Decoupling Multivariate Polynomials

Philippe Dreesen Mariya Ishteva Johan Schoukens

Vrije Universiteit Brussel Department ELEC

philippe.dreesen@vub.ac.be

BIL 2014 *Workshop on Data-driven Modeling Methods and Applications*

July 14–15, 2014 KU Leuven, BE

Outline

1. Motivation and Problem Statement
2. Method: algorithm and results
3. Smoothness-inducing regularization
4. Conclusions and Future Work

Outline

1 Motivation and Problem Statement

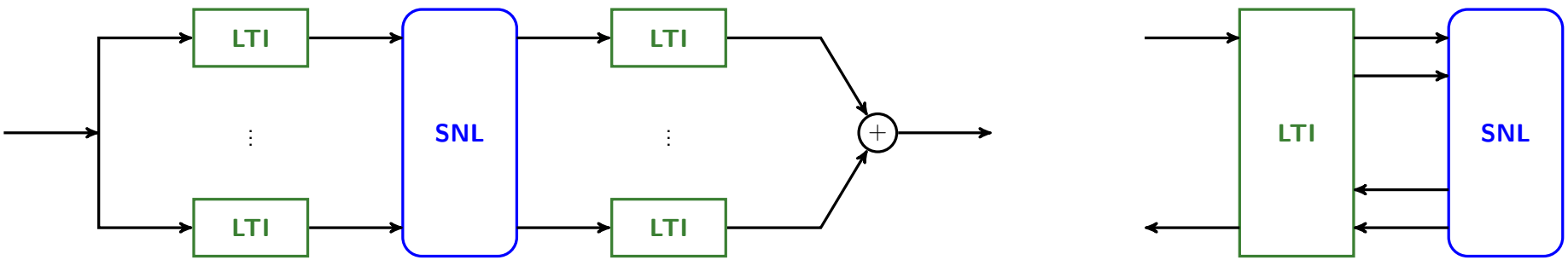
