigma_{uy} \\ \Sigma_{yu} & 0 \end{bmatrix} \begin{bmatrix} v_i \\ w_i \end{bmatrix} = \rho_i \begin{bmatrix} \Sigma_{uu} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix} \begin{bmatrix} v_i \\ w_i \end{bmatrix}$$

giving that $i \in \{1, \ldots n_{\mathrm{x}}\}$ with $0 < n_{\mathrm{x}} \leq n$.

Solved via sample based approximation of Σ and SVD.

Stochastic interpretation, e.g., [Bach and Jordan 2006]:



Optimal conditional entropy realization:

$$x \sim \mathcal{N}(0, I_{n_{\mathrm{x}}})$$

$$u \mid x \sim \mathcal{N}(\tilde{V} I_{n_{\mathrm{x}}}, \Psi_{\mathrm{u}}) \qquad \tilde{V} = V\Pi^{1/2};$$

$$y \mid x \sim \mathcal{N}(\tilde{W} I_{n_{\mathrm{x}}}, \Psi_{\mathrm{y}}) \qquad \tilde{W} = W\Pi^{1/2};$$

Estimation of $(\tilde{V}, \tilde{W}, \Psi_{\mathrm{u}}, \Psi_{\mathrm{y}})$ via sampled Σ is an ML solution under Gaussian noise.

Kernelized Canonical Correlation Analysis (KCCA)

What happens if we have a nonlinear relationship?

Let $\mathcal{F}$ be a function space and consider the correlates

$$\left.\begin{aligned} z_{\mathrm{u}} &= f_{\mathrm{v}}(u) \\ z_{\mathrm{y}} &= f_{\mathrm{w}}(y) \end{aligned}\right\} \underset{\text{target}}{\Rightarrow} \quad \begin{aligned} x &= \tilde{f}_{\mathrm{v}}(u) \\ x &= \tilde{f}_{\mathrm{w}}(y) \end{aligned}$$

Objective: maximize the $\mathcal{F}$-correlation:

$$\rho_{\mathcal{F}} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \operatorname{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y)) = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \frac{\operatorname{cov}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y))}{(\operatorname{var} f_{\mathrm{v}}(u))^{1/2} (\operatorname{var} f_{\mathrm{w}}(y))^{1/2}}$$

Kernelized Canonical Correlation Analysis (KCCA)

What happens if we have a nonlinear relationship?

Let $\mathcal{F}$ be a function space and consider the correlates

$$\left.\begin{array}{rcl} z_{\mathrm{u}} & = & f_{\mathrm{v}}(u) \\ z_{\mathrm{y}} & = & f_{\mathrm{w}}(y) \end{array}\right\} \underset{\text{target}}{\Rightarrow} \begin{array}{rcl} x & = & \tilde{f}_{\mathrm{v}}(u) \\ x & = & \tilde{f}_{\mathrm{w}}(y) \end{array}$$

Objective: maximize the $\mathcal{F}$-correlation:

$$\rho_{\mathcal{F}} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \mathrm{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y)) = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \frac{\mathrm{cov}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y))}{(\mathrm{var}\ f_{\mathrm{v}}(u))^{1/2} (\mathrm{var}\ f_{\mathrm{w}}(y))^{1/2}}$$

Let $\mathcal{F}$ be an RKHS on $\mathbb{R}^n$ and let $\mathcal{K}: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the associated symmetric, (continuous) positive definite kernel and $\langle \cdot, \cdot \rangle$ the corresponding inner product (induced norm).

$\phi_z = \mathcal{K}(\cdot, z)$ is a so called feature map satisfying

$$f(z) = \langle \phi_z, f \rangle, \quad \forall f \in \mathcal{F}, \ \forall z \in \mathbb{R}^n$$

$$\Downarrow$$

$$\mathrm{corr}(f_1(z_1), f_2(z_2)) = \mathrm{corr}(\langle \phi_{z_1}, f_1 \rangle, \langle \phi_{z_2}, f_2 \rangle)$$

which projects the previous $\mathcal{F}$-correlation w.r.t. the feature space.

For observations $z_{1,1}, \ldots z_{1,N}$ and $z_{2,1}, \ldots, z_{2,N}$ centered in feature space, i.e., $\sum_{k=1}^{N} \phi_{z_{1,k}} = 0$ and $\sum_{k=1}^{N} \phi_{z_{2,k}} = 0$:

$$f_1 = \sum_{k=1}^{N} \alpha_{1,k} \phi_{z_{1,k}} + f_1^{\perp}, \quad f_2 = \sum_{k=1}^{N} \alpha_{2,k} \phi_{z_{2,k}} + f_2^{\perp}$$

corresponding to projection on subspaces.

Let $\mathcal{F}$ be an RKHS on $\mathbb{R}^n$ and let $\mathcal{K} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the associated symmetric, (continuous) positive definite kernel and $\langle \cdot, \cdot \rangle$ the corresponding inner product (induced norm).

$\phi_z = \mathcal{K}(\cdot, z)$ is a so called feature map satisfying

$$f(z) = \langle \phi_z, f \rangle, \quad \forall f \in \mathcal{F}, \ \forall z \in \mathbb{R}^n$$

$$\Downarrow$$

$$\text{corr}(f_1(z_1), f_2(z_2)) = \text{corr}(\langle \phi_{\mathbf{z_1}}, f_1 \rangle, \langle \phi_{\mathbf{z_2}}, f_2 \rangle)$$

which projects the previous $\mathcal{F}$-correlation w.r.t. the feature space.

For observations $z_{1,1}, \ldots z_{1,N}$ and $z_{2,1}, \ldots, z_{2,N}$ centered in feature space, i.e., $\sum_{k=1}^{N} \phi_{z_{\mathbf{1},k}} = 0$ and $\sum_{k=1}^{N} \phi_{z_{\mathbf{2},k}} = 0$:

$$f_1 = \sum_{k=1}^{N} \alpha_{1,k} \phi_{z_{\mathbf{1},k}} + f_1^{\perp}, \quad f_2 = \sum_{k=1}^{N} \alpha_{2,k} \phi_{z_{\mathbf{2},k}} + f_2^{\perp}$$

corresponding to projection on subspaces.

