DISTRIBUTED PREDICTIVE CONTROL BASED ON GAUSSIAN PROCESS MODELS

Alexandra Grancharova

Department of Industrial Automation
University of Chemical Technology and Metallurgy, Sofia, Bulgaria

CT CPNL Scientific Day, Valence, France

CONTENTS

1 Distributed MPC versus centralized MPC

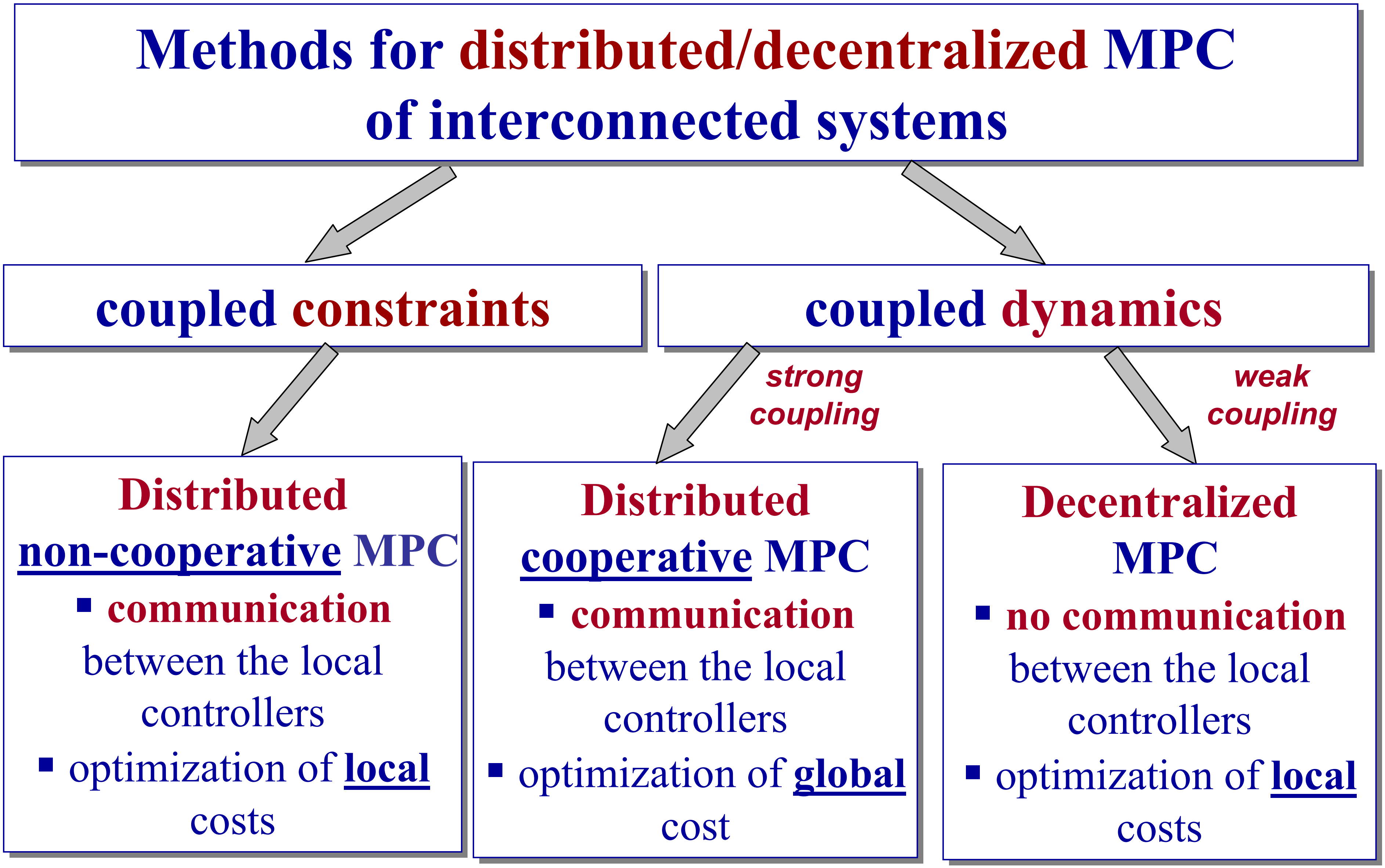
Complexity of the centralized solution of NMPC problems for medium- and large-scale systems

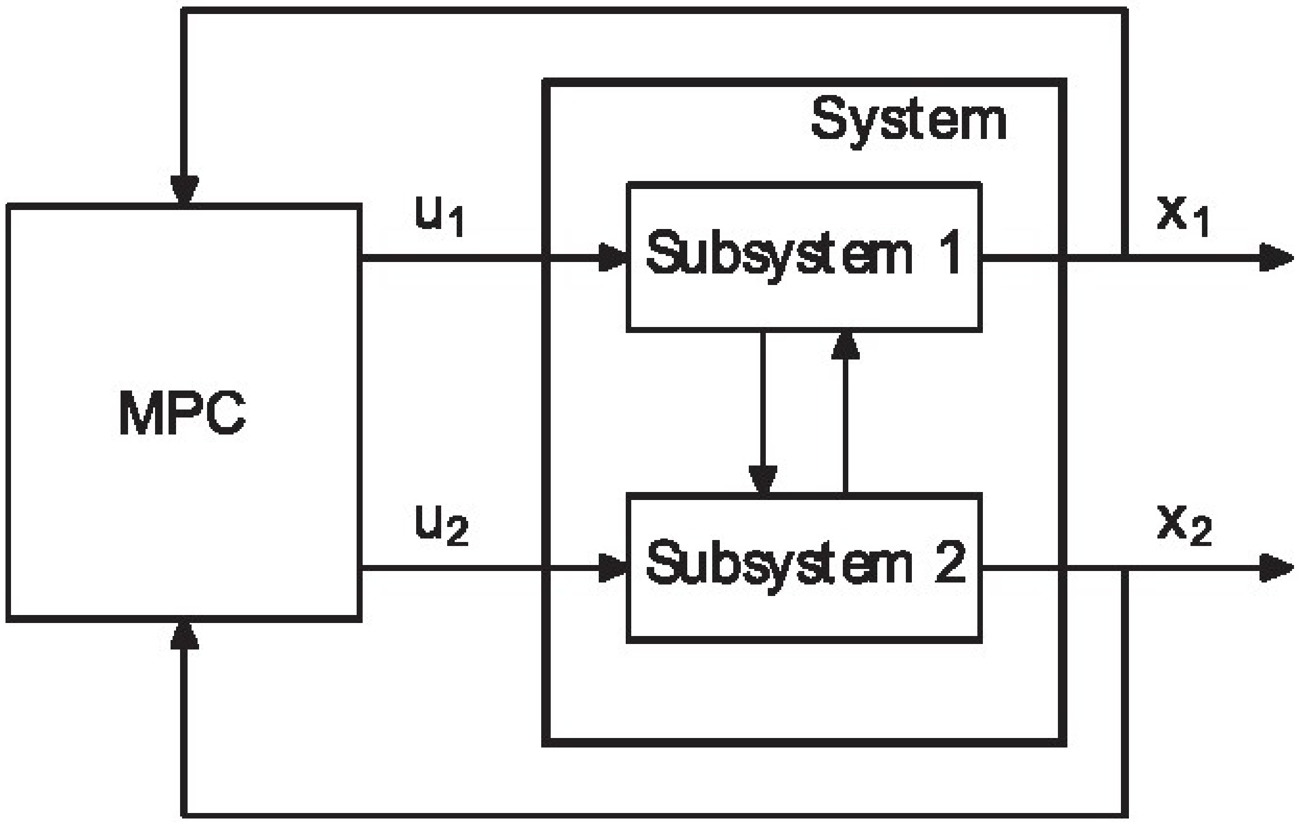
- Complexity of the NLP problem;
- Topology of the plant and data communication;
- Large number of decision variables.

↓

Advantages of the distributed solution of NMPC problems

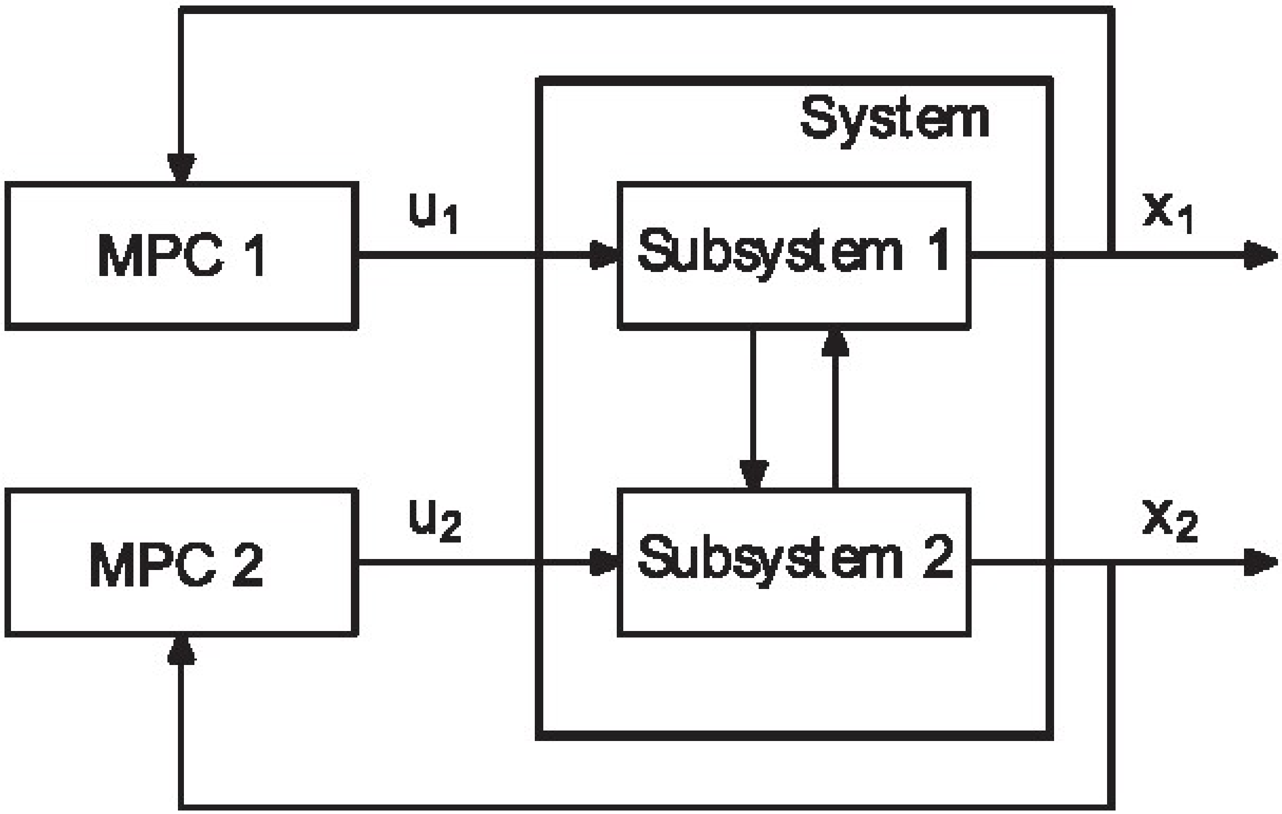
- Reducing the original, large size, optimization problem into a number of smaller and more tractable ones;
- Allows the systems to solve autonomously their local optimization problems (coordination required for global optimization).