2 Method: algorithm and results

3 Smoothness-inducing regularization

4 Conclusions and Future Work

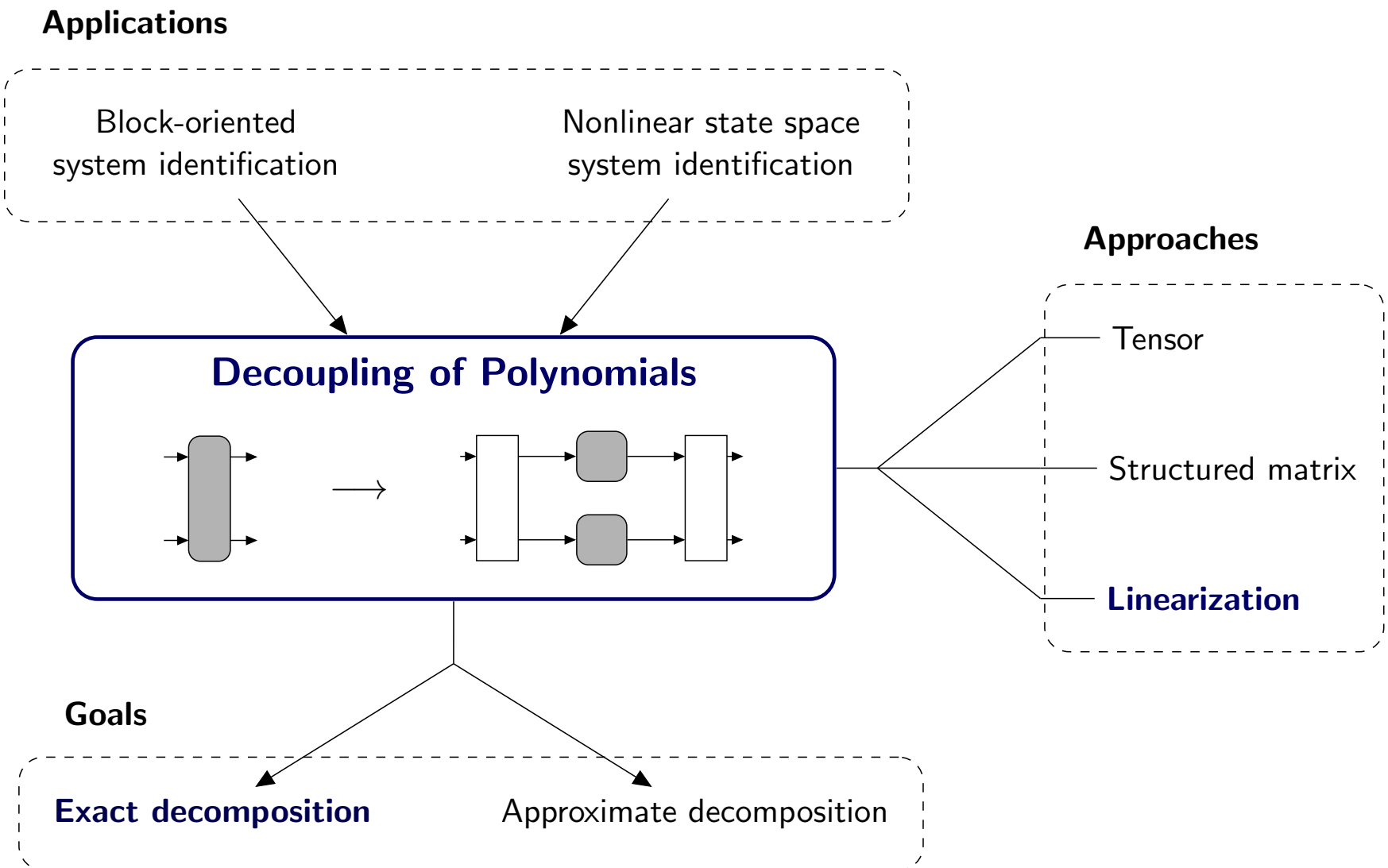
Motivation

- From linear to **nonlinear system identification**
  - block-oriented, NL state space models
  - Volterra kernel series
  - NARX, NARMAX, ...
  - (recurrent) NN, SVM, ...

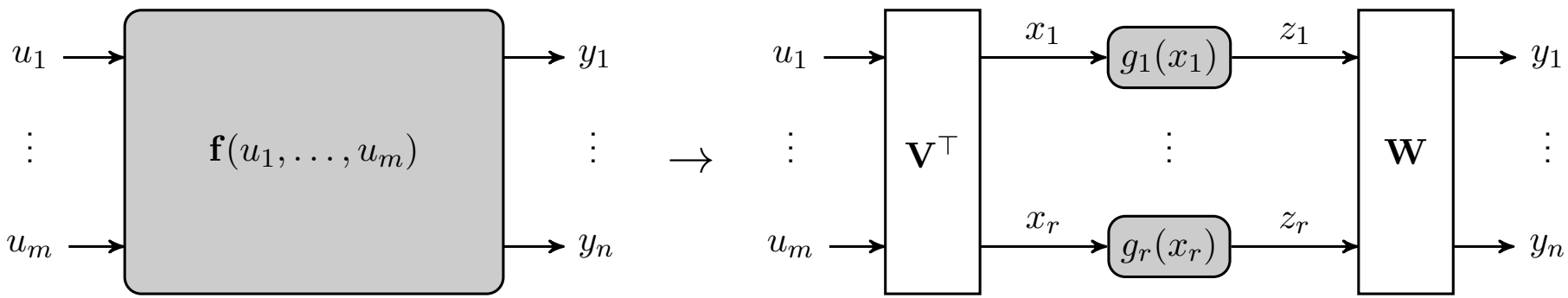


- Obtain physical insight
- Simplification (number of parameters)
- Get handle on nonlinearities