Let $\mathcal{F}$ be an RKHS on $\mathbb{R}^n$ and let $\mathcal{K} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the associated symmetric, (continuous) positive definite kernel and $\langle \cdot, \cdot \rangle$ the corresponding inner product (induced norm).

$\phi_z = \mathcal{K}(\cdot, z)$ is a so called feature map satisfying

$$f(z) = \langle \phi_z, f \rangle, \quad \forall f \in \mathcal{F}, \; \forall z \in \mathbb{R}^n$$

$$\Downarrow$$

$$\mathrm{corr}(f_1(z_1), f_2(z_2)) = \mathrm{corr}(\langle \phi_{z_1}, f_1 \rangle, \langle \phi_{z_2}, f_2 \rangle)$$

which projects the previous $\mathcal{F}$-correlation w.r.t. the feature space.

For observations $z_{1,1}, \ldots z_{1,N}$ and $z_{2,1}, \ldots, z_{2,N}$ centered in feature space, i.e., $\sum_{k=1}^N \phi_{z_{\mathbf{1},k}} = 0$ and $\sum_{k=1}^N \phi_{z_{\mathbf{2},k}} = 0$:

$$f_1 = \sum_{k=1}^N \alpha_{1,k} \phi_{z_{\mathbf{1},k}} + f_1^\perp, \quad f_2 = \sum_{k=1}^N \alpha_{2,k} \phi_{z_{\mathbf{2},k}} + f_2^\perp$$

corresponding to projection on subspaces.

Sampled $\mathcal{F}$ correlation:

$$\hat{\rho}_{\mathcal{F}} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \operatorname{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y) \mid \{u_k\}_{k=1}^N, \{y_k\}_{k=1}^N)$$

$$= \max_{\alpha_{\mathrm{v}}, \alpha_{\mathrm{w}} \in \mathbb{R}^N} \frac{\alpha_{\mathrm{v}}^\top K_{uy} \alpha_{\mathrm{w}}}{(\alpha_{\mathrm{v}}^\top K_{uu} \alpha_{\mathrm{v}})^{1/2} (\alpha_{\mathrm{w}}^\top K_{yy} \alpha_{\mathrm{w}})^{1/2}}$$

where $[K_u]_{i,j} = \mathcal{K}(u_i, u_j)$ and $K_{uy} = K_u K_y$.

By the Lagrangian method, the solutions correspond to the GEP problem:

$$\begin{bmatrix} 0 & K_{uy} \\ K_{yu} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix} = \hat{\rho}_{\mathcal{F}} \begin{bmatrix} K_{uu} & 0 \\ 0 & K_{yy} \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix}$$

Resulting correlates: $z_u = \hat{f}_{\mathrm{v}}(u) = \sum_{k=1}^N \alpha_{\mathrm{v},k} \mathcal{K}(u_k, u)$ and $z_y = \hat{f}_{\mathrm{w}}(y) = \sum_{k=1}^N \alpha_{\mathrm{w},k} \mathcal{K}(y_k, y)$

Note that the Gram matrices can be centered in case of non-centered data. All correlation moments are obtained as all solutions of the GEP.

Sampled $\mathcal{F}$ correlation:

$$\hat{\rho}_{\mathcal{F}} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \operatorname{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y) \mid \{u_k\}_{k=1}^{N}, \{y_k\}_{k=1}^{N})$$

$$= \max_{\alpha_{\mathrm{v}}, \alpha_{\mathrm{w}} \in \mathbb{R}^N} \frac{\alpha_{\mathrm{v}}^\top K_{uy} \alpha_{\mathrm{w}}}{(\alpha_{\mathrm{v}}^\top K_{uu} \alpha_{\mathrm{v}})^{1/2} (\alpha_{\mathrm{w}}^\top K_{yy} \alpha_{\mathrm{w}})^{1/2}}$$

where $[K_u]_{i,j} = \mathcal{K}(u_i, u_j)$ and $K_{uy} = K_u K_y$.

By the Lagrangian method, the solutions correspond to the GEP problem:

$$\begin{bmatrix} 0 & K_{uy} \\ K_{yu} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix} = \hat{\rho}_{\mathcal{F}} \begin{bmatrix} K_{uu} & 0 \\ 0 & K_{yy} \end{bmatrix} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix}$$

Resulting correlates: $z_u = \hat{f}_{\mathrm{v}}(u) = \sum_{k=1}^{N} \alpha_{\mathrm{v},k} \mathcal{K}(u_k, u)$ and $z_y = \hat{f}_{\mathrm{w}}(y) = \sum_{k=1}^{N} \alpha_{\mathrm{w},k} \mathcal{K}(y_k, y)$

Note that the Gram matrices can be centered in case of non-centered data. All correlation moments are obtained as all solutions of the GEP.

Sampled $\mathcal{F}$ correlation:

$$\hat{\rho}_{\mathcal{F}} = \max_{f_{\rm v}, f_{\rm w} \in \mathcal{F}} \text{corr}(f_{\rm v}(u), f_{\rm w}(y) \mid \{u_k\}_{k=1}^N, \{y_k\}_{k=1}^N)$$

$$= \max_{\alpha_{\rm v}, \alpha_{\rm w} \in \mathbb{R}^N} \frac{\alpha_{\rm v}^\top K_{uy} \alpha_{\rm w}}{(\alpha_{\rm v}^\top K_{uu} \alpha_{\rm v})^{1/2} (\alpha_{\rm w}^\top K_{yy} \alpha_{\rm w})^{1/2}}$$

where $[K_u]_{i,j} = \mathcal{K}(u_i, u_j)$ and $K_{uy} = K_u K_y$.

By the Lagrangian method, the solutions correspond to the GEP problem:

$$\begin{bmatrix} 0 & K_{uy} \\ K_{yu} & 0 \end{bmatrix} \begin{bmatrix} \alpha_{\rm v} \\ \alpha_{\rm w} \end{bmatrix} = \hat{\rho}_{\mathcal{F}} \begin{bmatrix} K_{uu} & 0 \\ 0 & K_{yy} \end{bmatrix} \begin{bmatrix} \alpha_{\rm v} \\ \alpha_{\rm w} \end{bmatrix}$$

Resulting correlates: $z_u = \hat{f}_{\rm v}(u) = \sum_{k=1}^N \alpha_{{\rm v},k} \mathcal{K}(u_k, u)$ and $z_y = \hat{f}_{\rm w}(y) = \sum_{k=1}^N \alpha_{{\rm w},k} \mathcal{K}(y_k, y)$

Note that the Gram matrices can be centered in case of non-centered data. All correlation moments are obtained as all solutions of the GEP.