Methods for distributed/decentralized MPC of interconnected systems
coupled constraints
coupled dynamics
strong coupling
weak coupling
Distributed non-cooperative MPC
▪ communication between the local controllers
▪ optimization of local costs
Distributed cooperative MPC
▪ communication between the local controllers
▪ optimization of global cost
Decentralized MPC
▪ no communication between the local controllers
▪ optimization of local costs



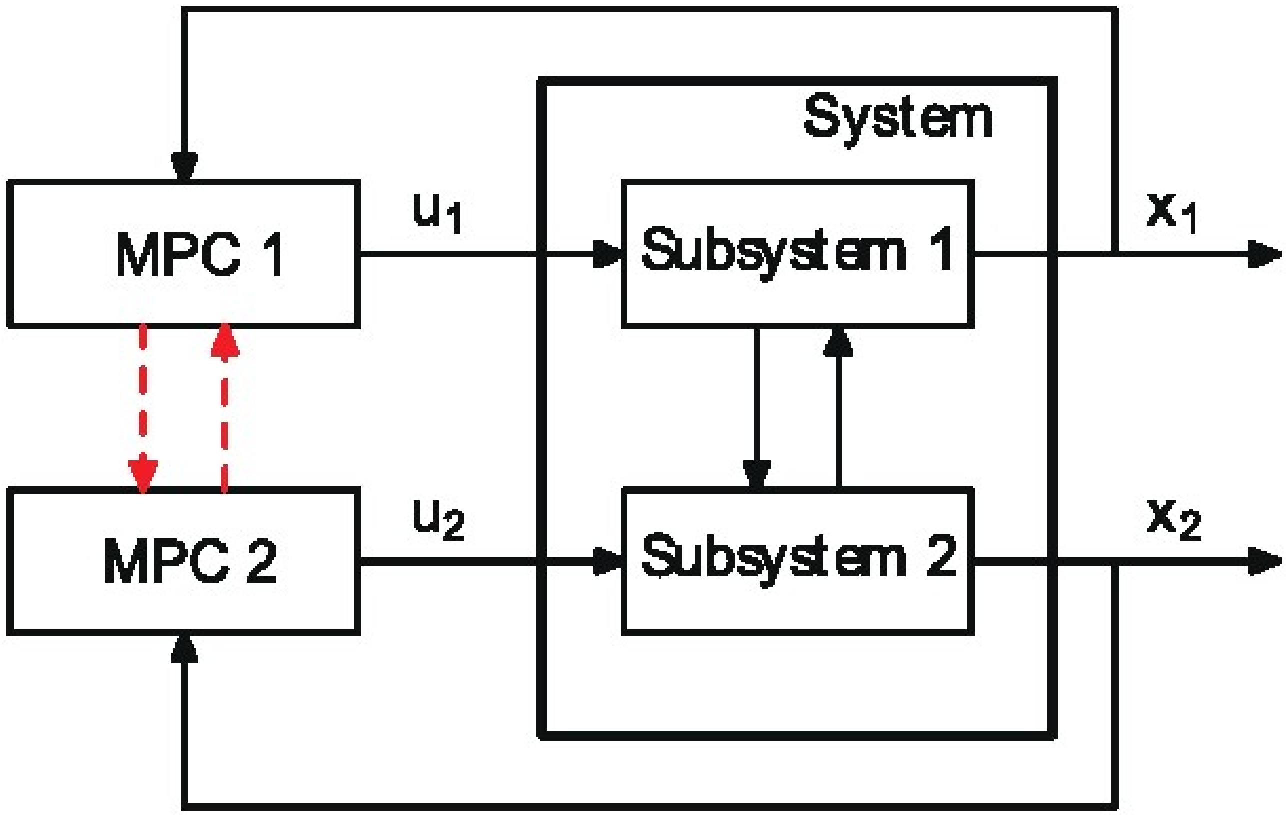
Centralized MPC

(Christofides et al. (2013))



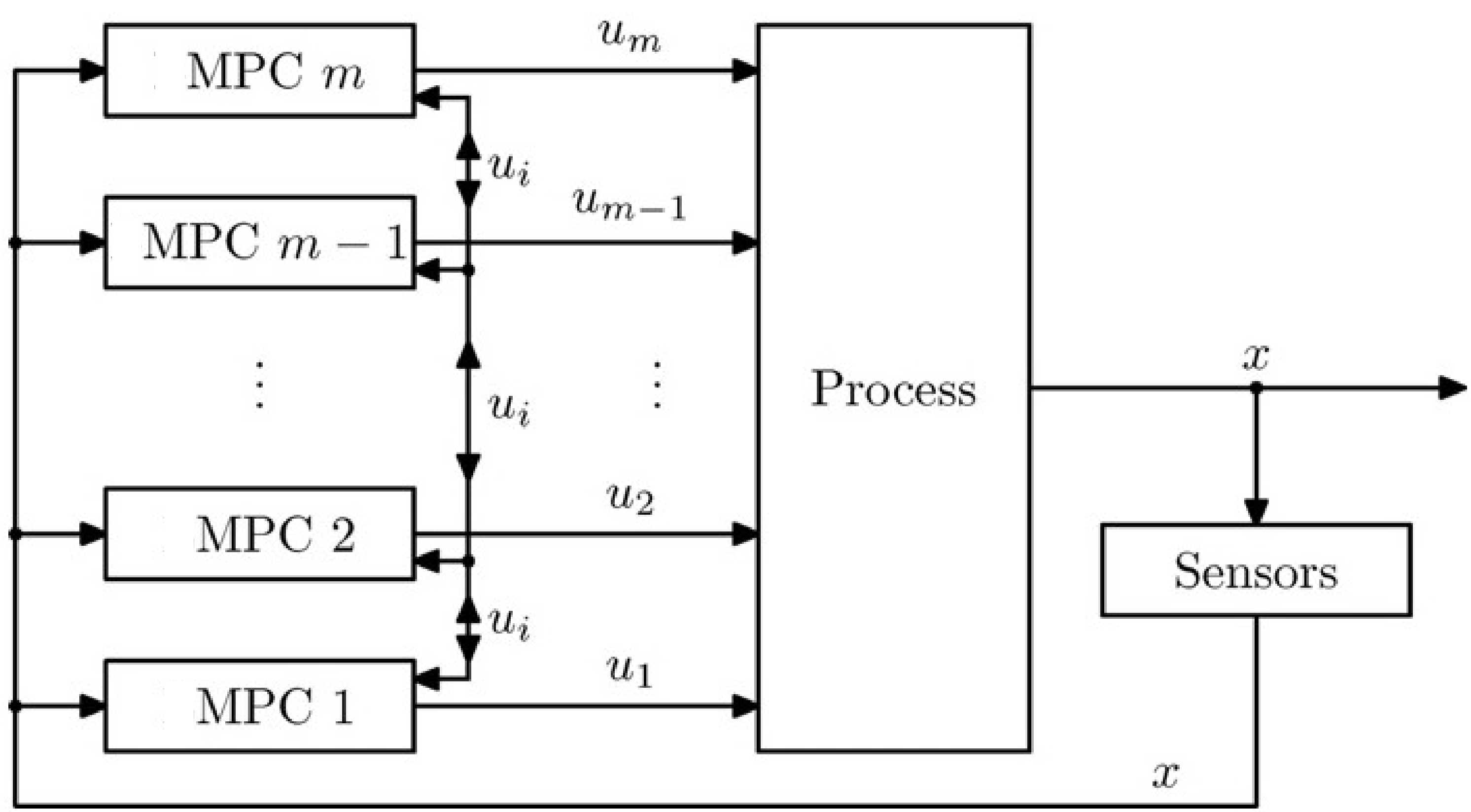
Decentralized MPC

(Christofides et al. (2013))



Distributed non-cooperative MPC

(Christofides et al. (2013))



Distributed cooperative MPC

(Christofides et al. (2013))

Focus

Distributed cooperative MPC for nonlinear interconnected systems:

- **coupled dynamics;**
- **decoupled state and input constraints;**
- **optimizing global cost function.**

2 Application of Gaussian processes to the modeling of complex dynamic systems

Preliminaries on modeling with Gaussian processes
(Rasmussen & Williams, 2006)

$$y = f(z) + \xi$$

$$z \in \mathbb{R}^D \; , \; y \in \mathbb{R} \; , \; \xi - \text{stochastic noise}$$

Gaussian Process
(a collection of random variables which have a joint multivariate Gaussian distribution)

$$y(1), y(2), \ldots, y(L) \sim \mathcal{N}(0, \mathbf{\Sigma})$$

mean *covariance*

Gaussian Process is specified by:

$$\mu(z) \text{ (usually } \mu(z)=0) - \text{mean value}$$

$$\Sigma_{pq} = \mathrm{Cov}(y(p), y(q)) = C(z(p), z(q)) - \text{covariance function}$$

Covariance function :

$$C(z(p), z(q)) = v_1 \exp\left[-\frac{1}{2}\sum_{i=1}^{D} w_i (z_i(p) - z_i(q))^2\right] + v_0 \delta_{pq}$$

'Hyperparameters' of the covariance function :

$$\Theta = [w_1, \ldots, w_D, v_0, v_1]$$

obtained by maximum likelihood optimization

Given a data set :

$$\mathbf{Z} = [z(1), z(2), \ldots, z(L)]^T \text{ , } Y = [y(1), y(2), \ldots, y(L)]^T$$

$$\mathbf{K} \in \mathbb{R}^{L \times L} \text{ – covariance matrix}$$

Estimate the probability distribution of:

$$y^* \mid z^*, (\mathbf{Z}, Y)$$

$$E\{y^*\} = \mu(z^*) = c(z^*)^{\mathrm{T}} \mathbf{K}^{-1} Y$$

$$\mathrm{var}\{y^*\} = \sigma^2(z^*) = c_0(z^*) - c(z^*)^{\mathrm{T}} \mathbf{K}^{-1} c(z^*) + v_0 \qquad \textbf{(1)}$$

$c(z^*) = [C(z(1), z^*), \ldots, C(z(L), z^*)]^{\mathrm{T}}$ – vector of covariances between the test and training cases

