Polynomial Chaos Expansion: Part II

Thierry A. Mara
thierry.mara@univ-reunion.fr

University of Reunion

12th SAMO Summer School
June 24-28 2024
University of Parma

Outline

How to compute the PCE coefficients $\boldsymbol{a}_{\mathcal{A}}$?

$\boldsymbol{a}_{\mathcal{A}}$? Given $(\boldsymbol{X}, \boldsymbol{y}, \mathcal{A})$

Assumptions: Given an independent input sample $\boldsymbol{X}$ of size N, the associated vector of output responses $\boldsymbol{y}$, and the sparse PCE structure $\mathcal{A}$,
Objective: Find $\boldsymbol{a}_{\mathcal{A}}$ such that,

$$y \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$\Leftrightarrow \boldsymbol{y} \approx \boldsymbol{\Psi}_{\mathcal{A}} \boldsymbol{a}_{\mathcal{A}}$$

$\boldsymbol{a}_{\mathcal{A}}$? Given $(\boldsymbol{X}, \boldsymbol{y}, \mathcal{A})$

Assumptions: Given an independent input sample $\boldsymbol{X}$ of size N, the associated vector of output responses $\boldsymbol{y}$, and the sparse PCE structure $\mathcal{A}$,

Objective: Find $\boldsymbol{a}_{\mathcal{A}}$ such that,

$$y \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$\Leftrightarrow \boldsymbol{y} \approx \boldsymbol{\Psi}_{\mathcal{A}} \boldsymbol{a}_{\mathcal{A}}$$

Well-posing the problem: We assume that

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \varepsilon$$

$$\Leftrightarrow \boldsymbol{y} = \boldsymbol{\Psi}_{\mathcal{A}} \boldsymbol{a}_{\mathcal{A}} + \boldsymbol{\varepsilon}$$

$\boldsymbol{a}_{\mathcal{A}}$? Given $(\boldsymbol{X}, \boldsymbol{y}, \mathcal{A})$

Assumptions: Given an independent input sample $\boldsymbol{X}$ of size N, the associated vector of output responses $\boldsymbol{y}$, and the sparse PCE structure $\mathcal{A}$,

Objective: Find $\boldsymbol{a}_{\mathcal{A}}$ such that,

$$y \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$$

$$\Leftrightarrow \boldsymbol{y} \approx \boldsymbol{\Psi}_{\mathcal{A}} \boldsymbol{a}_{\mathcal{A}}$$

Well-posing the problem: We assume that

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \varepsilon$$

$$\Leftrightarrow \boldsymbol{y} = \boldsymbol{\Psi}_{\mathcal{A}} \boldsymbol{a}_{\mathcal{A}} + \boldsymbol{\varepsilon}$$

To proceed we must define the RV ε

Method I:
The Maximum Likelihood Estimate

MLE

OLS: If it is assumed that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ then the **Ordinary Least Squares** estimate of $\boldsymbol{a}_\mathcal{A}$ is the best solution, that is

$$\boldsymbol{a}_\mathcal{A}^{MLE} = \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A}\right)^{-1} \boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{y} \tag{1}$$

MLE

<u>OLS</u>: If it is assumed that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ then the **Ordinary Least Squares** estimate of $\boldsymbol{a}_\mathcal{A}$ is the best solution, that is

$$\boldsymbol{a}_\mathcal{A}^{MLE} = \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A}\right)^{-1} \boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{y} \tag{1}$$

<u>ML</u>: Actually in a probabilistic framework the OLS solution is also the **Maximum Likelihood** estimate and the joint pdf of $\boldsymbol{a}_\mathcal{A}$ can even be estimated,

$$\hat{\boldsymbol{a}}_\mathcal{A} | \mathcal{A}, \boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{a}_\mathcal{A}^{MLE}, \hat{\boldsymbol{C}}_{\boldsymbol{aa}}\right) \tag{2}$$

$$\hat{\boldsymbol{C}}_{\boldsymbol{aa}} = \hat{\sigma}_\varepsilon^2 \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A}\right)^{-1} \tag{4}$$

MLE

OLS: If it is assumed that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ then the **Ordinary Least Squares** estimate of $\boldsymbol{a}_\mathcal{A}$ is the best solution, that is

$$\boldsymbol{a}_\mathcal{A}^{MLE} = \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A}\right)^{-1} \boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{y} \tag{1}$$