Need for regularization:

K_{uu} and K_{yy} are centered, hence under general kernel functions and data they can be singular.

Regularized $\mathcal{F}$-correlation ($\kappa > 0$):

$$\rho_{\mathcal{F}}^{\kappa} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \mathrm{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y)) = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \frac{\mathrm{cov}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y))}{(\mathrm{var}\ f_{\mathrm{v}}(u) + \kappa \|f_1\|_{\mathcal{F}}^2)^{1/2} (\mathrm{var}\ f_{\mathrm{w}}(y) + \kappa \|f_2\|_{\mathcal{F}}^2)^{1/2}}$$

which gives the GEP:

$$\underbrace{\begin{bmatrix} 0 & K_{uy} \\ K_{yu} & 0 \end{bmatrix}}_{\mathfrak{K}_\kappa} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix} = \lambda \underbrace{\begin{bmatrix} (K_u + \frac{N\kappa}{2} I)^2 & 0 \\ 0 & (K_y + \frac{N\kappa}{2} I)^2 \end{bmatrix}}_{\mathfrak{D}_\kappa} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix}$$

Consistent estimator of the regularized $\mathcal{F}$-correlation [Bach and Jordan 2003].

Effect of regularization: $\lambda \rightarrow \frac{\lambda}{\lambda + \frac{N\kappa}{2}}$

Need for regularization:

K_{uu} and K_{yy} are centered, hence under general kernel functions and data they can be singular.

Regularized $\mathcal{F}$-correlation ($\kappa > 0$):

$$\rho_{\mathcal{F}}^{\kappa} = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \operatorname{corr}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y)) = \max_{f_{\mathrm{v}}, f_{\mathrm{w}} \in \mathcal{F}} \frac{\operatorname{cov}(f_{\mathrm{v}}(u), f_{\mathrm{w}}(y))}{(\operatorname{var} f_{\mathrm{v}}(u) + \kappa \|f_1\|_{\mathcal{F}}^2)^{1/2} (\operatorname{var} f_{\mathrm{w}}(y) + \kappa \|f_2\|_{\mathcal{F}}^2)^{1/2}}$$

which gives the GEP:

$$\underbrace{\begin{bmatrix} 0 & K_{uy} \\ K_{yu} & 0 \end{bmatrix}}_{\mathfrak{K}_\kappa} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix} = \lambda \underbrace{\begin{bmatrix} (K_u + \frac{N\kappa}{2} I)^2 & 0 \\ 0 & (K_y + \frac{N\kappa}{2} I)^2 \end{bmatrix}}_{\mathfrak{D}_\kappa} \begin{bmatrix} \alpha_{\mathrm{v}} \\ \alpha_{\mathrm{w}} \end{bmatrix}$$

Consistent estimator of the regularized $\mathcal{F}$-correlation [Bach and Jordan 2003].

Effect of regularization: $\lambda \rightarrow \frac{\lambda}{\lambda + \frac{N\kappa}{2}}$

Choice of kernel

Gaussian kernel:

$$\mathcal{K}(u, y) = \exp\left(-\frac{\|u - y\|_2^2}{2\sigma^2}\right)$$

where $\sigma > 0$ is a hyper-parameter of the kernel width.

Polynomial kernel:

$$\mathcal{K}(u, y) = (x^\top y + c)^l$$

where $c \in \mathbb{R}$ and $l \in \mathbb{N}$ are hyper-parameters.

Hermite polynomial kernels, etc.

Choice of kernel

Gaussian kernel:

$$\mathcal{K}(u, y) = \exp\left(-\frac{\|u - y\|_2^2}{2\sigma^2}\right)$$

where $\sigma > 0$ is a hyper-parameter of the kernel width.

Polynomial kernel:

$$\mathcal{K}(u, y) = (x^\top y + c)^l$$

where $c \in \mathbb{R}$ and $l \in \mathbb{N}$ are hyper-parameters.

Hermite polynomial kernels, etc.

Choice of kernel

Gaussian kernel:

$$\mathcal{K}(u, y) = \exp\left(-\frac{\|u - y\|_2^2}{2\sigma^2}\right)$$

where $\sigma > 0$ is a hyper-parameter of the kernel width.

Polynomial kernel:

$$\mathcal{K}(u, y) = (x^\top y + c)^l$$

where $c \in \mathbb{R}$ and $l \in \mathbb{N}$ are hyper-parameters.

Hermite polynomial kernels, etc.

Choice of kernel

Gaussian kernel:

$$\mathcal{K}(u, y) = \exp\left(-\frac{\|u - y\|_2^2}{2\sigma^2}\right)$$

where $\sigma > 0$ is a hyper-parameter of the kernel width.

Polynomial kernel:

$$\mathcal{K}(u, y) = (x^\top y + c)^l$$

where $c \in \mathbb{R}$ and $l \in \mathbb{N}$ are hyper-parameters.

Hermite polynomial kernels, etc.

How to chose hyper-parameters θ associated with the kernel $\mathcal{K}$?

Limited literature, mainly based on various forms of Cross Validation (CV).

Minimization of the conditional entropy of u and y given z_u and z_y [Bach and Jordan 2003]?

Mutual information between the projected variables z_u and z_y in case of jointly Gaussian variables via the CCA:

$$\mathcal{I} = -\frac{1}{2}\sum_{i=1}^{n_x} log(1-\rho_i^2)$$

Approximation for KCCA in terms of the contrast function (generalized variance):

$$\hat{\mathcal{I}}_{\mathcal{F}} = \frac{\det(\mathfrak{K}_\kappa)}{\det \mathfrak{D}_\kappa}$$

Then compute derivative w.r.t. hyper-parameters and minimize it.

How to chose hyper-parameters θ associated with the kernel $\mathcal{K}$?