$c_0(z^*) = C(z^*, z^*)$ – the covariance between the test input and itself

Gaussian processes can be used for modelling of dynamic systems

↓

Autoregressive (NARX) model :

$$z(t) = [\hat{y}(t-1), \ldots, \hat{y}(t-l), u(t-1), \ldots, u(t-l)]^T$$

$$\hat{y}(t) = f(z(t)) + \eta(t)$$

t – consecutive number of data sample

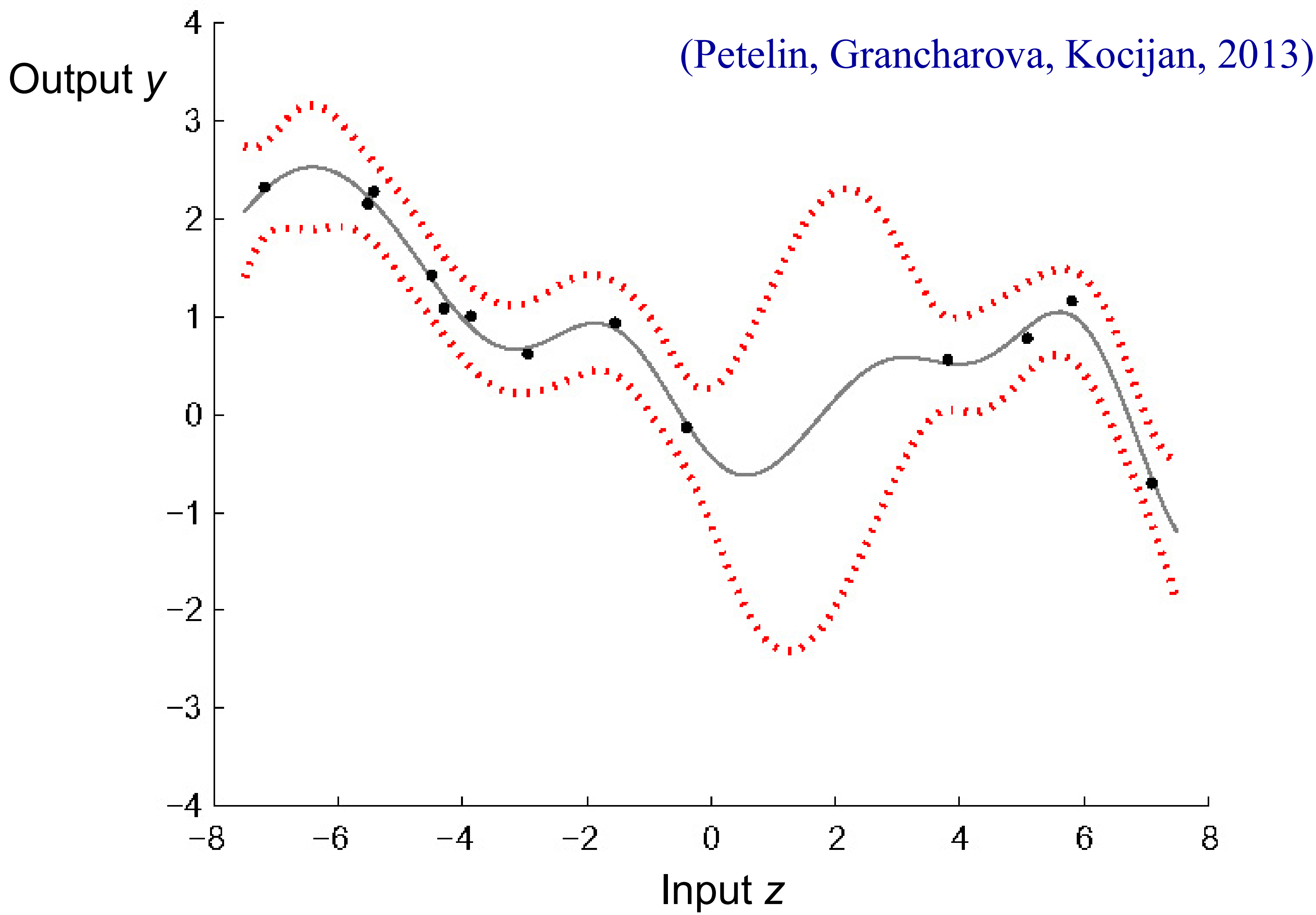
l – given lag

$\eta(t)$ – prediction error

Quality of prediction :

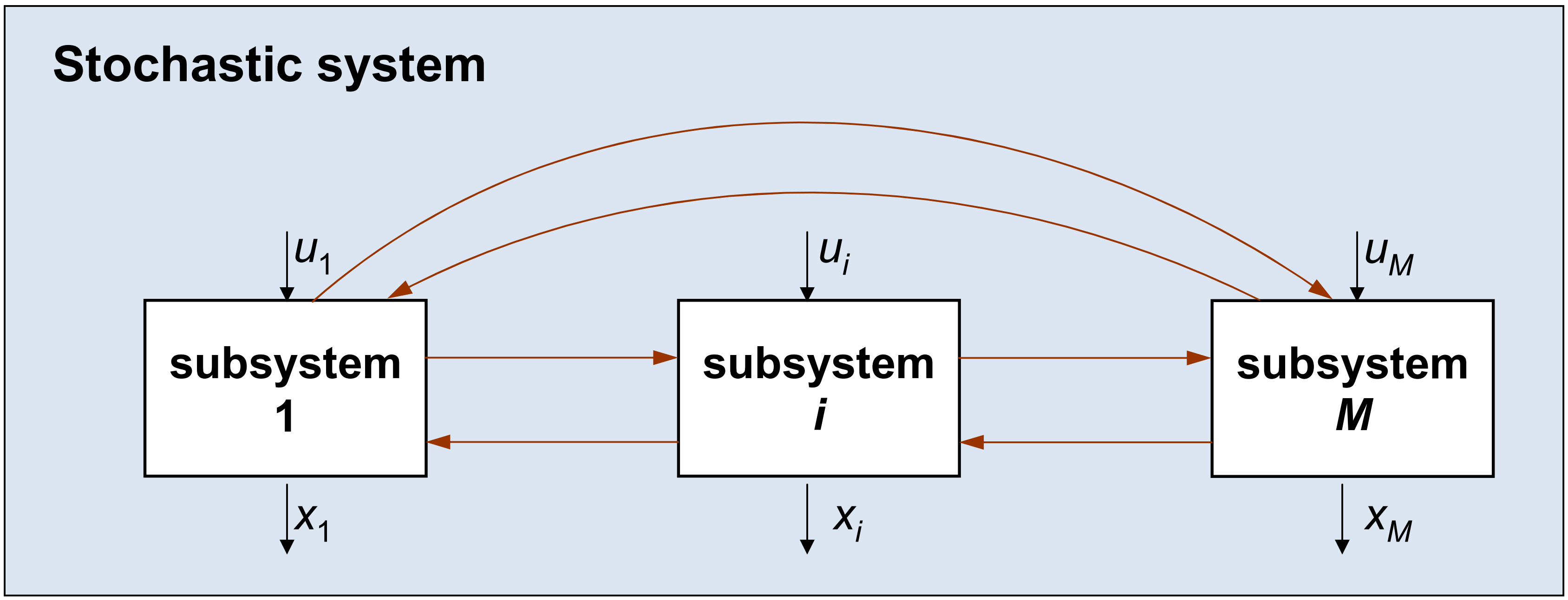
$$ASE = \frac{1}{L}\sum_{i=1}^{L}[\mu(\hat{y}(i)) - y(i)]^2$$

$$LDE = \frac{1}{2L}\sum_{i=1}^{L}\log(2\pi) + \log[\sigma^2(\hat{y}(i))] + \frac{[\mu(\hat{y}(i)) - y(i)]^2}{\sigma^2(\hat{y}(i))}$$



Using GP models: in addition to the prediction mean value (full line), we obtain a 95% confidence region (dotted lines) for the underlying function *f*.

Gaussian processes based model of interconnected dynamic systems



i-th subsystem:

$$x_i(t) \in \mathbb{R}^{n_i} \; - \text{ state}, \;\; u_i(t) \in \mathbb{R}^{m_i} \; - \text{ control input}$$

Overall system :

$$x(t) = [x_1(t), x_2(t), \ldots, x_M(t)] \in \mathbb{R}^n, \; n = \sum_{i=1}^{M} n_i$$

$$u(t) = [u_1(t), u_2(t), \ldots, u_M(t)] \in \mathbb{R}^m, \; m = \sum_{i=1}^{M} m_i$$

Uncertain nonlinear discrete - time models :

$$x_i(t+1) = h_i(\underbrace{x(t), u(t)}_{\text{coupled dynamics}}) + \xi_i(t), \; i = 1, \ldots, M \qquad \textbf{(2)}$$

coupled dynamics

$h_i(x(t), u(t))$ – nonlinear continuous function

$\xi_i(t)$ – Gaussian disturbances

Assumptions:

- **known topology** of interactions between subsystems;
- **uncertainty** about the functions h_i and the disturbances ξ_i, $i = 1, \ldots, M$.

Nonlinear discrete-time models:

$$y_i(t) = h_i(z(t)) + \xi_i(t),\ i = 1, \ldots, M \qquad (3)$$

$$y_i(t) = x_i(t+1),\ i = 1, 2, \ldots, M$$

$$z(t) = [x(t),\ u(t)]$$

Given a data set :

$$\mathbf{Z} = [z(0), z(1), \ldots, z(L-1)],$$

$$Y_{i,j} = [y_{i,j}(0), y_{i,j}(1), \ldots, y_{i,j}(L-1)],\ i = 1, \ldots, M,\ j = 1, \ldots, n_i$$

Relationship (3) is approximated with Gaussian processes with distributions :

$$Y_{i,1} \sim \mathcal{N}(0, \mathbf{\Sigma}_{i,1}),\ Y_{i,2} \sim \mathcal{N}(0, \mathbf{\Sigma}_{i,2}),\ \ldots,\ Y_{i,n_i} \sim \mathcal{N}(0, \mathbf{\Sigma}_{i,n_i})$$

$$i = 1, 2, \ldots, M$$

Known covariance functions:

$$\Sigma_{i,j,pq} = \mathrm{Cov}_{i,j}(y_{i,j}(p), y_{i,j}(q)) = C_{i,j}(z(p), z(q))$$

$$i = 1, \ldots, M;\ j = 1, \ldots, n_i;\ p, q = 0, 1, \ldots, L-1$$

Probability distribution of the output $y_i(L) = [y_{i,1}(L), \ldots, y_{i,n_i}(L)]$ **corresponding to a new input** $z(L)$ **:**

$$y_{i,1}(L) | z(L), (\mathbf{Z}, Y_{i,1}) \sim \mathcal{N}(\mu(y_{i,1}(L)), \sigma^2(y_{i,1}(L)))$$

$$\vdots$$

$$y_{i,n_i}(L) | z(L), (\mathbf{Z}, Y_{i,n_i}) \sim \mathcal{N}(\mu(y_{i,n_i}(L)), \sigma^2(y_{i,n_i}(L)))$$

$$i = 1, 2, \ldots, M$$

$\mu(y_{i,j}(L))$, $\sigma^2(y_{i,j}(L))$ – mean and variance of $y_{i,j}(L)$

$$i = 1, 2, \ldots, M$$

$$j = 1, 2, \ldots, n_i$$

Introducing :

$$\mu_{y_i}(L) = [\mu(y_{i,1}(L)), \; \ldots \; , \mu(y_{i,n_i}(L))]$$

$$\sigma^2_{y_i}(L) = [\sigma^2(y_{i,1}(L)), \; \ldots \; , \sigma^2(y_{i,n_i}(L))]$$

$$\mathbf{Y}_i = [Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_i}]$$

The relation (3) is represented :

$$y_i(L) | z(L), (\mathbf{Z}, \mathbf{Y}_i) \sim \mathcal{N}(\mu_{y_i}(L), \sigma^2_{y_i}(L)), \; i = 1, 2, \; \ldots \; , M$$