Problem Statement



Problem Statement



A simple example

The 'coupled' polynomials $f_1(u_1, u_2)$ and $f_2(u_1, u_2)$

$$f_1(u_1, u_2) = 3{u_1}^3 - 27{u_1}^2 u_2 - 22{u_1}^2 + 81 u_1 {u_2}^2 + 20 u_1 u_2 - 3u_1 - 81{u_2}^3 - 58{u_2}^2 - 33u_2$$

$$f_2(u_1, u_2) = {u_1}^3 - 9{u_1}^2 u_2 + 6{u_1}^2 + 27 u_1 {u_2}^2 + 20 u_1 u_2 + 9u_1 - 27{u_2}^3 - 16{u_2}^2 - 6u_2$$

can be decoupled as

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} {x_1}^3 - 2{x_1}^2 + 3x_1 \\ 2{x_2}^2 - 3x_2 \end{bmatrix}}_{g_i(x_i)},$$

with

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -3 \\ -2 & -1 \end{bmatrix}}_{\mathbf{V}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Outline

1 Motivation and Problem Statement

2 **Method: algorithm and results**

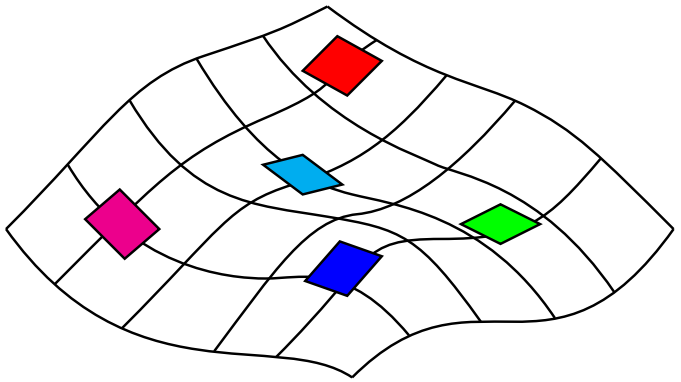
3 Smoothness-inducing regularization

4 Conclusions and Future Work

Linearization approach: main idea

- **Small-signal model** of $\mathbf{f}(\mathbf{u})$ in $\mathbf{u}^\star$

$$\mathbf{u}^\star = \left\{ \mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \mathbf{u}^{(4)}, \mathbf{u}^{(5)}, \ldots \right\}$$



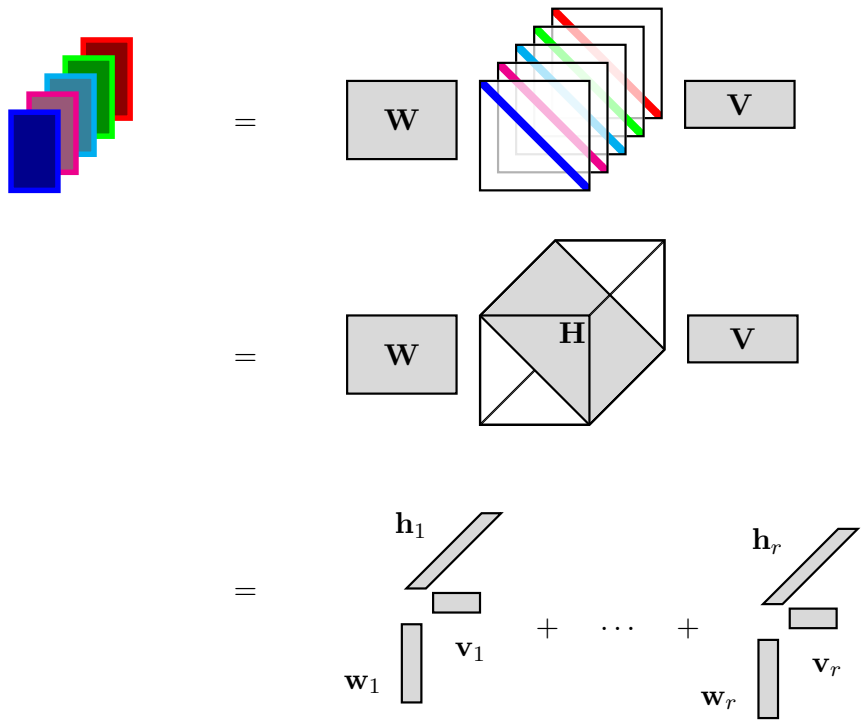
- Jacobian of $\mathbf{f}(\mathbf{u})$

$$\begin{aligned} \mathbf{f}(\mathbf{u}) &= \mathbf{W} \left[g_i(\mathbf{v}_i^\top \mathbf{u}) \right] \\ &\Downarrow \\ \mathbf{J}_\mathbf{f}(\mathbf{u}) &= \mathbf{W} \begin{bmatrix} g_1'(\mathbf{v}_1^\top \mathbf{u}) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_r'(\mathbf{v}_r^\top \mathbf{u}) \end{bmatrix} \mathbf{V}^\top \end{aligned}$$