Limited literature, mainly based on various forms of Cross Validation (CV).

Minimization of the conditional entropy of u and y given z_u and z_y [Bach and Jordan 2003]?

Mutual information between the projected variables z_u and z_y in case of jointly Gaussian variables via the CCA:

$$\mathcal{I} = -\frac{1}{2}\sum_{i=1}^{n_x} log(1-\rho_i^2)$$

Approximation for KCCA in terms of the contrast function (generalized variance):

$$\hat{\mathcal{I}}_{\mathcal{F}} = \frac{\det(\mathfrak{K}_\kappa)}{\det \mathfrak{D}_\kappa}$$

Then compute derivative w.r.t. hyper-parameters and minimize it.

How to chose hyper-parameters θ associated with the kernel $\mathcal{K}$?

Limited literature, mainly based on various forms of Cross Validation (CV).

Minimization of the conditional entropy of u and y given z_u and z_y [Bach and Jordan 2003]?

Mutual information between the projected variables z_u and z_y in case of jointly Gaussian variables via the CCA:

$$\mathcal{I} = -\frac{1}{2}\sum_{i=1}^{n_x} log(1-\rho_i^2)$$

Approximation for KCCA in terms of the contrast function (generalized variance):

$$\hat{\mathcal{I}}_{\mathcal{F}} = \frac{\det(\mathfrak{K}_\kappa)}{\det \mathfrak{D}_\kappa}$$

Then compute derivative w.r.t. hyper-parameters and minimize it.

Forward and backward prediction of the state

Notation to characterize past and future:

$$\bar{p}_k^d := \begin{bmatrix} p_{k-d}^\top & \cdots & p_{k-1}^\top \end{bmatrix}^\top \in \mathbb{R}^{dn_\mathrm{p}} \qquad \text{past data}$$

$$\bar{p}_{k+d}^d := \begin{bmatrix} p_k^\top & \cdots & p_{k+d-1}^\top \end{bmatrix}^\top \in \mathbb{R}^{dn_\mathrm{p}} \qquad \text{future data}$$

$\bar{u}_k^d \in \mathbb{R}^{dn_\mathrm{u}}$, $\bar{y}_k^d \in \mathbb{R}^{dn_\mathrm{y}}$, $\bar{u}_{k+d}^d \in \mathbb{R}^{dn_\mathrm{u}}$, and $\bar{y}_{k+d}^d \in \mathbb{R}^{dn_\mathrm{y}}$ are defined in a similar way.

we also define $\bar{z}_k^d = \begin{bmatrix} \bar{u}_k^d \\ \bar{y}_k^d \end{bmatrix}$, $\bar{z}_{k+d}^d = \begin{bmatrix} \bar{u}_{k+d}^d \\ \bar{y}_{k+d}^d \end{bmatrix} \in \mathbb{R}^{d(n_\mathrm{u}+n_\mathrm{y})}$.

$$\begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+d-1} \end{bmatrix} = \underbrace{\begin{bmatrix} C(p_k) \\ C(p_{k+1})\tilde{A}(p_k) \\ \vdots \\ C(p_{k+d-1})\prod_{l=2}^{d}\tilde{A}(p_{k+d-l}) \end{bmatrix}}_{(\mathcal{O}_{\mathrm{f}}^{d}\diamond p)(k)} x_k + \underbrace{\begin{bmatrix} D(p_k) & 0 & \cdots & 0 \\ C(p_{k+1})\tilde{B}(p_k) & D(p_{k+1}) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ C(p_{k+d-1})\prod_{l=2}^{d-1}\tilde{A}(p_{k+d-l})\tilde{B}(p_k) & C(p_{k+d-1})\prod_{l=2}^{d-2}\tilde{A}(p_{k+d-l})\tilde{B}(p_{k+1}) & \cdots & D(p_{k+d-1}) \end{bmatrix}}_{(\mathcal{H}_{\mathrm{f}}^{d}\diamond p)(k)} \begin{bmatrix} u_k \\ u_{k+1} \\ \vdots \\ u_{k+d-1} \end{bmatrix}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ C(p_{k+1})K(p_k) & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ C(p_{k+d-1})\prod_{l=2}^{d-1}\tilde{A}(p_{k+d-l})K(p_k) & C(p_{k+d-1})\prod_{l=2}^{d-2}\tilde{A}(p_{k+d-l})K(p_{k+1}) & \cdots & 0 \end{bmatrix}}_{(\mathcal{L}_{\mathrm{f}}^{d}\diamond p)(k)} \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+d-1} \end{bmatrix} + \begin{bmatrix} e_k \\ e_{k+1} \\ \vdots \\ e_{k+d-1} \end{bmatrix}.$$

Forward data equation

$$\bar{y}^{d}_{k+d} = (\mathcal{O}^{d}_{\mathrm{f}} \diamond p)(k) \cdot x_k + (\mathcal{H}^{d}_{\mathrm{f}} \diamond p)(k) \cdot \bar{u}^{d}_{k+d} + (\mathcal{L}^{d}_{\mathrm{f}} \diamond p)(k) \cdot \bar{y}^{d}_{k+d} + \bar{e}^{d}_{k+d},$$

- $(\mathcal{O}^{d}_{\mathrm{f}} \diamond p)(k) \in \mathbb{R}^{dn_{\mathrm{y}} \times n}$ is the time-varying d-step forward observability matrix
- $(\mathcal{H}^{d}_{\mathrm{f}} \diamond p)(k) \in \mathbb{R}^{dn_{\mathrm{y}} \times dn_{\mathrm{u}}}$ is a matrix with Infinite Impulse Response (IIR) coefficients
- $(\mathcal{L}^{d}_{\mathrm{f}} \diamond p)(k) \in \mathbb{R}^{dn_{\mathrm{y}} \times dn_{\mathrm{y}}}$ is a lower triangular matrix
- Assumption: *structural observability* in the deterministic sense

Definition (Structural observability)

The LPV-SS representation with state-dimension n_{x} is called structurally observable, if there exists a scheduling trajectory $p \in \mathbb{P}^{\mathbb{Z}}$, such that the n_{x}-step observability matrix $(\mathcal{O}^{n_{\mathrm{x}}}_{\mathrm{f}} \diamond p)(k)$ is full (column) rank for all $k \in \mathbb{Z}$.

