PSO for Convolved GP Models

Particle Swarm Optimization for Convolved Gaussian Process Models

2014 IEEE World Congress on Computaional Intelligence (IJCNN Program)

Gang Cao

School of Engineering and Advanced Technology
Massey University

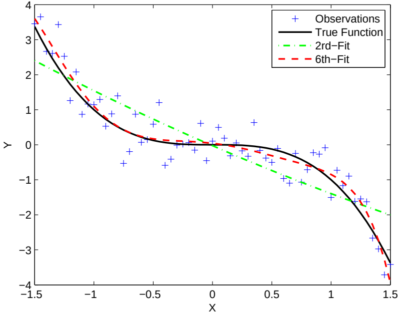
May 9, 2016

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Outline

Introduction

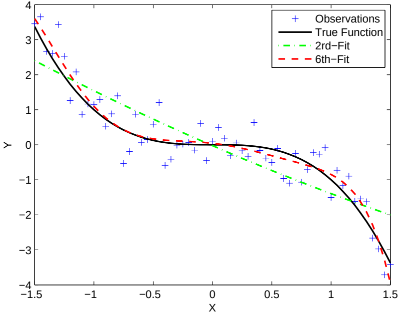
Learning from Data



A comparison of different model for learning question.
Original model is $y = x^3$ and corrupted by a noise $\epsilon \sim \mathcal{N}(0, 0.5)$.

Introduction

Learning from Data

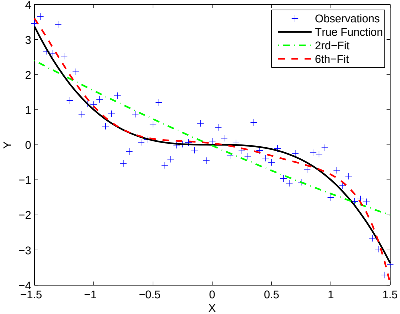


A comparison of different model for learning question.
Original model is $y = x^3$ and corrupted by a noise $\epsilon \sim \mathcal{N}(0, 0.5)$.

Questions?

Introduction

Learning from Data



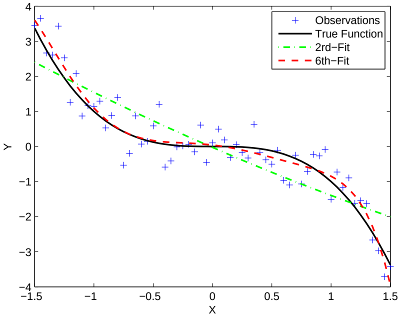
A comparison of different model for learning question.
Original model is $y = x^3$ and corrupted by a noise $\epsilon \sim \mathcal{N}(0, 0.5)$.

Questions?

- Best Model?

Introduction

Learning from Data



A comparison of different model for learning question.
Original model is $y = x^3$ and corrupted by a noise $\epsilon \sim \mathcal{N}(0, 0.5)$.

Questions?

- Best Model?
- Errors? and Measurements?

Introduction

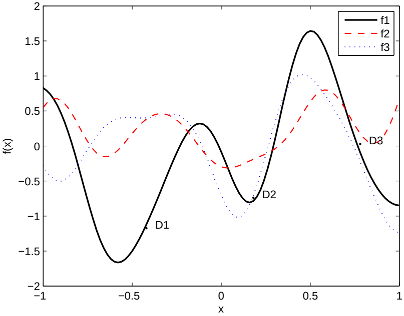
Bayesian Model

- Use prior
- Consider all possible models

Introduction

Bayesian Model

- Use prior
- Consider all possible models

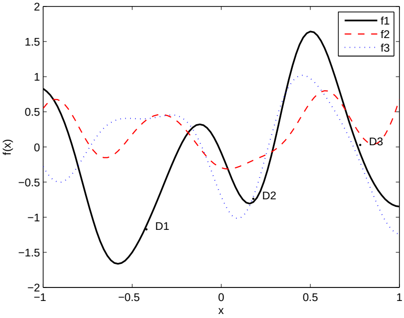


Bayesian Regression Technique
Contributions: $f1 > f3 > f2$

Introduction

Bayesian Model

- Use prior
- Consider all possible models



Bayesian Regression Technique
Contributions: $f1 > f3 > f2$

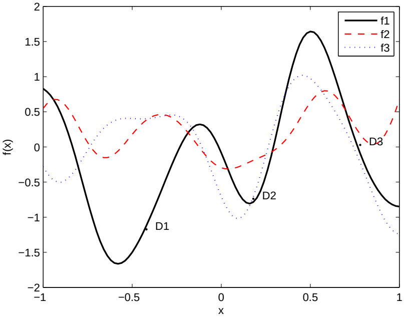
Questions?

- Defining Prior

Introduction

Bayesian Model

- Use prior
- Consider all possible models



Bayesian Regression Technique
Contributions: $f1 > f3 > f2$

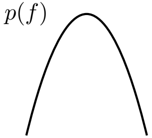
Questions?