ML: Actually in a probabilistic framework the OLS solution is also the **Maximum Likelihood** estimate and the joint pdf of $\boldsymbol{a}_\mathcal{A}$ can even be estimated,

$$\hat{\boldsymbol{a}}_\mathcal{A} | \mathcal{A}, \boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{a}_\mathcal{A}^{MLE}, \hat{\boldsymbol{C}}_{\boldsymbol{aa}}\right) \tag{2}$$

$$\hat{\sigma}_\varepsilon^2 | \mathcal{A}, \boldsymbol{y}, \hat{\boldsymbol{a}}_\mathcal{A} \sim \Gamma\left(\frac{N+2}{2}, \frac{\left(\boldsymbol{y} - \boldsymbol{\Psi}_\mathcal{A}\hat{\boldsymbol{a}}_\mathcal{A}\right)^T \left(\boldsymbol{y} - \boldsymbol{\Psi}_\mathcal{A}\hat{\boldsymbol{a}}_\mathcal{A}\right)}{2}\right) \tag{3}$$

$$\hat{\boldsymbol{C}}_{\boldsymbol{aa}} = \hat{\sigma}_\varepsilon^2 \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A}\right)^{-1} \tag{4}$$

$N > \mathrm{Card}\,(\mathcal{A})$ being the size of the sample.

MLE

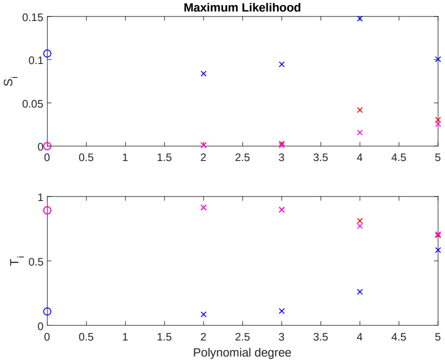
Example: Let us consider the simple model
$y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$

Let noise the output data to simulate a non-polynomial model.
Set, $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}\left[y\right])$

MLE

Example: Let us consider the simple model
$y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$

on-polynomial model.



Solution: Divergence at high-polynomial degrees = Overfitting

Conclusion

Truncated full MLE PCE is hampered by the following drawbacks

- A full PCE of polynomial degree $p_{\mathcal{A}}$ contains $\mathrm{Card}\left(\mathcal{A}\right) = \frac{(d+p_{\mathcal{A}})!}{d!p_{\mathcal{A}}!}$ elements
- The higher $p_{\mathcal{A}}$ the higher the risk of overfitting
- Choice of $p_{\mathcal{A}}$ not obvious

Solution: Constrain the PCE coefficients and/or Reduce $\mathrm{Card}\left(\mathcal{A}\right)$ by keeping only the relevant monomials with the help of a Model Selection Criterion (sparse PCE)

Method II:
The Maximum A Posteriori

MAP

MAP: If we assume that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and we impose that $\boldsymbol{a}_\mathcal{A} \sim \mathcal{N}\left(0, \boldsymbol{C}_{\boldsymbol{aa}}\right)$, then the Maximum a Posteriori estimate of $\boldsymbol{a}_\mathcal{A}$ is the best solution, that is

$$\boldsymbol{a}_\mathcal{A}^{MAP} = \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A} + \hat{\sigma}_\varepsilon^2 \boldsymbol{C}_{\boldsymbol{aa}}^{-1}\right)^{-1} \boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{y} \tag{5}$$

MAP

<u>MAP</u>: If we assume that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and we impose that $\boldsymbol{a}_\mathcal{A} \sim \mathcal{N}\left(0, \boldsymbol{C}_{\boldsymbol{aa}}\right)$, then the Maximum a Posteriori estimate of $\boldsymbol{a}_\mathcal{A}$ is the best solution, that is

$$\boldsymbol{a}_\mathcal{A}^{MAP} = \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A} + \hat{\sigma}_\varepsilon^2 \boldsymbol{C}_{\boldsymbol{aa}}^{-1}\right)^{-1} \boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{y} \tag{5}$$