$$\rightarrow \mathbf{J}_\mathbf{f}^{(1)}, \mathbf{J}_\mathbf{f}^{(2)}, \mathbf{J}_\mathbf{f}^{(3)}, \mathbf{J}_\mathbf{f}^{(4)}, \mathbf{J}_\mathbf{f}^{(5)}, \ldots$$

Simultaneous Matrix Diagonalization: tensors

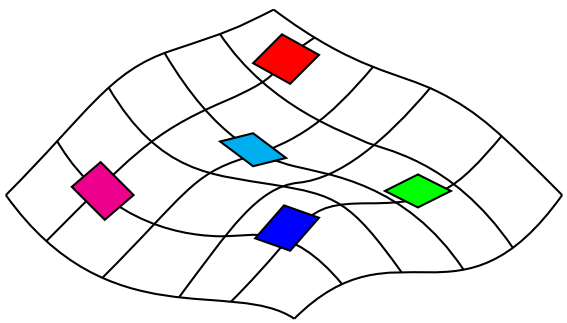
- Tensor of Jacobians $\mathcal{J}_{\mathbf{f}} = \left\{ \mathbf{J}_{\mathbf{f}}^{(1)}, \mathbf{J}_{\mathbf{f}}^{(2)}, \mathbf{J}_{\mathbf{f}}^{(3)}, \mathbf{J}_{\mathbf{f}}^{(4)}, \mathbf{J}_{\mathbf{f}}^{(5)}, \ldots \right\}$
- **Simultaneous matrix diagonalization**: sum of rank-one terms
- **Tool:** Canonical Polyadic Decomposition (CPD)
  (CanDecomp, PARAFAC)

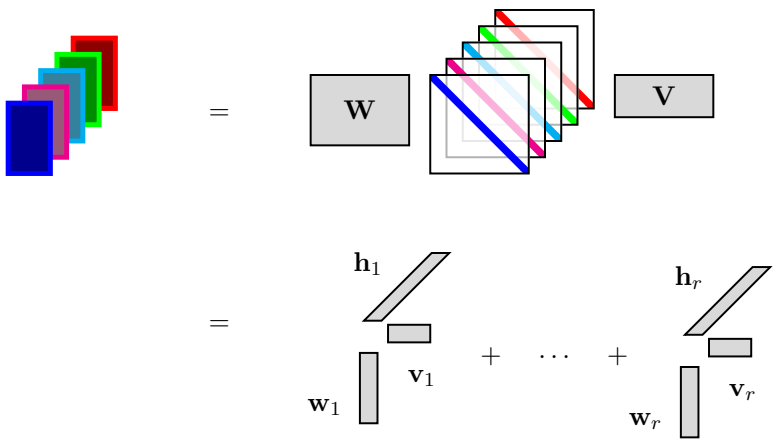


Algorithm

1. Construct tensor of Jacobians $\mathcal{J}_{\mathbf{f}} = \left\{ \mathbf{J}_{\mathbf{f}}^{(1)}, \mathbf{J}_{\mathbf{f}}^{(2)}, \mathbf{J}_{\mathbf{f}}^{(3)}, \mathbf{J}_{\mathbf{f}}^{(4)}, \mathbf{J}_{\mathbf{f}}^{(5)}, \ldots \right\}$
2. CPD of $\mathcal{J}_{\mathbf{f}}$ gives $\mathbf{W}$, $\mathbf{V}$ and $\mathbf{H}$
3. Retrieve $g_i(\cdot)$ from $h_{ij} = g_j'(\mathbf{v}_j^\top \mathbf{u}^{(i)})$
   (interpolation, integration, linear system)
4. Complete structure:

$$\mathbf{f}(\mathbf{u}) = \mathbf{W} \left[g_i(\mathbf{v}_i^\top \mathbf{u}) \right]$$





Remarks

- Choice of operating points $\mathbf{u}^\star = \left\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \mathbf{u}^{(4)}, \mathbf{u}^{(5)}, \ldots\right\}$
- Identifiability: 'unique' under mild conditions (Kruskal):

$$\min(m, r) + \min(n, r) \geq r + 2$$

("OK if there are not too many internal branches")

- Determining r is open problem
- Structure of $g(\cdot)$ not used!