Multi - step ahead prediction :

$$y_i(L+k) | z(L+k), (\mathbf{Z}, \mathbf{Y}_i) \sim \mathcal{N}(\mu_{y_i}(L+k), \sigma^2_{y_i}(L+k))$$

$$i = 1, 2, \; \ldots \; , M \; , \; k = 0, 1, \; \ldots \; , N-1$$

N – prediction horizon

GP-based prediction models of the subsystems:

$$x_{i,t+k+1|t} \mid x_{t+k|t}, u_{t+k} \sim \mathcal{N}(\mu(x_{i,t+k+1|t}), \sigma^2(x_{i,t+k+1|t})) \quad (4)$$

$$i = 1, 2, \dots, M, \; k = 0, 1, \dots, N-1$$

The 95% confidence interval of $x_{i,t+k+1|t}$:

$$[\mu(x_{i,t+k+1|t}) - 2\sigma(x_{i,t+k+1|t}); \mu(x_{i,t+k+1|t}) + 2\sigma(x_{i,t+k+1|t})]$$

Predictions of the mean values are represented as:

$$\mu(x_{i,t+k+1|t}) = E\{f_{\mathrm{GP},i}(x_{t+k|t}, u_{t+k})\} \quad (5)$$

$$i = 1, 2, \dots, M, \; k = 0, 1, \dots, N-1$$

$f_{\mathrm{GP},i}(x_{t+k|t}, u_{t+k})$ – defined by the GP model (1)

Constraints :

$$x_i(t) \in \mathcal{X}_i \,,\; u_i(t) \in \mathcal{U}_i \,,\; i = 1, 2, \,\dots\, , M$$

decoupled constraints

Assumption :

A1. The admissible sets $\mathcal{X}_i$ and $\mathcal{U}_i$ are bounded polyhedral sets:

$$\mathcal{X}_i = \{x_i \in \mathbb{R}^{n_i} \mid C_i^x x_i \leq d_i^x\}$$

$$\mathcal{U}_i = \{u_i \in \mathbb{R}^{m_i} \mid C_i^u u_i \leq d_i^u\}$$

3 Distributed cooperative MPC based on Gaussian process models

Formulation of centralized GP-NMPC problem

Optimal regulation problem

Overall state at the current time t :

$$\bar{x} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M]$$

→

Set-point :

$$x_{sp} = [x_{1,sp}, x_{2,sp}, \ldots, x_{M,sp}]$$

Problem P1 (Centralized GP-NMPC):

$$V^{\text{opt}}(\overline{x}) = \min_{U} J(U, \overline{x})$$

subject to $x_{t|t} = \overline{x}$ and:

$$\mu(x_{i,t+k|t}) - 2\sigma(x_{i,t+k|t}) \in \mathcal{X}_i \ , \ i = 1,\ldots,M \ , \ k = 1,\ldots,N$$

$$\mu(x_{i,t+k|t}) + 2\sigma(x_{i,t+k|t}) \in \mathcal{X}_i \ , \ i = 1,\ldots,M \ , \ k = 1,\ldots,N$$

$$u_{i,t+k} \in \mathcal{U}_i \ , \ i = 1,\ldots,M \ , \ k = 0,1,\ldots,N-1$$

$$x_{i,t+k+1|t} \mid x_{t+k|t}, u_{t+k} \sim \mathcal{N}(\mu(x_{i,t+k+1|t}), \sigma^2(x_{i,t+k+1|t}))$$
$$i = 1, \ldots, M, \ k = 0, 1, \ldots, N-1$$

$$x_{t+k|t} = [x_{1,t+k|t}, x_{2,t+k|t}, \ldots, x_{M,t+k|t}], \ k = 0,1,\ldots,N$$

$$u_{t+k} = [u_{1,t+k}, u_{2,t+k}, \ldots, u_{M,t+k}], \ k = 0,1,\ldots,N-1$$

Optimization variables :

$$U = [u_t, u_{t+1}, \dots, u_{t+N-1}]$$

Global cost function :

$$J(U,\bar{x}) = \sum_{i=1}^{M} J_i(U,\bar{x}) = \sum_{i=1}^{M}\sum_{k=0}^{N} l_i(\mu(x_{i,t+k|t}), u_{i,t+k})$$

$$l_i(\mu(x_{i,t+k|t}), u_{i,t+k}) = \| \mu(x_{i,t+k|t}) - x_{i,\mathrm{sp}} \|_{Q_i}^2 + \| u_{i,t+k} - u_{i,\mathrm{sp}} \|_{R_i}^2$$

Weighting matrices :

$$Q_i,\ R_i \succ 0$$

$u_{i,\mathrm{sp}}$ – the steady state value of the control input of the subsystem corresponding to $x_{i,\mathrm{sp}}$

Distributed cooperative GP-NMPC by using dual decomposition

(Grancharova & Johansen, 2014, 2016, 2018)

1) Approximation of the GP-NMPC problem by a linear MPC problem

GP - based prediction models of the subsystems :

$$\mu(x_{i,t+k+1|t}) = E\{f_{\mathrm{GP},i}(x_{t+k|t}, u_{t+k})\}$$

$$i = 1, 2, \ldots, M,\ k = 0, 1, \ldots, N-1$$

Deviations from the set - point values :

$$\tilde{x}_{i,t+k} = x_{i,t+k} - x_{i,\mathrm{sp}},\ \tilde{u}_{i,t+k} = u_{i,t+k} - u_{i,\mathrm{sp}}$$

$$k = 0, 1, \ldots, N-1,\ i = 1, \ldots, M$$

GP - based prediction models :

$$\mu(\tilde{x}_{i,t+k+1|t}) = -x_{i,\mathrm{sp}} + E\{f_{\mathrm{GP},i}(\tilde{x}_{t+k|t} + x_{\mathrm{sp}}, \tilde{u}_{t+k} + u_{\mathrm{sp}})\} \quad \textbf{(6)}$$

$$k = 0,1, \ldots, N-1, \; i = 1, \ldots, M$$

$$\tilde{u}_{t+k} = [\tilde{u}_{1,t+k}, \ldots, \tilde{u}_{M,t+k}]$$

$$\tilde{x}_{t+k} = [\tilde{x}_{1,t+k}, \ldots, \tilde{x}_{M,t+k}]$$

$$u_{\mathrm{sp}} = [u_{1,\mathrm{sp}}, \ldots, u_{M,\mathrm{sp}}]$$

$$x_{\mathrm{sp}} = [x_{1,\mathrm{sp}}, \ldots, x_{M,\mathrm{sp}}]$$

Given trajectories of the deviated control input

and the mean of deviated state at time t :

$$\tilde{U}_i^0 = [\tilde{u}_{i,t}^0, \tilde{u}_{i,t+1}^0, \ldots, \tilde{u}_{i,t+N-1}^0], \quad \tilde{X}_i^0 = [\tilde{x}_{i,t|t}^0, \tilde{x}_{i,t+1|t}^0, \ldots, \tilde{x}_{i,t+N-1|t}^0]$$

The nonlinear models (6) are locally approximated

by linear models about the point $(\tilde{U}_i^0, \tilde{X}_i^0)$:

$$\mu(\tilde{x}_{i,t+k+1}) = E\{\sum_{j=1}^{M} (A_{ij,t+k}\tilde{x}_{j,t+k} + B_{ij,t+k}\tilde{u}_{j,t+k}) + g_{i,t+k}\} \qquad \textbf{(7)}$$

$$k = 0, 1, \ldots, N-1, \; i = 1, \ldots, M$$

Linear time - varying approximation of the original model :

$$A_{ij,t+k} = \nabla_{\tilde{x}_j} f_{\mathrm{GP},i}(\tilde{x}^0_{t+k|t} + x_{\mathrm{sp}}, \tilde{u}^0_{t+k} + u_{\mathrm{sp}})$$

$$B_{ij,t+k} = \nabla_{\tilde{u}_j} f_{\mathrm{GP},i}(\tilde{x}^0_{t+k|t} + x_{\mathrm{sp}}, \tilde{u}^0_{t+k} + u_{\mathrm{sp}})$$

$$g_{i,t+k} = -x_{i,\mathrm{sp}} - \sum_{j=1}^{M}(A_{ij,t+k}\tilde{x}^0_{j,t+k} + B_{ij,t+k}\tilde{u}^0_{j,t+k}) + f_{\mathrm{GP},i}(\tilde{x}^0_{t+k} + x_{\mathrm{sp}}, \tilde{u}^0_{t+k} + u_{\mathrm{sp}})$$

$$k = 0,1, \dots, N-1,\ i, j = 1, \dots, M$$

$$\tilde{u}^0_{t+k} = [\tilde{u}^0_{1,t+k}, \dots, \tilde{u}^0_{M,t+k}],\ \tilde{x}^0_{t+k|t} = [\tilde{x}^0_{1,t+k|t}, \dots, \tilde{x}^0_{M,t+k|t}]$$

Assumption :

A2. The standard deviations $\sigma(x_{i,t+k+1|t})$ predicted with the GP model (4) satisfy:

$$\sigma(x_{i,t+k+1|t}) \leq \sigma_{\max,i}\ ,\ i = 1,\ \ldots\ , M\ ,\ k = 0,\ \ldots\ , N-1$$

where $\sigma_{\max,i}\ ,\ i = 1,\ \ldots\ , M$ are known.

interpreted as the upper bound on admissible trust in model predictions

Tightened constraint sets (as in Giselsson & Rantzer, 2014) :

$$(1-\delta)\mathcal{X}_i = \{x_i \in \mathbb{R}^{n_i} \mid C_i^x x_i \le (1-\delta) d_i^x\}$$

$$(1-\delta)\mathcal{U}_i = \{u_i \in \mathbb{R}^{m_i} \mid C_i^u u_i \le (1-\delta) d_i^u\}$$

$\delta \in (0, 1)$ – relative constraint tightening

Tightened constraint sets of the state and the control input deviations from their set-point values :

$$\tilde{\mathcal{X}}_i = \{\tilde{x}_i \in \mathbb{R}^{n_i} \mid C_i^x \tilde{x}_i \le (1-\delta) d_i^x - C_i^x x_{i,\mathrm{sp}}\}$$

$$\tilde{\mathcal{U}}_i = \{\tilde{u}_i \in \mathbb{R}^{m_i} \mid C_i^u \tilde{u}_i \le (1-\delta) d_i^u - C_i^u u_{i,\mathrm{sp}}\}$$