<u>ML</u>: Actually in a Bayesian framework the joint pdf of $\boldsymbol{a}_\mathcal{A}$ is,

$$\hat{\boldsymbol{a}}_\mathcal{A} | \mathcal{A}, \boldsymbol{y}, \hat{\sigma}_\varepsilon^2 \sim \mathcal{N}\left(\boldsymbol{a}_\mathcal{A}^{MAP}, \hat{\sigma}_\varepsilon^2 \left(\boldsymbol{\Psi}_\mathcal{A}^T \boldsymbol{\Psi}_\mathcal{A} + \hat{\sigma}_\varepsilon^2 \boldsymbol{C}_{\boldsymbol{aa}}^{-1}\right)^{-1}\right) \tag{6}$$

$$\hat{\sigma}_\varepsilon^2 | \mathcal{A}, \boldsymbol{y}, \hat{\boldsymbol{a}}_\mathcal{A} \sim \Gamma\left(\frac{N+2}{2}, \frac{\left(\boldsymbol{y} - \boldsymbol{\Psi}_\mathcal{A} \hat{\boldsymbol{a}}_\mathcal{A}\right)^T \left(\boldsymbol{y} - \boldsymbol{\Psi}_\mathcal{A} \hat{\boldsymbol{a}}_\mathcal{A}\right)}{2}\right) \tag{7}$$

MAP

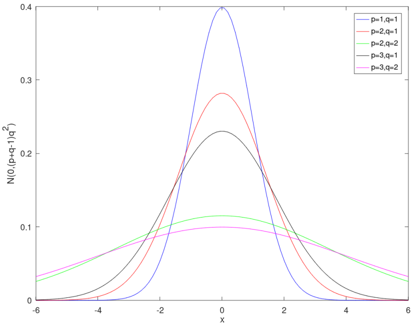
Example: Let us consider the simple model $y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$ with the noisy data $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}\left[y\right])$

Setting $\boldsymbol{C_{aa}} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma^2_{\mathrm{Card}(\mathcal{A})})$ with $\sigma_k^2 = (p_k + q_k - 1)q_k^2$ compute the MAP estimates of the Sobol' indices

MAP

Example: Let us consider the simple model
$y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$ with the noisy data $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}[y])$

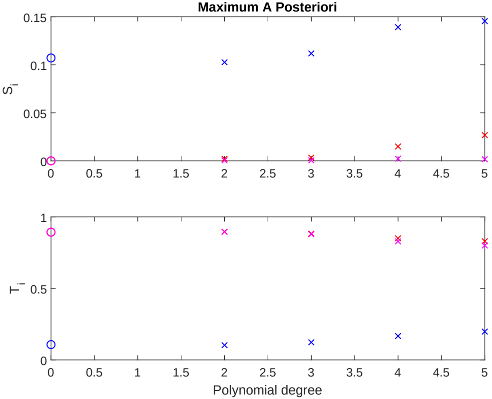
Setting $\boldsymbol{C_{aa}} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma^2_{\mathrm{Card}(\mathcal{A})})$ with $\sigma_k^2 = (p_k + q_k - 1)q_k^2$ compute the MAP estimates of the Sobol' indices



MAP

Example: Let us consider the simple model
$y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$ with the noisy data $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}[y])$

Setting $\boldsymbol{C_{aa}} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma^2_{\mathrm{Card}(\mathcal{A})})$ with $\sigma_k^2 = (p_k + q_k - 1)q_k^2$ compute the MAP estimates of the Sobol' indices



Conclusion

- Imposing a Gaussian prior on the PCE coefficients $\boldsymbol{a}_{\mathcal{A}}$ can alleviate the overfitting
- Nevertheless, the choice of the prior Covariance matrix $\boldsymbol{C}_{\boldsymbol{aa}}$ is arbitrary
- There are still $\mathrm{Card}\left(\mathcal{A}\right) = \frac{(d+p_{\mathcal{A}})!}{d!p_{\mathcal{A}}!}$ coefficients most of which are non-significant

Solution: Reduce $\mathrm{Card}\left(\mathcal{A}\right)$ by keeping only the relevant monomials with the help of a Model Selection Criterion

Model Selection
The Kashyap Information Criterion

KIC

Given a subset $\mathcal{A} \subset \mathbb{N}^d$ and assuming

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \varepsilon$$

with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, MLE and MAP are two possible estimates of the PC coefficients.

However, if $\mathrm{Card}\,(\mathcal{A})$ is too large MLE and MAP are not very accurate (overfitting).