Simple example revisited

The 'coupled' polynomials $f_1(u_1, u_2)$ and $f_2(u_1, u_2)$

$$f_1(u_1, u_2) = 3{u_1}^3 - 27{u_1}^2 u_2 - 22{u_1}^2 + 81 u_1 {u_2}^2 + 20 u_1 u_2 - 3u_1 - 81{u_2}^3 - 58{u_2}^2 - 33u_2$$

$$f_2(u_1, u_2) = {u_1}^3 - 9{u_1}^2 u_2 + 6{u_1}^2 + 27 u_1 {u_2}^2 + 20 u_1 u_2 + 9u_1 - 27{u_2}^3 - 16{u_2}^2 - 6u_2$$

can be decoupled as

$$\begin{bmatrix} f_1(u_1, u_2) \\ f_2(u_1, u_2) \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} {x_1}^3 - 2{x_1}^2 + 3x_1 \\ 2{x_2}^2 - 3x_2 \end{bmatrix}}_{g_i(x_i)},$$

with

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -3 \\ -2 & -1 \end{bmatrix}}_{\mathbf{V}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Simple example revisited

$$f_1(u_1, u_2) = 3{u_1}^3 - 27{u_1}^2 u_2 - 22{u_1}^2 + 81 u_1 {u_2}^2 + 20 u_1 u_2 - 3u_1 - 81{u_2}^3 - 58{u_2}^2 - 33u_2$$

$$f_2(u_1, u_2) = {u_1}^3 - 9{u_1}^2 u_2 + 6{u_1}^2 + 27 u_1 {u_2}^2 + 20 u_1 u_2 + 9u_1 - 27{u_2}^3 - 16{u_2}^2 - 6u_2$$

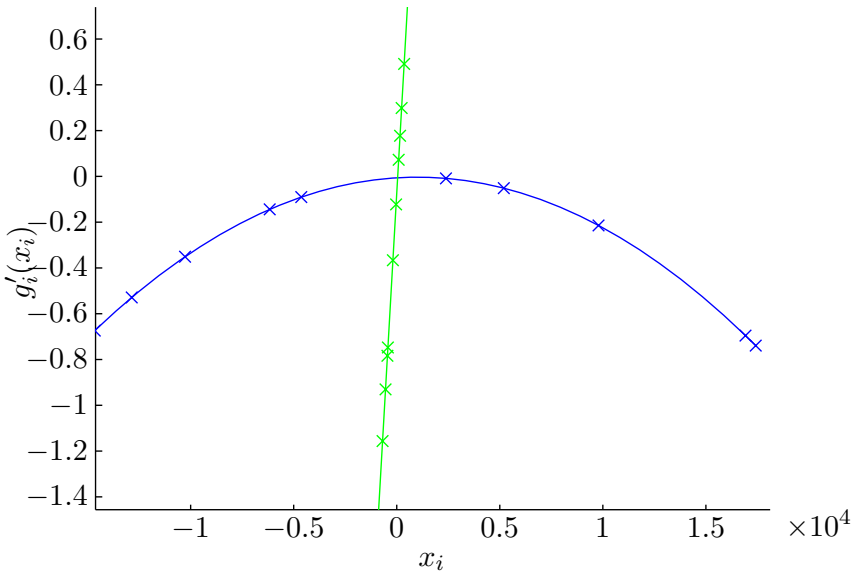
1. Collect $N = 50$ data points $\mathbf{u}^{(i)}$, $\mathbf{y}^{(i)}$ and Jacobians $\mathbf{J_f}(\mathbf{u}^{(i)})$
2. Perform CPD of tensor of Jacobians $\mathcal{J}_\mathbf{f}$ to find $\mathbf{W}$, $\mathbf{V}$ and $\mathbf{H}$
   (relative error on tensor is $\mathbf{7.27 \times 10^{-17}}$)