This gives a necessary PE condition on $\mathcal{D}_N$.

Forward state transfer:

(based on the forward output equation)

$$x_k = \left(\mathcal{O}_{\mathrm{f}}^d(k)\right)^{\dagger}\left(\left(I - \mathcal{L}_{\mathrm{f}}^d(k)\right)\bar{y}_{k+d}^d - \mathcal{H}_{\mathrm{f}}^d(k)\bar{u}_{k+d}^d\right) - \left(\mathcal{O}_{\mathrm{f}}^d(k)\right)^{\dagger}\bar{e}_{k+d}^d,$$

Backward state transfer:

(based on the state equation)

$$x_k = \underbrace{\left(\prod_{i=1}^{d}\tilde{A}(p_{k-i})\right)}_{\mathcal{X}_{\mathrm{p}}^d(k)} x_{k-d} + \mathcal{R}_{\mathrm{p}}^d(k)\bar{u}_k^d + \mathcal{V}_{\mathrm{p}}^d(k)\bar{y}_k^d,$$

Using a similar definition of a d-step backward reachability matrix $\mathcal{R}_{\mathrm{p}}^d(k)$ depending on $p_{k-d}, \ldots, p_{k-1}$ and its counterpart $\mathcal{V}_{\mathrm{p}}^d(k)$ with respect to $K(p_k)$.

Assumption: *structural state controllability*.

Main data equation

(relation of future and past IO data)

$$\bar{y}^d_{k+d} = \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{R}^d_{\mathrm{p}}(k)\bar{u}^d_k + \mathcal{H}^d_{\mathrm{f}}(k)\bar{u}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{V}^d_{\mathrm{p}}(k)\bar{y}^d_k + (\mathcal{L}^d_{\mathrm{f}} \diamond p)(k) \cdot \bar{y}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{X}^d_{\mathrm{p}}(k)x_{k-d} + \bar{e}^d_{k+d}$$

where d is chosen such that $\mathcal{X}^d_{\mathrm{p}}(k) \approx 0$ due to the asymptotic stability of the predictor form.

Central equation in LPV subspace identification.

Recent unified subspace theory under affine dependence [Cox 2017].
CCA-affine case $\Rightarrow$ ML estimator. Parametrization $\Rightarrow$ curse of dimensionality.

How to estimate the state without parameterization of the dependency?

Main data equation

(relation of future and past IO data)

$$\bar{y}^d_{k+d} = \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{R}^d_{\mathrm{p}}(k)\bar{u}^d_k + \mathcal{H}^d_{\mathrm{f}}(k)\bar{u}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{V}^d_{\mathrm{p}}(k)\bar{y}^d_k + (\mathcal{L}^d_{\mathrm{f}} \diamond p)(k) \cdot \bar{y}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{X}^d_{\mathrm{p}}(k)x_{k-d} + \bar{e}^d_{k+d}$$

where d is chosen such that $\mathcal{X}^d_{\mathrm{p}}(k) \approx 0$ due to the asymptotic stability of the predictor form.

Central equation in LPV subspace identification.

Recent unified subspace theory under affine dependence [Cox 2017].
CCA-affine case $\Rightarrow$ ML estimator. Parametrization $\Rightarrow$ curse of dimensionality.

How to estimate the state without parameterization of the dependency?

Main data equation

(relation of future and past IO data)

$$\bar{y}^d_{k+d} = \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{R}^d_{\mathrm{p}}(k)\bar{u}^d_k + \mathcal{H}^d_{\mathrm{f}}(k)\bar{u}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{V}^d_{\mathrm{p}}(k)\bar{y}^d_k + (\mathcal{L}^d_{\mathrm{f}} \diamond p)(k) \cdot \bar{y}^d_{k+d} + \mathcal{O}^d_{\mathrm{f}}(k)\mathcal{X}^d_{\mathrm{p}}(k)x_{k-d} + \bar{e}^d_{k+d}$$

where d is chosen such that $\mathcal{X}^d_{\mathrm{p}}(k) \approx 0$ due to the asymptotic stability of the predictor form.

Central equation in LPV subspace identification.

Recent unified subspace theory under affine dependence [Cox 2017].
CCA-affine case $\Rightarrow$ ML estimator. Parametrization $\Rightarrow$ curse of dimensionality.

How to estimate the state without parameterization of the dependency?

State estimation via KCCA

Application of KCCA

e is an *independent and identically distributed* (i.i.d) zero-mean process,
the expected value of the last term in the forward equation is zero:

$$\hat{x}_k = \underbrace{\left(\mathcal{O}_\mathrm{f}^d(k)\right)^\dagger \begin{bmatrix} -\mathcal{H}_\mathrm{f}^d(k) & I - \mathcal{L}_\mathrm{f}^d(k) \end{bmatrix}}_{\varphi_\mathrm{f}(\bar{p}_{k+d}^d)} \bar{z}_{k+d}^d \qquad \text{forward predictor}$$

$$\hat{x}_k = \underbrace{\begin{bmatrix} \mathcal{R}_\mathrm{p}^d(k) & \mathcal{V}_\mathrm{p}^d(k) \end{bmatrix}}_{\varphi_\mathrm{p}(\bar{p}_k^d)} \bar{z}_k^d, \qquad \text{backward predictor}$$

KCCA will change the state basis to maximize correlation of the predictors under a chosen $\mathcal{F}$.

Note that $f_\mathrm{f}(\bar{p}_{k+d}^d, \bar{z}_{k+d}^d) = \varphi_\mathrm{f}(\bar{p}_{k+d}^d)\bar{z}_{k+d}^d$ and $f_\mathrm{p}(\bar{p}_k^d, \bar{z}_k^d) = \varphi_\mathrm{p}(\bar{p}_k^d)\bar{z}_k^d$ correspond to $\mathcal{F}$.