KIC

Given a subset $\mathcal{A} \subset \mathbb{N}^d$ and assuming

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) + \varepsilon$$

with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, MLE and MAP are two possible estimates of the PC coefficients.

However, if $\mathrm{Card}\,(\mathcal{A})$ is too large MLE and MAP are not very accurate (overfitting).

Question: How can we optimize the choice of $\mathcal{A}$?

Answer: Use a Model Selection Criterion

KIC

Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_M\}$ be M competing PC representations. We denote by $\hat{\boldsymbol{a}}_k$ the estimated vector of PC coefficients associated with $\mathcal{A}_k$ and $P_k = \mathrm{Card}\,(\mathcal{A}_k)$. The former being obtained either with MLE or MAP.

KIC

Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_M\}$ be M competing PC representations. We denote by $\hat{\boldsymbol{a}}_k$ the estimated vector of PC coefficients associated with $\mathcal{A}_k$ and $P_k = \mathrm{Card}\,(\mathcal{A}_k)$. The former being obtained either with MLE or MAP.

According to Rass & Raftery (1995) the best model is the one maximizing the Bayesian Model Evidence defined as,

$$p(y|\mathcal{A}) = \int p(y|\boldsymbol{a}, \mathcal{A}) p(\boldsymbol{a}|\mathcal{A}) \mathrm{d}\boldsymbol{a} \tag{8}$$

KIC

Let $\{\mathcal{A}_1, \dots, \mathcal{A}_M\}$ be M competing PC representations. We denote by $\hat{\boldsymbol{a}}_k$ the estimated vector of PC coefficients associated with $\mathcal{A}_k$ and $P_k = \mathrm{Card}\,(\mathcal{A}_k)$. The former being obtained either with MLE or MAP.

According to Rass & Raftery (1995) the best model is the one maximizing the Bayesian Model Evidence defined as,

$$p(y|\mathcal{A}) = \int p(y|\boldsymbol{a}, \mathcal{A}) p(\boldsymbol{a}|\mathcal{A}) \mathrm{d}\boldsymbol{a} \tag{8}$$

In can be proven that, under our assumptions, this integral becomes (Schroniger et al. 2014),

$$p(\boldsymbol{y}|\mathcal{A}_k) = p(\boldsymbol{y}|\hat{\boldsymbol{a}}_{\boldsymbol{k}}, \mathcal{A}_k) p(\hat{\boldsymbol{a}}_{\boldsymbol{k}}|\mathcal{A}_k) (2\pi)^{P_k/2} |\hat{\boldsymbol{C}}_{\boldsymbol{a_k a_k}}|^{1/2} \tag{9}$$

where $\hat{\boldsymbol{C}}_{\boldsymbol{a_k a_k}}$ is the (posterior) covariance of $\hat{\boldsymbol{a}}_{\boldsymbol{k}}$.

KIC

The Kashyap information criterion is defined as the deviance (Kashyap, 1982), namely,

$$KIC_{\mathcal{A}_k} = -2\ln\left(p(\boldsymbol{y}, \hat{\boldsymbol{a}}_{\boldsymbol{k}}|\mathcal{A}_k)\right) - P_k \ln(2\pi) - \ln\left(|\hat{\boldsymbol{C}}_{\boldsymbol{a_k}\,\boldsymbol{a_k}}|\right) \qquad (10)$$

KIC

The Kashyap information criterion is defined as the deviance (Kashyap, 1982), namely,

$$KIC_{\mathcal{A}_k} = -2\ln\left(p(\boldsymbol{y}, \hat{\boldsymbol{a}}_{\boldsymbol{k}}|\mathcal{A}_k)\right) - P_k \ln(2\pi) - \ln\left(|\hat{\boldsymbol{C}}_{\boldsymbol{a_k}\boldsymbol{a_k}}|\right) \tag{10}$$

The $KIC^{MLE}_{\mathcal{A}_k}$ is obtained when the MLE approach is used, in that case (see Eq.(2)),

$$KIC^{MLE}_{\mathcal{A}_k} = N\ln\hat{\sigma}^2_{\varepsilon_k} - P_k\ln(2\pi) - \ln|\hat{\sigma}^2_{\varepsilon_k}\left(\boldsymbol{\Psi}^T_{\mathcal{A}_k}\boldsymbol{\Psi}_{\mathcal{A}_k}\right)^{-1}| \tag{11}$$