$$\mathbf{W}_0 = \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} \quad \Rightarrow \quad \hat{\mathbf{W}} = \begin{bmatrix} 0.9514 & 1.2649 \\ 0.3171 & -0.6324 \end{bmatrix}$$

$$\mathbf{V}_0 = \begin{bmatrix} 1 & -2 \\ -3 & -1 \end{bmatrix} \quad \Rightarrow \quad \hat{\mathbf{V}} = \begin{bmatrix} 0.9669 & -0.0854 \\ -2.9008 & -0.0427 \end{bmatrix}$$

Simple example revisited

3. Extract $g_i(\cdot)$ from $h_{ij} = g'_j(\mathbf{v}_j^\top \mathbf{u}^{(i)})$

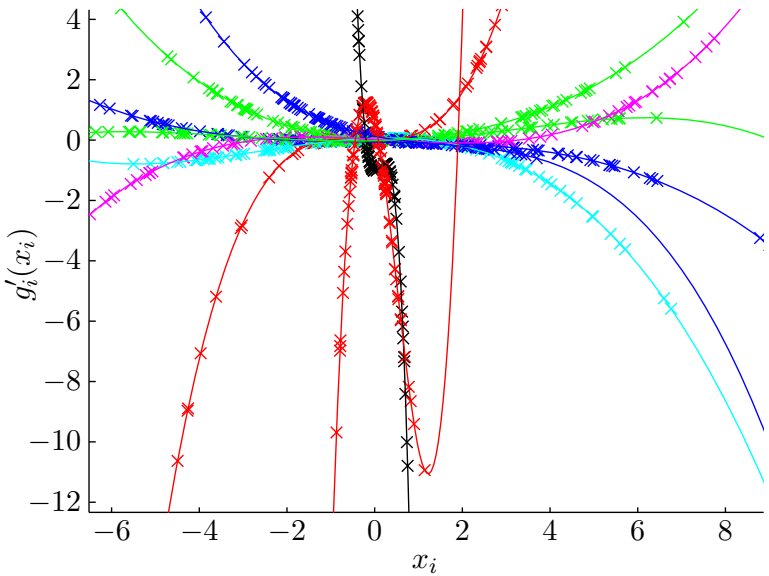


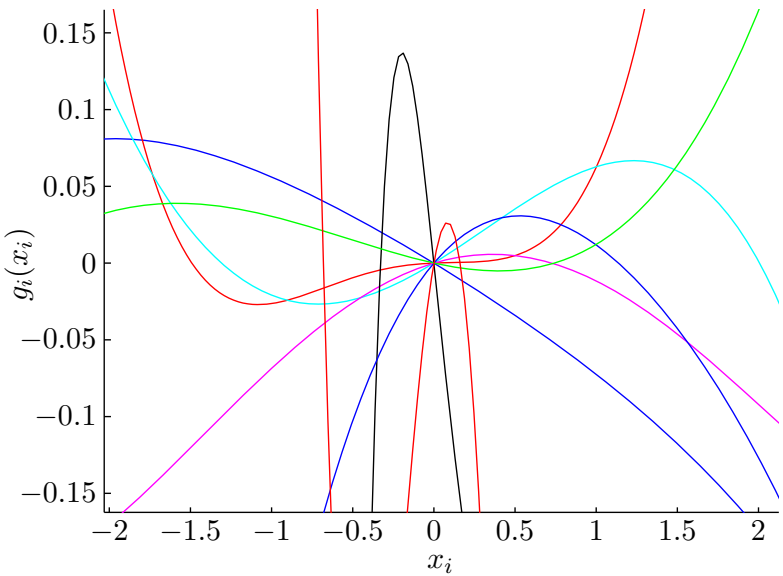
4. Reconstruct $\mathbf{f}(\mathbf{u}) = \mathbf{W}\left[g_i(\mathbf{v}_i^\top \mathbf{u})\right]$