Problem P2 (Centralized linear MPC):

$$V^*(\tilde{x}^0) = \min_{\tilde{U}} J(\tilde{U}, \tilde{x}^0)$$

s.t. $\tilde{x}_{t|t} = \tilde{x}^0 = [\tilde{x}^0_{1,t|t}, \tilde{x}^0_{2,t|t}, \ldots, \tilde{x}^0_{M,t|t}]$ and:

$$\mu(\tilde{x}_{i,t+k|t}) - 2\sigma_{\max,i} \in \tilde{\mathcal{X}}_i \, , \; i = 1,\ldots,M \, , \; k = 1,\ldots,N$$

$$\mu(\tilde{x}_{i,t+k|t}) + 2\sigma_{\max,i} \in \tilde{\mathcal{X}}_i \, , \; i = 1,\ldots,M \, , \; k = 1,\ldots,N$$

$$\tilde{u}_{i,t+k} \in \tilde{\mathcal{U}}_i \, , \; i = 1,\ldots,M \, , \; k = 0,1,\ldots,N-1$$

$$\mu(\tilde{x}_{i,t+k+1}) = E\{\sum_{j=1}^{M}(A_{ij,t+k}\tilde{x}_{j,t+k} + B_{ij,t+k}\tilde{u}_{j,t+k}) + g_{i,t+k}\}$$

$$k = 0,1, \ldots, N-1, \; i = 1, \ldots, M$$

Optimization variables :

$$\tilde{U} = [\tilde{u}_t, \tilde{u}_{t+1}, \ldots, \tilde{u}_{t+N-1}]$$

Global cost function :

$$J(\tilde{U}, \tilde{x}^0) = \sum_{i=1}^{M} \sum_{k=0}^{N} [\| \mu(\tilde{x}_{i,t+k|t}) \|_{Q_i}^2 + \| \tilde{u}_{i,t+k} \|_{R_i}^2]$$

Weighting matrices :

$$Q_i, R_i \succ 0$$

2) Representation and solution of the linear MPC problem as a distributed Quadratic Programming problem

Stacking all decision variables into one vector :

$$S = [\mu(\tilde{x}_{1,t+1|t}), \tilde{u}_{1,t}, \mu(\tilde{x}_{1,t+2|t}), \tilde{u}_{1,t+1}, \ \dots \ , \mu(\tilde{x}_{1,t+N|t}), \tilde{u}_{1,t+N-1},$$

$$\vdots$$

$$\mu(\tilde{x}_{M,t+1|t}), \tilde{u}_{M,t}, \mu(\tilde{x}_{M,t+2|t}), \tilde{u}_{M,t+1}, \ \dots \ , \mu(\tilde{x}_{M,t+N|t}), \tilde{u}_{M,t+N-1}]$$

$$S \in \mathbb{R}^{n_S}, \ n_S = \sum_{i=1}^{M} N(n_i + m_i)$$

Problem P3 (QP problem):

$$V^*(\tilde{x}^0) = \min_S \frac{1}{2} S^{\mathrm{T}} \overline{H} S$$

$$\text{s.t.} \quad \overline{A} S = \overline{B} \tilde{x}^0 - \overline{G}$$

$$\overline{C} S \le \overline{d}$$

$$\overline{H} = \mathrm{diag}\{\overline{H}_1, \overline{H}_2, \ldots, \overline{H}_M\}$$

$$\overline{A} = [\overline{A}_1 \,|\, \overline{A}_2 \,|\, \ldots \,|\, \overline{A}_M]^{\mathrm{T}}, \; \overline{B} = [\overline{B}_1 \,|\, \overline{B}_2 \,|\, \ldots \,|\, \overline{B}_M]^{\mathrm{T}}, \; \overline{G} = [\overline{G}_1 \,|\, \overline{G}_2 \,|\, \ldots \,|\, \overline{G}_M]^{\mathrm{T}}$$

$$\overline{C} = \mathrm{diag}\{\overline{C}_1 \,|\, \overline{C}_2 \,|\, \ldots \,|\, \overline{C}_M\}, \; \overline{d} = \mathrm{diag}\{\overline{d}_1 \,|\, \overline{d}_2 \,|\, \ldots \,|\, \overline{d}_M\}$$

For the i - th subsystem :

$$\bar{H}_i = \text{diag}\{\underbrace{W_i, W_i, \ldots, W_i}_{N \text{ elements}}\},\; W_i = \begin{bmatrix} Q_i & 0 \\ 0 & R_i \end{bmatrix}$$

$$\bar{A}_i = \begin{bmatrix} \bar{A}_{i,1,1} & \bar{A}_{i,1,2} & \cdots & \bar{A}_{i,1,M} \\ \bar{A}_{i,2,1} & \bar{A}_{i,2,2} & \cdots & \bar{A}_{i,2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{A}_{i,N-1,1} & \bar{A}_{i,N-1,2} & \cdots & \bar{A}_{i,N-1,M} \end{bmatrix}, \bar{B}_i = \begin{bmatrix} -A_{i1,t} & -A_{i2,t} & \cdots & -A_{iM,t} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$\bar{A}_{i,k,j}$, $k = 1, 2, \ldots, N-1,\; j = 1, 2, \ldots, M$ depend on the matrices $A_{ij,t+k}$ and $B_{ij,t+k}$ of the linear model (7)

$A_{i1,t}, A_{i2,t}, \ldots, A_{iM,t}$ are the matrices $A_{ij,t+k}$ for $k = 0,\; j = 1, 2, \ldots, M$

For the i - th subsystem :

$$\bar{G}_i = [g_{i,t}, g_{i,t+1}, \dots, g_{i,t+N-1}]^{\mathrm{T}}$$

$$\bar{C}_i = \mathrm{diag}\{\underbrace{C_i^{x,u}, C_i^{x,u}, \dots, C_i^{x,u}}_{N \text{ elements}}\}, \; C_i^{x,u} = \begin{bmatrix} C_i^x & 0 \\ 0 & C_i^u \end{bmatrix}$$

$$\bar{d}_i = \mathrm{diag}\{\underbrace{d_i^{x,u}, d_i^{x,u}, \dots, d_i^{x,u}}_{N \text{ elements}}\}, \; d_i^{x,u} = [d_i^x, d_i^u]^{\mathrm{T}}$$

The QP problem P3 is solved distributedly by applying the dual accelerated gradient algorithm in Giselsson et al. (2013)

The distribution is enabled by solving the dual problem to problem P3:

- introducing dual variables $\lambda \in \mathbb{R}^{n_{\bar{A}}}$ for the equality constraints;
- introducing dual variables $\eta \in \mathbb{R}^{n_{\bar{C}}}$ for the inequality constraints;

Dual problem:

$$\max_{\lambda,\eta\geq 0} D(\tilde{x}^0, \lambda, \eta)$$

Dual cost function:

$$D(\tilde{x}^0, \lambda, \eta) = -\frac{1}{2}(\bar{A}^{\mathrm{T}}\lambda + \bar{C}^{\mathrm{T}}\eta)^{\mathrm{T}}\,\bar{H}^{-1}(\bar{A}^{\mathrm{T}}\lambda + \bar{C}^{\mathrm{T}}\eta)$$

$$-\lambda^{\mathrm{T}}(\bar{B}\tilde{x}^0 - \bar{G}) - \eta^{\mathrm{T}}\bar{d}$$

Decision variables, associated to the i - th subsystem :

(enables distributed iterations of the dual accelerated gradient method)

$$S_i = [\mu(\tilde{x}_{i,t+1|t}), \tilde{u}_{i,t}, \mu(\tilde{x}_{i,t+2|t}), \tilde{u}_{i,t+1}, \ \dots \ , \mu(\tilde{x}_{i,t+N|t}), \tilde{u}_{i,t+N-1}]$$

Dual variables:

- $\lambda_i \in \mathbb{R}^{n_{\bar{A}_i}}$ for the equality constraints;
- $\eta_i \in \mathbb{R}^{n_{\bar{C}_i}}$ for the inequality constraints;

Distributed iterations of the dual accelerated gradient method (Giselsson et al. (2013))

$$S_i^r = -\bar{H}_i^{-1}\left(\sum_{j=1}^{M} \bar{A}_j^{i\mathrm{T}} \lambda_j^r + \bar{C}_i^{\mathrm{T}} \eta_i^r\right)$$

iteration number

$$\bar{S}_i^r = S_i^r + \frac{r-1}{r+2}\left(S_i^r - S_i^{r-1}\right)$$

$$\lambda_i^{r+1} = \lambda_i^r + \frac{r-1}{r+2}\left(\lambda_i^r - \lambda_i^{r-1}\right) + \frac{1}{L}\left(\bar{A}_i \bar{S}^r - \left(\bar{B}_i \tilde{x}_i^0 - \bar{G}_i\right)\right)$$

Lipschitz constant to the gradient of the dual function

$$\eta_i^{r+1} = \max\left(0, \eta_i^r + \frac{r-1}{r+2}\left(\eta_i^r - \eta_i^{r-1}\right) + \frac{1}{L}\left(\bar{C}_i \bar{S}_i^r - \bar{d}_i\right)\right)$$

$$i = 1, 2, \ldots, M$$

$\overline{H}_i$, $\overline{A}_i$, $\overline{B}_i$, $\overline{G}_i$, $\overline{C}_i$, $\overline{A}_j^i$, $\overline{d}_i$ – related to the i – th subsystem and represent submatrices / subvectors of $\overline{H}$, $\overline{A}$, $\overline{B}$, $\overline{G}$, $\overline{C}$, $\overline{d}$ in problem P3

$L = \left\| [\overline{A}^{\mathrm{T}}, \overline{C}^{\mathrm{T}}]^{\mathrm{T}} \overline{H}^{-1} [\overline{A}^{\mathrm{T}}, \overline{C}^{\mathrm{T}}] \right\|$ – the Lipschitz constant to the gradient of the dual function

r – the iteration number

Remarks:

- Because of the couplings in the dynamics models of the subsystems, the computation of the decision variables S_i^r for the *i*-th subsystem requires to have information about the dual variables λ^r for the whole system.

- The update of the dual variables λ_i associated to the *i*-th subsystem uses the information about the decision variables $\bar{S}^r$ for the entire system.

- Since there are no couplings in the control input and state constraints of the subsystems, the update of the dual variables η_i for the *i*-th subsystem requires information only about the decision variables $\bar{S}_i^r$ for this subsystem.

3) Algorithm for distributed GP-NMPC by sequential linearization and distributed Quadratic Programming

Based on the search for the Nash equilibrium between the interconnected subsystems by applying the distributed iterations

(modification of the distributed deterministic NMPC algorithm in Grancharova et al. (2016))

The iterations terminate if:

$$| J_i(U_2,\overline{x}) - J_i(U_1,\overline{x}) | < \varepsilon, \ \forall i = 1, \ldots, M$$

U_1, U_2 – the control input trajectories for the subsystems obtained in two sequential iterations in the outer loop

$J_i(U_1,\overline{x})$, $J_i(U_2,\overline{x})$ – the local cost function values

Algorithm:

1. Given ε, R and δ. Let $t = 0$, $U(t) = [u_{sp}, u_{sp}, \dots, u_{sp}]$.

2. Let the state at time t be $\bar{x} = [\bar{x}_1, \dots, \bar{x}_M]$.

3. Compute the state mean trajectory $\mu(X(t))$ corresponding to initial state $\bar{x}$ and control input trajectory $U(t)$ by using the GP model (12) and the associated local cost function values $J_{\mathrm{new},i} := J_i(U, \bar{x})$, $i = 1, \dots, M$ in (22). Form the vector $S(t)$ of decision variables.