KIC

The Kashyap information criterion is defined as the deviance (Kashyap, 1982), namely,

$$KIC_{\mathcal{A}_k} = -2\ln\left(p(\boldsymbol{y}, \hat{\boldsymbol{a}}_{\boldsymbol{k}}|\mathcal{A}_k)\right) - P_k\ln(2\pi) - \ln\left(|\hat{\boldsymbol{C}}_{\boldsymbol{a_k a_k}}|\right) \tag{10}$$

The $KIC^{MLE}_{\mathcal{A}_k}$ is obtained when the MLE approach is used, in that case (see Eq.(2)),

$$KIC^{MLE}_{\mathcal{A}_k} = N\ln\hat{\sigma}^2_{\varepsilon_k} - P_k\ln(2\pi) - \ln|\hat{\sigma}^2_{\varepsilon_k}\left(\boldsymbol{\Psi}^T_{\mathcal{A}_k}\boldsymbol{\Psi}_{\mathcal{A}_k}\right)^{-1}| \tag{11}$$

The $KIC^{MAP}_{\mathcal{A}_k}$ is obtained when the MAP approach is used, in that case (see Eq.(5))

$$\begin{aligned} KIC^{MAP}_{\mathcal{A}_k} = & N\ln\hat{\sigma}^2_{\varepsilon_k} + \ln|\boldsymbol{C}_{\boldsymbol{a_k a_k}}| + \hat{\boldsymbol{a}}^T_k\boldsymbol{C}^{-1}_{\boldsymbol{a_k a_k}}\hat{\boldsymbol{a}}_k \\ & - \ln|\hat{\sigma}^2_{\varepsilon_k}\left(\boldsymbol{\Psi}^T_{\mathcal{A}_k}\boldsymbol{\Psi}_{\mathcal{A}_k} + \hat{\sigma}^2_{\varepsilon_k}\boldsymbol{C}^{-1}_{\boldsymbol{a_k a_k}}\right)^{-1}| \end{aligned} \tag{12}$$

Bayesian Sparse PCE
The Algorithm (partially)

For more details see Shao et al. 2017

BSPCE

Algorithm of the BSPCE

Given $(\boldsymbol{X}, \boldsymbol{y})$, with $\boldsymbol{y}$ being standardised (i.e., mean=0, var=1),

1. *Initialization*: Set initial polynomial degree $p = 4$ and interaction level $q = 2$ (or $p = 2$, $q = 1$ if d high). Create the initial subset $\mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$.

BSPCE

Algorithm of the BSPCE

Given $(\boldsymbol{X}, \boldsymbol{y})$, with $\boldsymbol{y}$ being standardised (i.e., mean=0, var=1),

1. *Initialization*: Set initial polynomial degree $p = 4$ and interaction level $q = 2$ (or $p = 2$, $q = 1$ if d high). Create the initial subset $\mathcal{B} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$.
2. Set $k = 0$, $\mathcal{A}_k = \mathcal{B}_k$, and $KIC_k^{MLE} = +\infty$.

BSPCE

Algorithm of the BSPCE

Given $(\boldsymbol{X}, \boldsymbol{y})$, with $\boldsymbol{y}$ being standardised (i.e., mean=0, var=1),

1. *Initialization*: Set initial polynomial degree $p = 4$ and interaction level $q = 2$ (or $p = 2$, $q = 1$ if d high). Create the initial subset $\mathcal{B} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$.
2. Set $k = 0$, $\mathcal{A}_k = \mathcal{B}_k$, and $KIC_k = +\infty$.
3. *Model selection*: Set $k = k + 1$, $\mathcal{A}_k = \mathcal{A}_k \cup \mathcal{B}_k$. Compute $\boldsymbol{a}_{\mathcal{A}_k}$ and $KIC_{\mathcal{A}_k}$. If $KIC_{\mathcal{A}_k} > KIC_{\mathcal{A}_{k-1}}$ remove $\mathcal{B}_k$ from $\mathcal{A}_k$. Resume until $k = \mathrm{Card}(\mathcal{B})$.