   (mean relative coeff-wise error on f_i is $\mathbf{6.77 \times 10^{-14}}$)

Larger example

- $\mathbf{5} \to \mathbf{9} \to \mathbf{7}$, degree $d = 4$
- Kruskal satisfied
- $N = 100$ linearization points
- relative error on tensor is $\mathbf{6.05 \times 10^{-9}}$

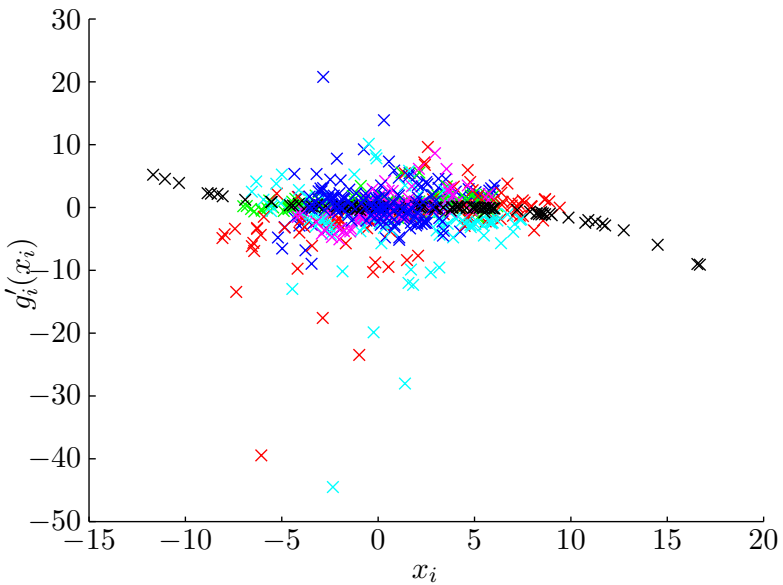


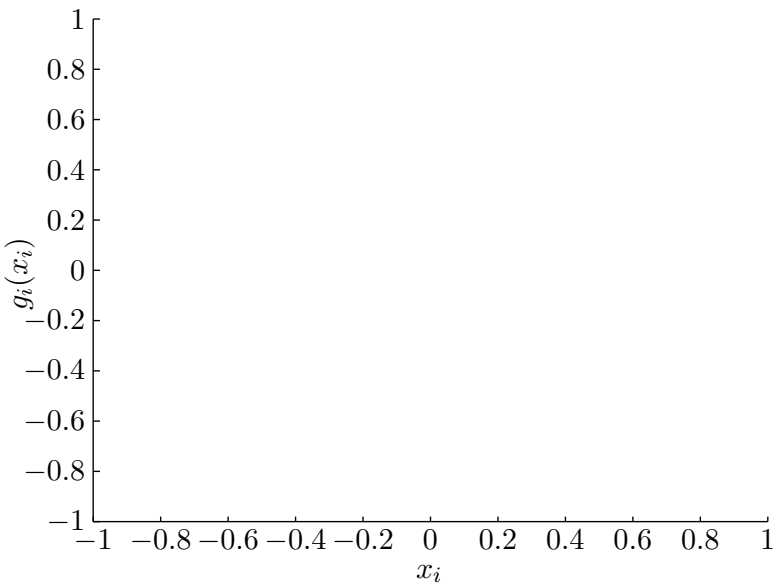


- mean relative coeff-wise error on f_i is $\mathbf{1.13 \times 10^{-7}}$

Many internal branches: identifiability issue

- $\mathbf{3} \to \mathbf{7} \to \mathbf{3}$, degree $d = 3$
- Kruskal **not satisfied**
- $N = 100$ linearization points
- relative error on tensor is $\mathbf{4.91 \times 10^{-16}}$