Define the past and future data sets $\Phi_\mathrm{p}, \Phi_\mathrm{f} \in \mathbb{R}^{N\times n_\mathrm{x}}$ as

$$\Phi_\mathrm{p} := \begin{bmatrix} \varphi_\mathrm{p}(\bar{p}_1^d)\bar{z}_1^d & \cdots & \varphi_\mathrm{p}(\bar{p}_N^d)\bar{z}_N^d \end{bmatrix}^\top,$$

$$\Phi_\mathrm{f} := \begin{bmatrix} \varphi_\mathrm{f}(\bar{p}_{1+d}^d)\bar{z}_{1+d}^d & \cdots & \varphi_\mathrm{f}(\bar{p}_{N+d}^d)\bar{z}_{N+d}^d \end{bmatrix}^\top,$$

where $\varphi_\mathrm{p} : \mathbb{R}^{dn_\mathrm{p}} \to \mathbb{R}^{n_\mathrm{x}\times d(n_\mathrm{u}+n_\mathrm{y})}$ and $\varphi_\mathrm{f} : \mathbb{R}^{dn_\mathrm{p}} \to \mathbb{R}^{n_\mathrm{x}\times d(n_\mathrm{u}+n_\mathrm{y})}$ represent unknown feature maps with unknown n_x.

Let's use the previous KCCA theory with

RKHS	Generating kernel	Kernel core	Regularization
$\mathcal{F}$	$\mathcal{K}(s_1, s_2) = \bar{z}_1^\top \mathcal{K}_\mathrm{c}(\bar{p}_1, \bar{p}_2)\bar{z}_2$	$\mathcal{K}_\mathrm{c} : \mathbb{R}^{dn_\mathrm{p}} \times \mathbb{R}^{dn_\mathrm{p}} \to \mathbb{R}$	$\kappa > 0$

(Here, $s = \begin{bmatrix} \bar{z}^\top & \bar{p}^\top \end{bmatrix}^\top$)

Estimates:

Potentially $2N$ solutions computed via the SVD of a $2N \times 2N$ matrix.
(no curse of dimensionality)

State estimates using the forward correlates:

$$\breve{x}_k^j = \eta_j^\top \begin{bmatrix} \bar{z}_{1+d}^{d\,\top} \breve{k}(\bar{p}_{1+d}^d, \bar{p}_{k+d}^d) \\ \vdots \\ \bar{z}_{N+d}^{d\,\top} \breve{k}(p_{N+d}^d, \bar{p}_{k+d}^d) \end{bmatrix} \bar{z}_{k+d}^d.$$

State dimension is selected via:
(a) magnitude of resulting singular values (b) AIC [Larimore 2005].

Properties?

State is estimated as $\breve{x}_k \approx (T \diamond p)(k) \cdot x_k$ where $T : \mathbb{R}^{dn_\mathrm{p}} \to \mathbb{R}^{\hat{n}_\mathrm{x} \times n_\mathrm{x}}$ is a state transformation

Equivalent realization:

$$\begin{aligned} \breve{x}_{k+1} &= (\tilde{A}_\mathrm{e} \diamond p)(k)\breve{x}_k + (\tilde{B}_\mathrm{e} \diamond p)(k)u_k + (K_\mathrm{e} \diamond p)(k)y_k, \\ y_k &= (C_\mathrm{e} \diamond p)(k)\breve{x}_k + (D_\mathrm{e} \diamond p)(k)u_k + e_k, \end{aligned}$$

where $T\tilde{A} = \tilde{A}_\mathrm{e} T$, $T\tilde{B} = \tilde{B}_\mathrm{e}$, $TK = K_\mathrm{e}$, $C = C_\mathrm{e} T$, and $D = D_\mathrm{e}$

Potential price to be paid

Free choice of state basis can easily lead to the state of an equivalent LPV-SS representation with dynamic dependence.

Nonparametric estimation of M

The concept:

Given the extended data set $\breve{\mathcal{D}} = \{u_k, y_k, \breve{x}_k, p_k\}_{k=1}^{N}$, estimation of the matrix functions in

$$\begin{aligned} \breve{x}_{k+1} &= (\tilde{A}_\mathrm{e} \diamond p)(k)\breve{x}_k + (\tilde{B}_\mathrm{e} \diamond p)(k)u_k + (K_\mathrm{e} \diamond p)(k)y_k, + \xi_k \\ y_k &= (C_\mathrm{e} \diamond p)(k)\breve{x}_k + D(p_k)u_k + \varphi_k, \end{aligned}$$

$$\tilde{M}(\cdot) = \begin{bmatrix} \tilde{A}_\mathrm{e}(\cdot) & \tilde{B}_\mathrm{e}(\cdot) & K_\mathrm{e}(\cdot) \\ C_\mathrm{e}(\cdot) & D(\cdot) & I \end{bmatrix} : \mathbb{P}^d \to \mathbb{R}^{(n_\mathrm{x}+n_\mathrm{y})\times(n_\mathrm{x}+n_\mathrm{u}+n_\mathrm{y})}$$

Corresponds to a non-parametric regression problem.

RKHS solutions: Empirical Bayesian via GP or LS-SVM
(dependency can be restricted to static to reduce model complexity)

The concept:

Given the extended data set $\breve{\mathcal{D}} = \{u_k, y_k, \breve{x}_k, p_k\}_{k=1}^{N}$, estimation of the matrix functions in

$$\breve{x}_{k+1} = (\tilde{A}_\mathrm{e} \diamond p)(k)\breve{x}_k + (\tilde{B}_\mathrm{e} \diamond p)(k)u_k + (K_\mathrm{e} \diamond p)(k)y_k, + \xi_k$$
$$y_k = (C_\mathrm{e} \diamond p)(k)\breve{x}_k + D(p_k)u_k + \varphi_k,$$

$$\tilde{M}(\cdot) = \begin{bmatrix} \tilde{A}_\mathrm{e}(\cdot) & \tilde{B}_\mathrm{e}(\cdot) & K_\mathrm{e}(\cdot) \\ C_\mathrm{e}(\cdot) & D(\cdot) & I \end{bmatrix} : \mathbb{P}^d \to \mathbb{R}^{(n_\mathrm{x}+n_\mathrm{y})\times(n_\mathrm{x}+n_\mathrm{u}+n_\mathrm{y})}$$

Corresponds to a non-parametric regression problem.