BSPCE

Algorithm of the BSPCE

Given $(\boldsymbol{X}, \boldsymbol{y})$, with $\boldsymbol{y}$ being standardised (i.e., mean=0, var=1),

1. *Initialization*: Set initial polynomial degree $p = 4$ and interaction level $q = 2$ (or $p = 2$, $q = 1$ if d high). Create the initial subset $\mathcal{B} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$.
2. Set $k = 0$, $\mathcal{A}_k = \mathcal{B}_k$, and $KIC_k = +\infty$.
3. *Model selection*: Set $k = k + 1$, $\mathcal{A}_k = \mathcal{A}_k \cup \mathcal{B}_k$. Compute $\boldsymbol{a}_{\mathcal{A}_k}$ and $KIC_{\mathcal{A}_k}$. If $KIC_{\mathcal{A}_k} > KIC_{\mathcal{A}_{k-1}}$ remove $\mathcal{B}_k$ from $\mathcal{A}_k$. Resume until $k = \mathrm{Card}(\mathcal{B})$.
4. *Enrichment of $\mathcal{A}_k$ or Stop*: From $\mathcal{A}_k$, get $(p_{\mathcal{A}_k}, q_{\mathcal{A}_k})$. If $p_{\mathcal{A}_k} < (p - 1)$ and $q_{\mathcal{A}_k} < q$, Stop.

BSPCE

Algorithm of the BSPCE

Given $(\boldsymbol{X}, \boldsymbol{y})$, with $\boldsymbol{y}$ being standardised (i.e., mean=0, var=1),

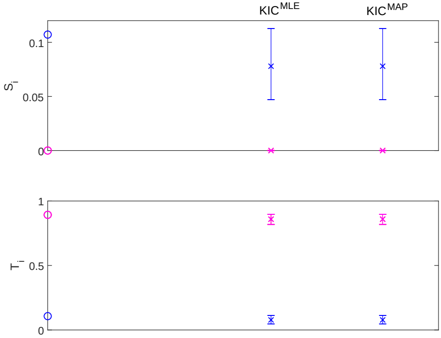
1. *Initialization*: Set initial polynomial degree $p = 4$ and interaction level $q = 2$ (or $p = 2$, $q = 1$ if d high). Create the initial subset $\mathcal{B} = \{\boldsymbol{\alpha} \in \mathbb{N}^d : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$.
2. Set $k = 0$, $\mathcal{A}_k = \mathcal{B}_k$, and $KIC_k = +\infty$.
3. *Model selection*: Set $k = k + 1$, $\mathcal{A}_k = \mathcal{A}_k \cup \mathcal{B}_k$. Compute $\boldsymbol{a}_{\mathcal{A}_k}$ and $KIC_{\mathcal{A}_k}$. If $KIC_{\mathcal{A}_k} > KIC_{\mathcal{A}_{k-1}}$ remove $\mathcal{B}_k$ from $\mathcal{A}_k$. Resume until $k = \mathrm{Card}(\mathcal{B})$.
4. *Enrichment of $\mathcal{A}_k$ or Stop*: From $\mathcal{A}_k$, get $(p_{\mathcal{A}_k}, q_{\mathcal{A}_k})$. If $p_{\mathcal{A}_k} < (p - 1)$ and $q_{\mathcal{A}_k} < q$, Stop. Otherwise, set $\mathcal{B} = \mathcal{A}_k$, enrich $\mathcal{B}$ with PC elements $\in [p_{\mathcal{A}_k}, p_{\mathcal{A}_k} + 2]\&[q_{\mathcal{A}_k}, q_{\mathcal{A}_k} + 1]$ and resume from 2.

BSPCE

Example: Let us consider the simple model $y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$ with the noisy data $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}[y])$

BSPCE

Example: Let us consider the simple model $y = x_1 + 10\left(x_2 - \frac{1}{2}\right)\left(x_3 - \frac{1}{2}\right)$ with $x_j \sim \mathcal{U}(0,1)$ with the noisy data $y_{br} = y + \mathcal{N}(0, 0.05\mathrm{Var}[y])$



References

- Kass, R. E., and A. E. Raftery. (1995), J. Am. Stat. Assoc., 773–795
- Schoniger A. et al. (2014), Water Resour. Res., 9484–9513.
- Kashyap R.L. (1982), IEEE Trans. Pattern Anal. Mach. Intell., 99–104
- Shao Q. at al. (2017), Comput. Methods Appl. Mech. Engrg., 474–496