- reconstruction of $g_i(\cdot)$ not possible

Outline

1 Motivation and Problem Statement

2 Method: algorithm and results

3 **Smoothness-inducing regularization**

4 Conclusions and Future Work

Many internal branches: identifiability issue

- Kruskal fails to guarantee **uniqueness** if $r \geq m + n - 2$
- Smoothness-inducing regularization scheme:

$$\mathbf{W}^{+} \leftarrow \min_{\mathbf{W}} \left\| \mathbf{T}_{(1)} - \mathbf{W}(\mathbf{V} \odot \mathbf{H})^{\top} \right\|_F^2$$

$$\mathbf{V}^{+} \leftarrow \min_{\mathbf{V}} \left\| \mathbf{T}_{(2)} - \mathbf{V}(\mathbf{W} \odot \mathbf{H})^{\top} \right\|_F^2$$

$$\mathbf{H}^{+} \leftarrow \min_{\mathbf{H}} \left\| \mathbf{T}_{(3)} - \mathbf{H}(\mathbf{V} \odot \mathbf{W})^{\top} \right\|_F^2 + \lambda \sum_{i=1}^{r} \|\mathbf{D}\mathbf{S}_i(\mathbf{H}\mathbf{e}_i)\|_2^2$$

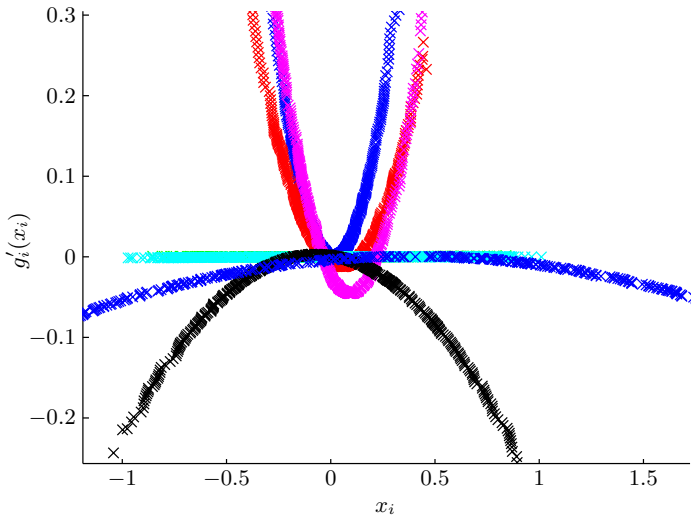
with

- $\mathbf{T}_{(i)}$: mode-i tensor unfolding
- $\mathbf{D}$: smoothing operator
- $\mathbf{S}_i$: permutation matrix (dep. on $\mathbf{V}$)
- $\mathbf{e}_i$: ith unit vector

Silverbox Benchmark

- $\mathbf{3} \to \mathbf{r} \to \mathbf{3}$ SNL in PNLSS
- nonlinear identification benchmark
- cubic nonlinearity

r	ordinary CPD		smooth ALS	
	$\frac{\|\mathcal{J}-\hat{\mathcal{J}}\|}{\|\mathcal{J}\|}$	$\frac{\Delta c}{c}$	$\frac{\|\mathcal{J}-\hat{\mathcal{J}}\|}{\|\mathcal{J}\|}$	$\frac{\Delta c}{c}$
1	0.74	0.64	0.78	0.64
2	0.45	0.37	0.47	0.36
3	0.26	0.23	0.30	0.20
4	0.17	0.28	0.23	0.16
5	0.09	0.52	0.13	0.08
6	0.03	0.81	0.08	0.05
7	0.00	0.92	0.03	0.02



- tuning λ: 3-fold cross-validation, 3 repetitions
- ordinary CPD loses (quasi-)smoothness for $r \geq 4$
- rank estimation $r = 7$

Outline

1 Motivation and Problem Statement

2 Method: algorithm and results

3 Smoothness-inducing regularization

4 Conclusions and Future Work

Conclusions and Future Work

- Decoupling of polynomials: linearization approach
- Simultaneous matrix diagonalization
- Imposing smoothness via regularization

Conclusions and Future Work

- Decoupling of polynomials: linearization approach
- Simultaneous matrix diagonalization
- Imposing smoothness via regularization

- Exact vs. approximate decoupling
  - noisy case
  - best rank-r approximation
  - weighted objective criterion

Thank you for your attention!