4. Do

5. $J_{\mathrm{old},i} := J_{\mathrm{new},i}$, $i = 1, \dots, M$

6. Obtain a linearized model (26)-(27) of the model (12) around the trajectories $(U(t), \mu(X(t)))$.

7. **For** $r = 0, 1, \dots, R$ **do**

8. **If** $r = 0$ **then**

9. Initialize iterations (46)-(49) with $S^{-1} = S(t)$, $\lambda^0 = \lambda^{-1} = 0$, $\eta^0 = \eta^{-1} = 0$.

10. **else**

11. Let $S^{r-1} := S^r$, $\lambda^{r-1} := \lambda^r$, $\lambda^r := \lambda^{r+1}$, $\eta^{r-1} := \eta^r$, $\eta^r := \eta^{r+1}$.

12. **end**

13. Run (46)-(49) distributedly by communicating S_i^r and λ_i^r, $i = 1, \ldots, M$ between interconnected subsystems and obtain S^r, λ^{r+1}, η^{r+1} for the overall system. Extract U^r from S^r.

14. **end**

15. Let $U(t) = U^R$.

16. Compute the state mean trajectory $\mu(X(t))$ corresponding to initial state $\bar{x}$ and control input trajectory $U(t)$ by using the GP model (12) and the associated local cost function values $J_{\text{new},i} := J_i(U, \bar{x})$, $i = 1, \ldots, M$ in (22). Form the vector $S(t)$ of decision variables.

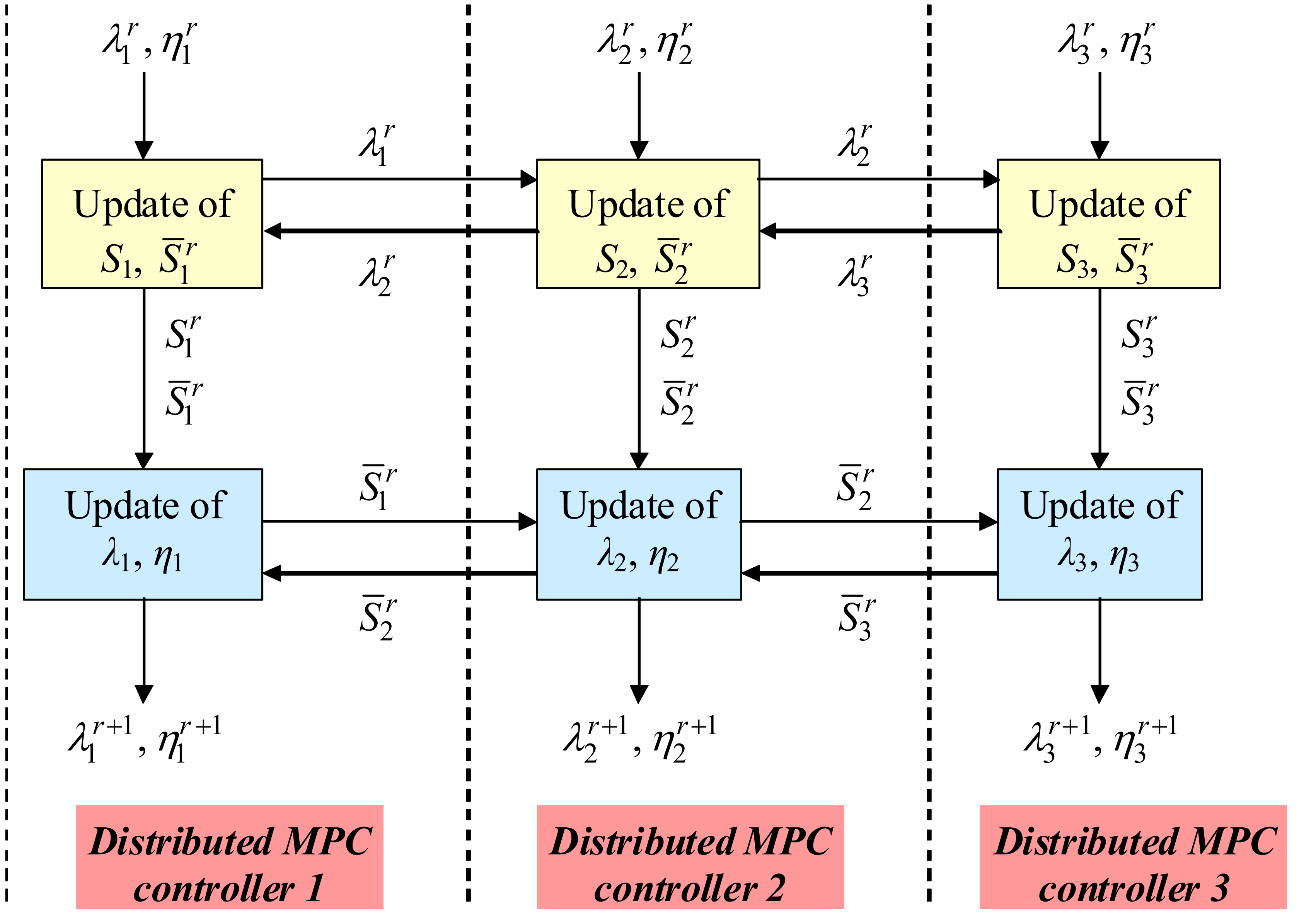
17. **while** Nash equilibrium is reached

$(|J_{\text{new},i} - J_{\text{old},i}| < \varepsilon, \forall i = 1, \ldots, M)$

18. Apply to the overall system the input $u(t) = [I\ 0 \ldots 0]U(t)$.

19. Let $t = t + 1$ and go to step 2.

Communication between the distributed MPC controllers

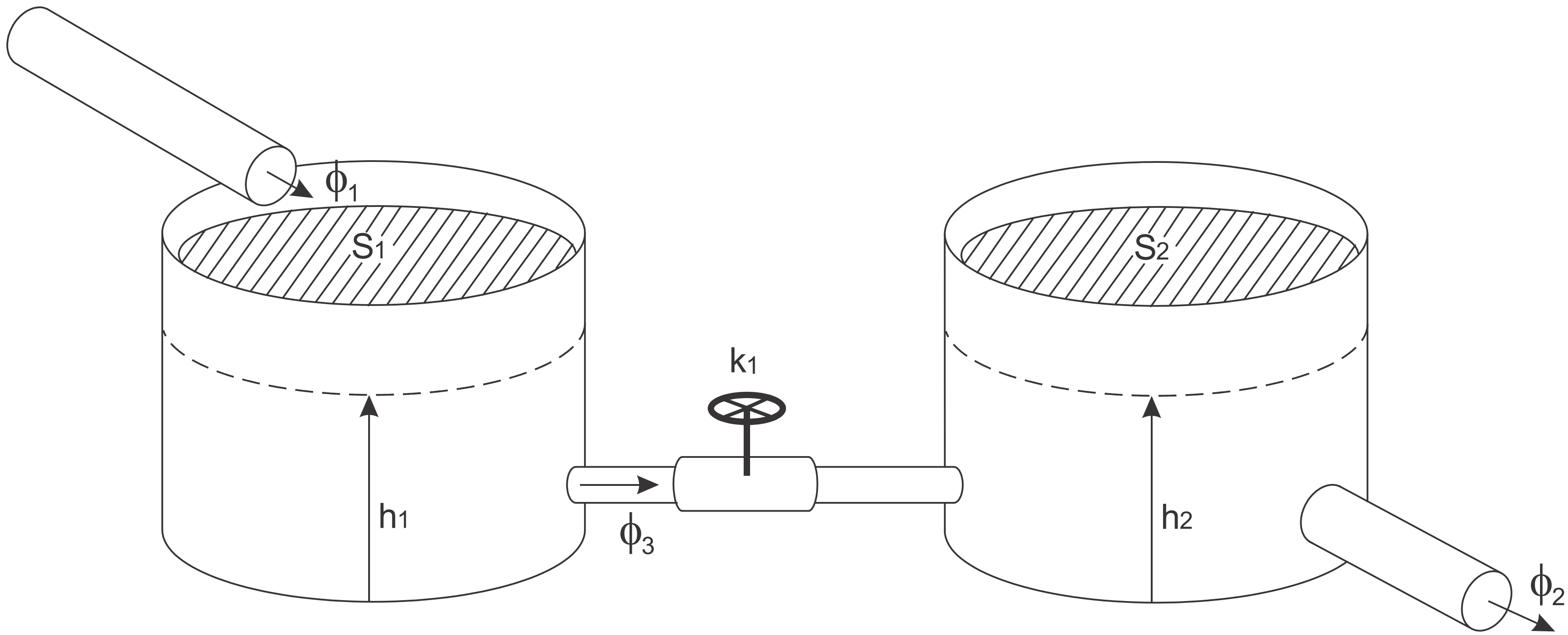


4 Numerical Example

System description - combined sewer networks

(Marinaki & Papageorgiou, 2005; Schütze & Beck, 2002)

- collect domestic and industrial sewage as well as rainwater drainage;
- **distributed systems** that consist of pipes, sewer stretches (in-line storage), retention reservoirs (off-line storage) with overflow capabilities, and nodes for merging of flows from different catchments;
- at **rain events**, the whole retention capacity of all reservoirs is used; during **dry weather conditions**, the sewer storage capacities can be used for the smoothing of peak discharges towards the wastewater treatment plant.