RKHS solutions: Empirical Bayesian via GP or LS-SVM
(dependency can be restricted to static to reduce model complexity)

Tuning of Hyper-parameters

Let θ be the collection of hyper-parameters:

- θ_{c} for KCCA state estimation: κ and kernel coefficients
- θ_{s} for estimation of $\tilde{M}$ by GP/LS-SVM: regularization parameters and kernel coefficients

Method 1: Cross validation via

$$\mathrm{BFR}(\theta) = 100\% \cdot \max\left(1 - \frac{\|y_k - \hat{y}_k(\theta)\|_2}{\|y_k - \bar{y}\|_2}, 0\right)$$

where $\hat{y}_k$ is the predicted or simulated response of the estimated model.

Method 2: Marginalized likelihood (only for level 2)

$$\log \bar{p}(Y|\breve{X}, U, P, \theta_{\mathrm{s}}) = -\frac{1}{2}\left(\sum_{i=1}^{n_{\mathrm{y}}} Y_i \Xi_i^{-1} Y_i^\top + \log|\Xi_i|\right) - \frac{1}{2}\left(\sum_{i=1}^{n_{\mathrm{x}}} \breve{X}_i \Omega_i^{-1} \breve{X}_i^\top + \log|\Omega_i|\right) - \frac{N}{2}\log 2\pi,$$

where the sub-kernel matrices Ξ_i and Ω_i are defined as $\Xi_i = \Xi + \psi_i^{-1} I_N$, $\Omega_i = \Omega + \gamma_i^{-1} I_N$.

Numerical examples

Example 1: Academic example

$$x_{k+1} = A(p_k)x_k + B(p_k)u_k + K(p_k)e_k,$$
$$y_k = C(p_k)x_k + e_k.$$

$$A(p_k) = \begin{bmatrix} \mathrm{sat}(p_k) & 1 & 0 & 0 \\ \frac{1}{2} & \frac{p_k^3}{8} & \frac{4}{10} & \frac{1}{5} \\ \frac{3}{10} & 0 & \frac{p_k^2}{5} & \frac{1}{8} \\ 0 & 0 & \frac{1}{2} & \frac{1}{5} \end{bmatrix}, \qquad B(p_k) = \begin{bmatrix} \frac{p_k^4}{5} & 0 & \frac{1}{5} & 0 \end{bmatrix}^\top,$$

$$K(p_k) = \begin{bmatrix} \frac{\tanh(p_k)}{3p_k} & 0 \\ 0 & 0 \\ 0 & \sin(2\pi p_k) + \cos(2\pi p_k) \\ 0 & 1 \end{bmatrix}, \qquad C(p_k) = \begin{bmatrix} \frac{p_k^2}{5} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$\mathrm{sat}(p_k)$ is a saturation function with limits at ± 0.5 and unity slope; $\mathbb{P} = [-1, +1]$

Data generation:

- $x_0 = [0\ 0\ 0\ 0]^\top$ (initial condition)
- $u_k \sim \mathcal{U}(-1, 1)$ and $p_k = \sin(0.3k)$
- $e_k \sim \mathcal{N}(0, I\sigma_\mathrm{e}^2)$, σ_e^2 chosen to guarantee a 20dB *signal-to-noise ratio* (SNR)
- $\mathcal{D}_N = \{u_k, y_k, p_k\}_{k=1}^{N}$ with $N = 1100$
- The data is divided into 800 and 300 samples for estimation $\mathcal{D}_{800}^{\text{est}}$ and validation $\mathcal{D}_{300}^{\text{val}}$

Identification via KCCA based LS-SVM for LPV-SS:

- future & past window size of $d = 4$
- Polynomial kernel (order 4) in KCCA
- RBF kernels in estimation of M (uniform choice of σ)
- CV tuning of hyper parameters θ

Data generation:

- $x_0 = [0\ 0\ 0\ 0]^\top$ (initial condition)
- $u_k \sim \mathcal{U}(-1, 1)$ and $p_k = \sin(0.3k)$
- $e_k \sim \mathcal{N}(0, I\sigma_{\mathrm{e}}^2)$, σ_{e}^2 chosen to guarantee a 20dB *signal-to-noise ratio* (SNR)
- $\mathcal{D}_N = \{u_k, y_k, p_k\}_{k=1}^N$ with $N = 1100$
- The data is divided into 800 and 300 samples for estimation $\mathcal{D}_{800}^{\text{est}}$ and validation $\mathcal{D}_{300}^{\text{val}}$

Identification via KCCA based LS-SVM for LPV-SS:

- future & past window size of $d = 4$
- Polynomial kernel (order 4) in KCCA
- RBF kernels in estimation of M (uniform choice of σ)
- CV tuning of hyper parameters θ

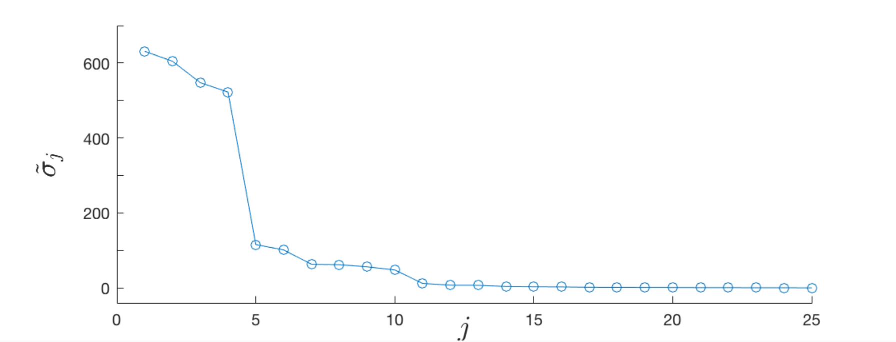


Figure 1: Singular values in the SVD based KCCA with poly-kernel.

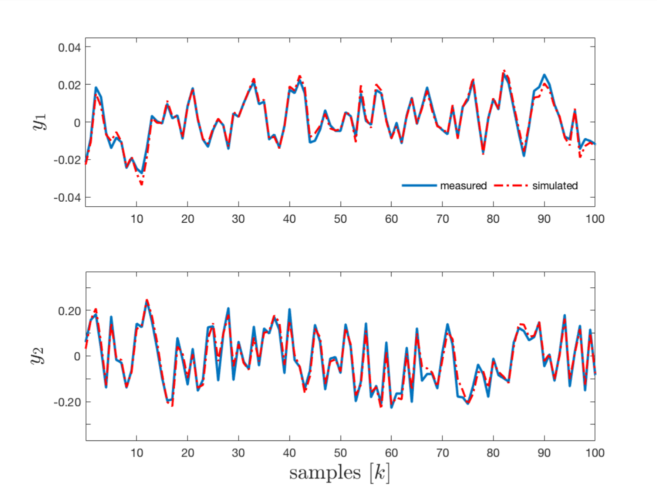


Figure 2: Validation results: simulated response computed on $\mathcal{D}_{300}^{\text{val}}$ (red) and the noise free response of the original system (blue), i.e., noise free y associated with $\mathcal{D}_{300}^{\text{val}}$.

Table 1: Monte-Carlo simulation results for Example 1.

	$\hat{n}$	Mean (BFR %)	Std. (BFR %)
SNR 25dB	4	85.15	1.12
	8	87.03	0.751
SNR 20dB	4	83.91	0.911
	8	86.31	0.022
	9	86.03	1.015

Sate order estimation

Inaccuracy of kernel selection and sub-optimality of hyper-parameter choice leads to increased state order of the estimate (well-known phenomenon in realization theory).

Example 2: CSTR simulation example

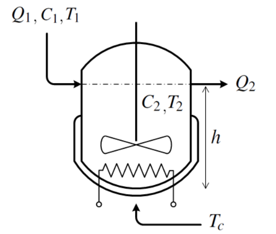


Figure 3: An ideal continuous stirred tank reactor.

First principles based dynamics:

$$\dot{T}_2 = \frac{Q_1}{V}(T_1 - T_2) - \frac{U_{\mathrm{HE}}}{A_{\mathrm{HE}}}(T_2 - T_{\mathrm{c}}) + \frac{\Delta H k_0}{\rho c_\rho} e^{-\frac{E_A}{RT_2}} C_2,$$

$$\dot{C}_2 = \frac{Q_1}{V}(C_1 - C_2) - k_0 e^{-\frac{E_A}{RT_2}} C_2,$$

Input	Output	Scheduling
$u = \begin{bmatrix} Q_1 & T_c \end{bmatrix}^\top$	$y = T_2$	$p = C_1$

Data generation:

- $u =$ PRBS with $\pm$ 10% of the nominal values
- OE scenario: Gaussian white noise is added such that 25dB SNR is maintained for the output T_2

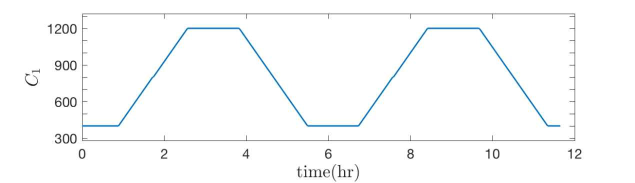


Figure 4: Scheduling trajectory $C_1(\text{kg/m}^3)$ for Example 2.

Identification via KCCA based LS-SVM for LPV-SS:

- future & past window size of $d = 4$
- Polynomial kernel (order 4) in KCCA
- RBF kernels in estimation of M (uniform choice of σ)
- CV tuning of hyper parameters θ

Data generation:

- $u =$ PRBS with $\pm$ 10% of the nominal values
- OE scenario: Gaussian white noise is added such that 25dB SNR is maintained for the output T_2

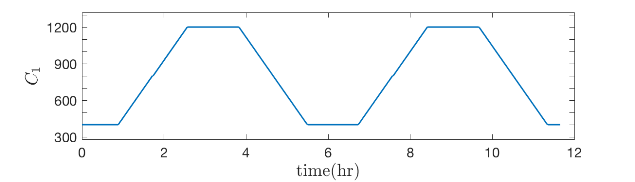


Figure 4: Scheduling trajectory $C_1(\mathrm{kg/m^3})$ for Example 2.

Identification via KCCA based LS-SVM for LPV-SS:

- future & past window size of $d = 4$
- Polynomial kernel (order 4) in KCCA
- RBF kernels in estimation of M (uniform choice of σ)
- CV tuning of hyper parameters θ

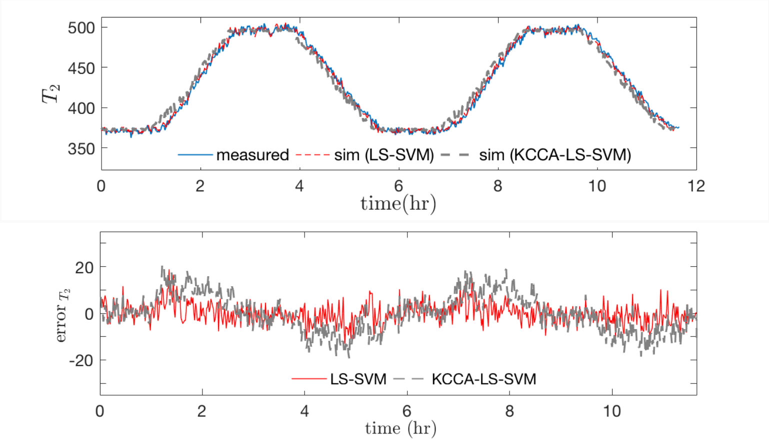


Figure 5: Example 2: Validation results for CSTR output temperature $T_2(^{\circ}\mathrm{C})$ using LS-SVM-based identification with (gray) and without (red) full states measurements.

Table 2: CSTR output fitness simulation results.

	SNR (dB)	BFR (%)
LS-SVM (full states measurement)	25	86.72
KCCA-based LS-SVM	25	83.23

Conclusion and outlook

Summary

- Applied KCCA to avoid curse of dimensionality in LPV subspace identification.
- Proof of concept via simulation studies.
- Work in progress. Suggestions are welcome.

Outlook

- Implement automatic kernel tuning and explore kernel selection for state transfer maps
- Investigate stochastic realization in the LPV case (NL dependence, closed-loop case, etc.)
- KCCA-SVM as a non-parametric estimate of the system to seed parametric PEM-SS