The concept of interacting reservoirs for simplified modelling of relatively flat sewers where backwater phenomena may occur

First-pricnciples model

Mathematical model :

$$S_1 \frac{dh_1}{d\tau} = -k_1 \sqrt{h_1(\tau) - h_2(\tau)} + \phi_1(\tau)$$

$$S_2 \frac{dh_2}{d\tau} = k_1 \sqrt{h_1(\tau) - h_2(\tau)} - \phi_2(\tau)$$

τ – the continuous time

h_1 , h_2 – the heights of liquid in the tanks

S_1 , S_2 – the cross sectional areas of the tanks

ϕ_1 , ϕ_2 – the volumetric flows

k_1 – valve characteristic

Discrete - time model :

$$h_1(t+1) = h_1(t) - (T_s/S_1)k_1\sqrt{h_1(t) - h_2(t)} + (T_s/S_1)\phi_1(t)$$

$$h_2(t+1) = h_2(t) + (T_s/S_2)k_1\sqrt{h_1(t) - h_2(t)} - (T_s/S_2)\phi_2(t)$$

t – discrete time, $T_s = 1$ min – sampling time

$S_1 = 2.3130 \text{ m}^2$, $S_2 = 2.1048 \text{ m}^2$

Control inputs :

$$u(t) = [\phi_2(t), k_1(t)]$$

States :

$$x(t) = [h_1(t), h_2(t)]$$

Stochastic disturbance :

$$v(t) = (T_s/S_1)\phi_1(t)$$

Subsystem 1:

$$x_1(t+1) = x_1(t) - 0.4323 u_1(t)\sqrt{x_1(t) - x_2(t)} + v(t)$$

Subsystem 2:

$$x_2(t+1) = x_2(t) + 0.4751 u_1(t)\sqrt{x_1(t) - x_2(t)} - 0.4751 u_2(t)$$

$v = \mathcal{N}(0,\ 0.02^2)$ – Gaussian disturbance

Control goal: **maintain the level of sewage in both tanks with minimal deviation from the set-point despite of input flow variations**

Gaussian process model

Gaussian covariance function :

$$C(z(t_p), z(t_q)) = v_1 \exp\left[-\frac{1}{2}\sum_{i=1}^{D} w_i (z_i(t_p) - z_i(t_q))^2\right]$$

GP - model of subsystem 1 :

$$x_{1,t+k+1|t} \mid z_{t+k|t} \sim \mathcal{N}(\mu(x_{1,t+k+1|t}), \sigma^2(x_{1,t+k+1|t}))$$

$$z_{t+k|t} = [x_{1,t+k|t},\ x_{2,t+k|t},\ u_{1,t+k|t}],\ k = 0, 1,\ \ldots\ , N-1$$

Estimated hyperparameters

(based on 1000 training samples) :

$$\Theta^1 = [w_1^1,\ w_2^1,\ w_3^1,\ v_1^1] = [0.4275,\ 0.1989,\ 0.0005,\ 1.0167]$$

GP - model of subsystem 2 :

$$x_{2,t+k+1|t} \mid z_{t+k|t} \sim \mathcal{N}(\mu(x_{2,t+k+1|t}), \sigma^2(x_{2,t+k+1|t}))$$

$$z_{t+k|t} = [x_{1,t+k|t},\ x_{2,t+k|t},\ u_{1,t+k|t},\ u_{2,t+k|t}],\ k = 0,1,\ \ldots\ , N-1$$

Estimated hyperparameters :

$$\Theta^2 = [w_1^2,\ w_2^2,\ w_3^2,\ w_4^2,\ v_1^2] = [6.7230,\ 6.8256,\ 0.0008,\ 0.0046,\ 0.1707]$$

Inputs for models' identification are generated as random values with uniform distribution for all regressors

Maximal values of standard deviations :

$$\sigma_{\max,1} = 0.025, \ \sigma_{\max,2} = 0.009$$

Set-point values of x_1, x_2 **and steady-state values of** u_1, u_2 **:**

$$x_{1,\mathrm{sp}} = 1.5 \text{ m}, \ x_{2,\mathrm{sp}} = 1.2 \text{ m}$$

$$u_{1,\mathrm{sp}} = 0.15 \text{ m}^2\sqrt{\text{m}} / \text{min}, \ u_{2,\mathrm{sp}} = 0.05 \text{ m}^3 / \text{min}$$

Constraints :

$$0.1 \le u_1(t) \le 0.2 \text{ m}^2\sqrt{\text{m}} / \text{min}$$

$$0 \le u_2(t) \le 0.1 \text{ m}^3 / \text{min}$$

$$x_1(t) \le 1.8 \text{ m}$$

Simulation results

NMPC parameters :

Horizon: $N = 5$

Weighting coefficients:

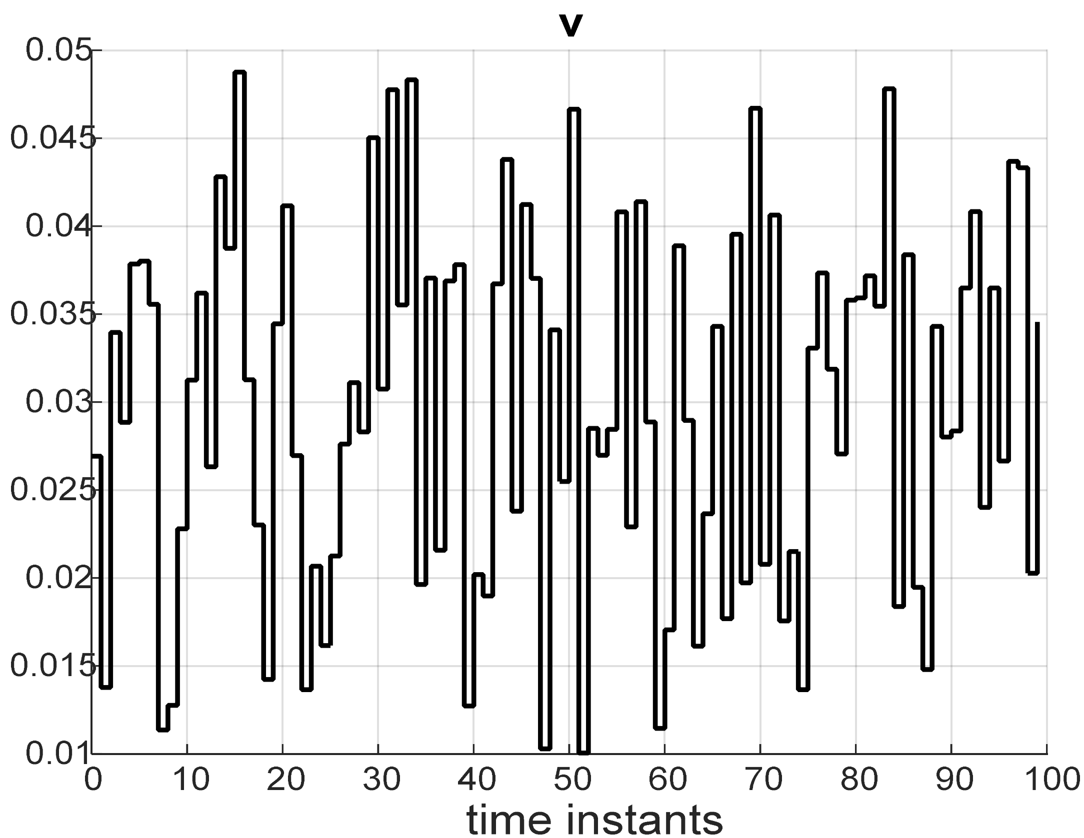
$Q_1 = Q_2 = 10,\ R_1 = R_2 = 0.1$

Design parameters in the algorithm :

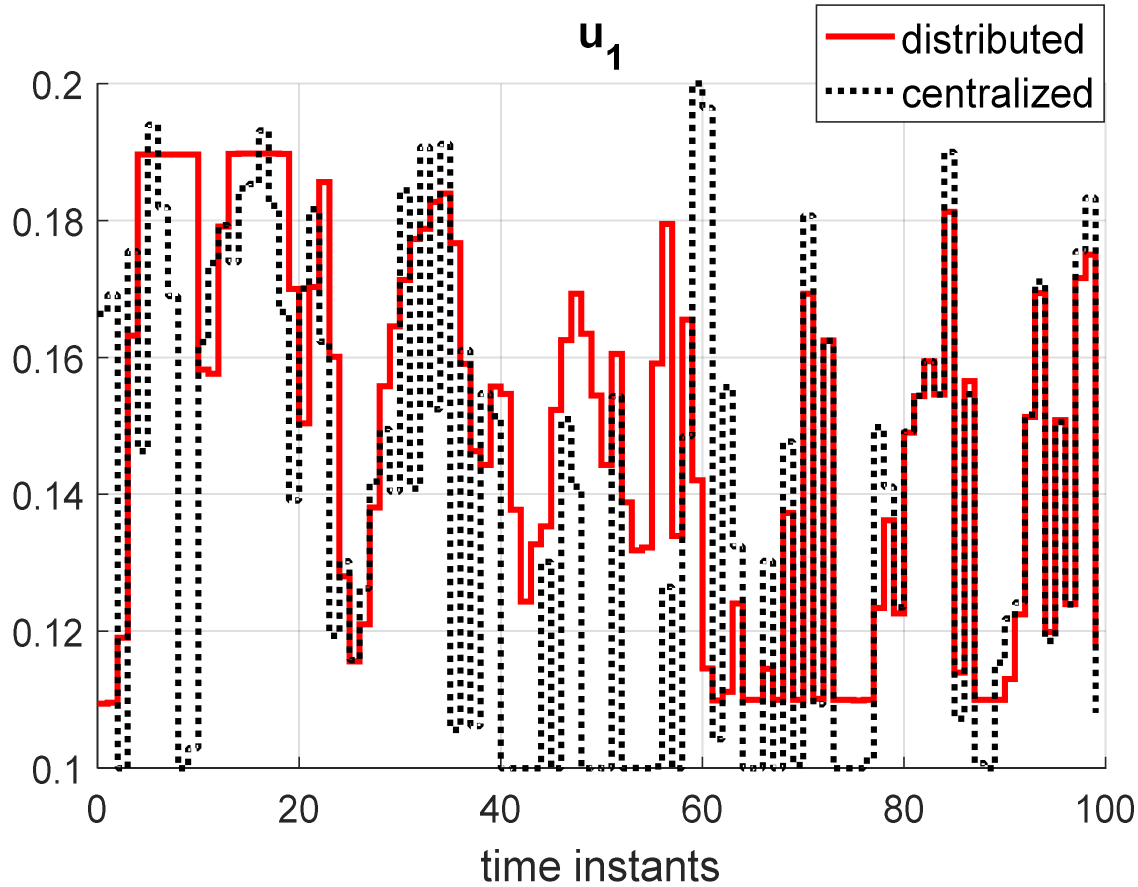
$\delta = 0.2,\ \varepsilon = 0.03,\ R = 70$

Initial states of the subsystems :

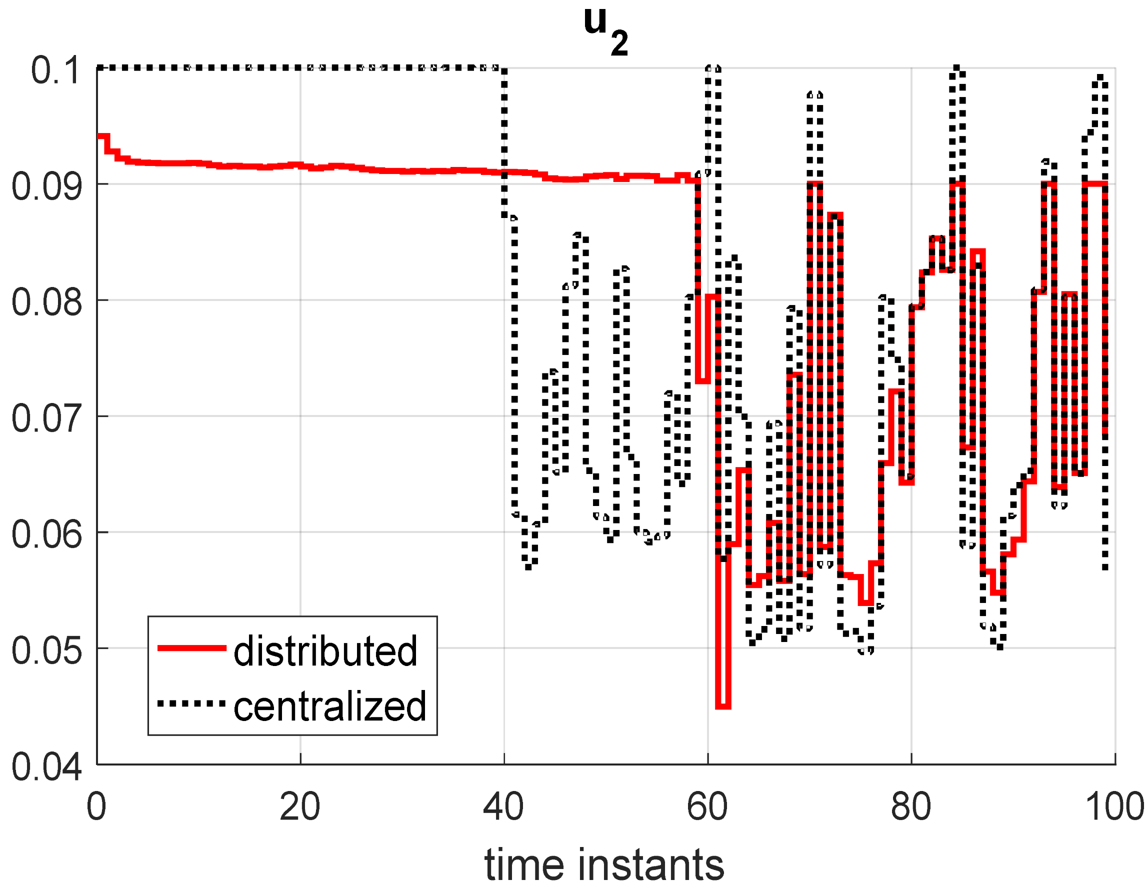
$x_1(0) = 1.73\ \text{m},\ x_2(0) = 1.6\ \text{m}$



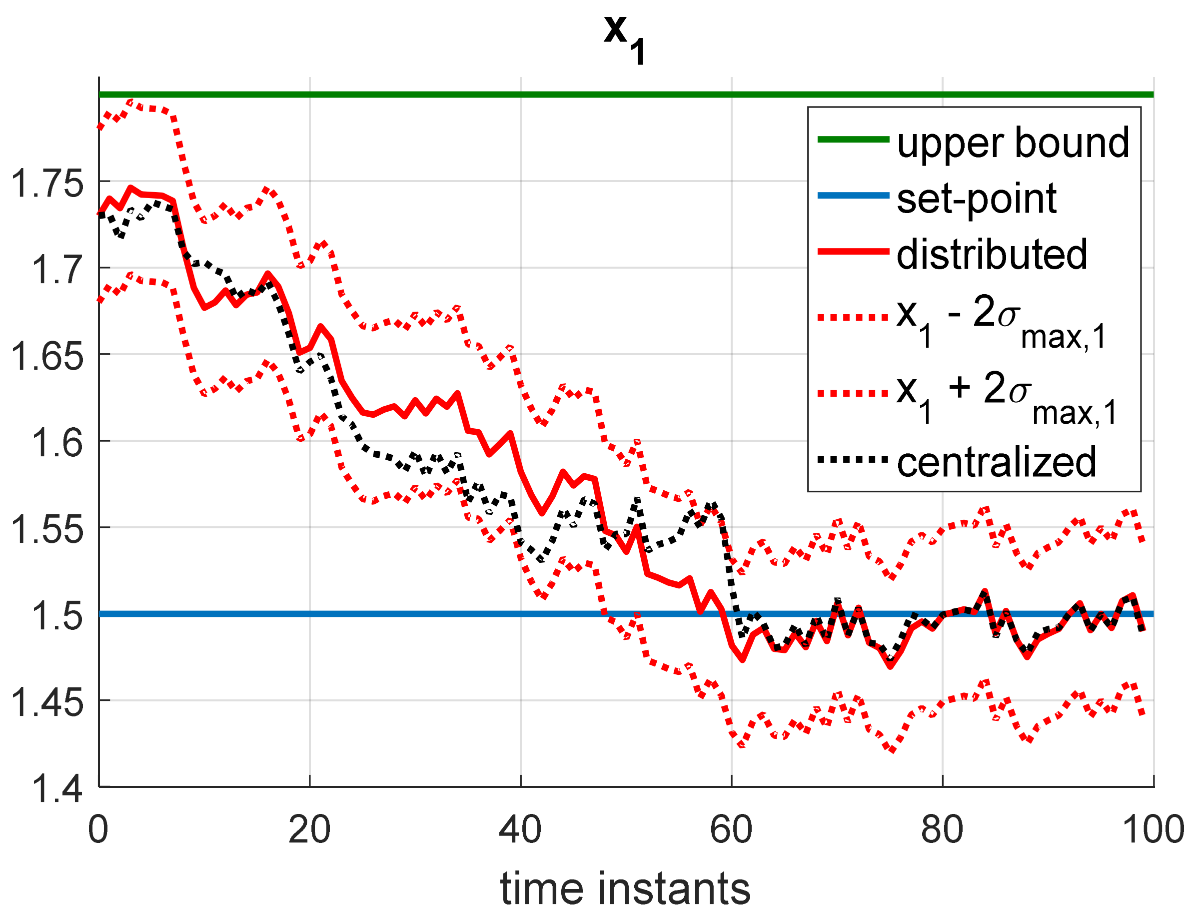
The persistent stochastic disturbance v (related to the input flow φ_1)



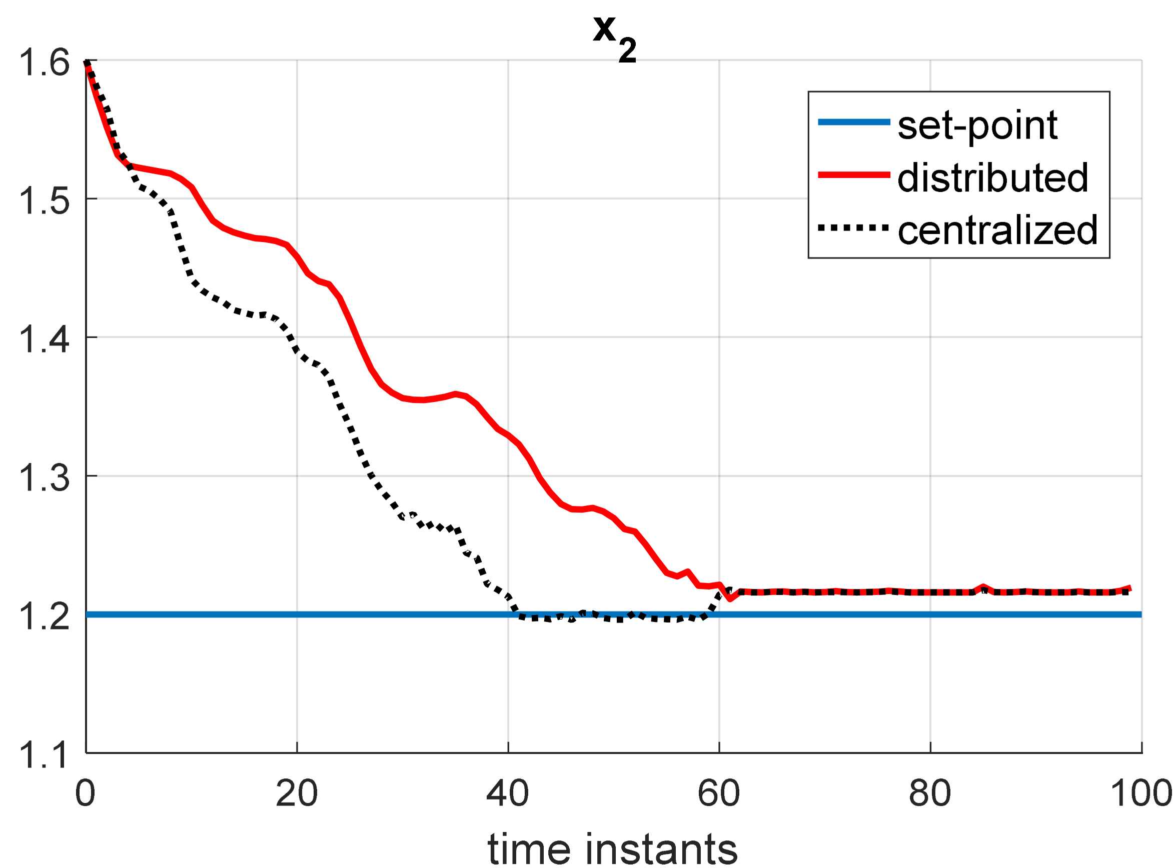
Trajectory of control input u_1



Trajectory of control input u_2



Trajectory of the state variable x_1



Trajectory of the state variable x_2

Computational cost per sampling time

(3.10 GHz AMD Ryzen 3 1200 quad-core processor)

Method	Average CPU time, s	Maximal CPU time, s
Distributed GP-NMPC	0.10	0.34
Centralized GP-NMPC	11.54	55.94

5 Conclusions

- **A suboptimal approach to distributed GP-NMPC is proposed based on Gaussian process models of the interconnected systems dynamics. It has a reduced complexity of the on-line computations and its simple software implementation makes it attractive for the implementation as embedded control.**

- **It allows the computation of the suboptimal control inputs to be done autonomously by the subsystems without the need for centralized optimization.**

References:

Christofides, P.D., Scattolini, R., Muñoz de la Peña, D., & Liu, J. (2013). Distributed model predictive control: A tutorial review and future research directions. *Computers & Chemical Engineering*, 51, 21–41.

Giselsson, P., Doan, M.D., Keviczky, T., De Schutter, B., & Rantzer, A. (2013). Accelerated gradient methods and dual decomposition in distributed model predictive control. *Automatica*, 49, 829–833.

Grancharova, A., Valkova, I., Hvala, N. & Kocijan, J. (2023). Distributed predictive control based on Gaussian process models. *Automatica*, vol. 149, 110807.

Grancharova, A., Johansen, T.A., & Petrova, V. (2016). Distributed nonlinear model predictive control by sequential linearization and accelerated gradient method. *IFAC-PapersOnLine*, 49, 597–602.

Grancharova, A., Kocijan, J., & Johansen, T.A. (2008). Explicit stochastic predictive control of combustion plants based on Gaussian process models. *Automatica*, 44, 1621–1631.

Kocijan, J., Grancharova, A. (2014). Application of Gaussian processes to the modeling and control in process engineering. In: V. E. Balas et al. (eds.), *Innovations in Intelligent Machines-5*, *Studies in Computational Intelligence 561*, Springer-Verlag, pp.155-190, 2014.

Kocijan, J. (2016). Modelling and control of dynamic systems using Gaussian process models. *Advances in Industrial Control*, Springer International Publishing.

Marinaki, M. & Papageorgiou, M. (2005). *Optimal real-time control of sewer networks*. Springer.

Rasmussen, C.E. & Williams, C.K.I. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge.

Schütze, M.R., & Beck, M.B. (2002). *Modelling, simulation and control of urban wastewater systems*. Springer.