- Defining Prior
- Analytically intractable

Gaussian Process Models

- Standard GP
- Multi-output GP
- Model Learning
- CGP Learning using PSO

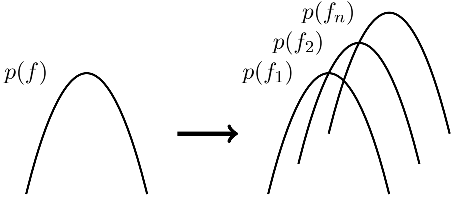
GP Model – Standard GP



Gaussian distribution

$f \sim \mathcal{N}(\mu, \sigma^2)$

GP Model – Standard GP



Gaussian distribution

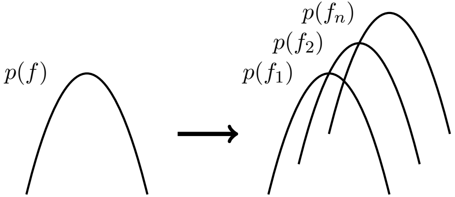
$f \sim \mathcal{N}(\mu, \sigma^2)$

Gaussian processes

$\mathbf{f} \sim \mathcal{GP}(\boldsymbol{\mu}(\cdot), \mathbf{cov}(\cdot))$

- GP: A collection of random variables with a joint Gaussian distribution among any finite number of them

GP Model – Standard GP



Gaussian distribution
$f \sim \mathcal{N}(\mu, \sigma^2)$

Gaussian processes
$\mathbf{f} \sim \mathcal{GP}(\boldsymbol{\mu}(\cdot), \mathbf{cov}(\cdot))$

- GP: A collection of random variables with a joint Gaussian distribution among any finite number of them
- Fully specified by mean and variance functions

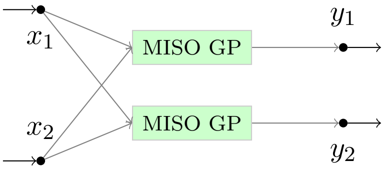
$$\begin{aligned}\mu(\mathbf{x}^*) &= \mathbf{K}_{\mathbf{f}^*,\mathbf{f}}(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma^2\mathbf{I})^{-1}\mathbf{y} \\ var(\mathbf{x}^*) &= \mathbf{K}_{\mathbf{f}^*,\mathbf{f}^*} + \sigma^2 - \mathbf{K}_{\mathbf{f}^*,\mathbf{f}}(\mathbf{K}_{\mathbf{f},\mathbf{f}} + \sigma^2\mathbf{I})^{-1}\mathbf{K}_{\mathbf{f},\mathbf{f}^*}.\end{aligned} \tag{1}$$

GP Model – Multi-Output GP

Question: MIMO problem using GP model?

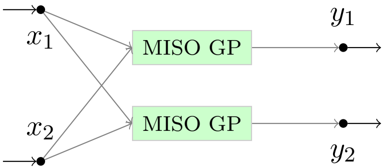
GP Model – Multi-Output GP

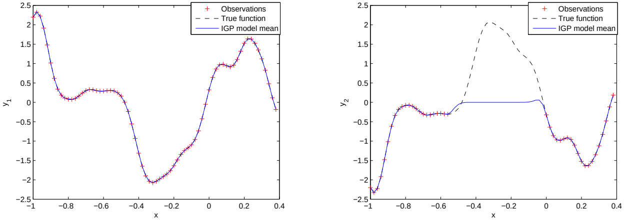
Question: MIMO problem using GP model?



GP Model – Multi-Output GP

Question: MIMO problem using GP model?

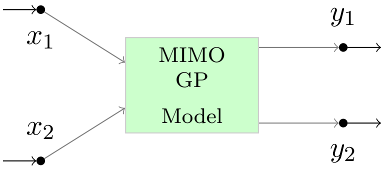




Modelling using IGP model

GP Model – Multi-Output GP

Question: MIMO problem using GP model?



GP Model – Multi-Output GP

Convolved Gaussian process models:

GP Model – Multi-Output GP

Convolved Gaussian process models:

- Construct a new covariance function

$$\begin{bmatrix} \mathbf{K_{11}} & \cdots & \mathbf{K_{1n}} \\ \mathbf{K_{21}} & \cdots & \mathbf{K_{2n}} \\ \vdots & \vdots & \vdots \\ \mathbf{K_{n1}} & \cdots & \mathbf{K_{nn}} \end{bmatrix} \tag{2}$$

GP Model – Multi-Output GP

Convolved Gaussian process models:

- Construct a new covariance function

$$\begin{bmatrix} \boxed{\mathbf{K_{11}}} & \cdots & \mathbf{K_{1n}} \\ \mathbf{K_{21}} & \boxed{\cdots} & \mathbf{K_{2n}} \\ \vdots & \vdots & \vdots \\ \mathbf{K_{n1}} & \cdots & \boxed{\mathbf{K_{nn}}} \end{bmatrix} \tag{2}$$

  - input-output correlations

PSO for Convolved GP Models

GP Model – Multi-Output GP

Convolved Gaussian process models:

- Construct a new covariance function

$$\begin{bmatrix} \mathbf{K_{11}} & \cdots & \mathbf{K_{1n}} \\ \mathbf{K_{21}} & \cdots & \mathbf{K_{2n}} \\ \vdots & \vdots & \vdots \\ \mathbf{K_{n1}} & \cdots & \mathbf{K_{nn}} \end{bmatrix} \tag{2}$$

  - input-output correlations
  - correlations between outputs

GP Model – Multi-Output GP

Convolved Gaussian process models:

- Construct a new covariance function

$$\begin{bmatrix} \mathbf{K_{11}} & \cdots & \mathbf{K_{1n}} \\ \mathbf{K_{21}} & \cdots & \mathbf{K_{2n}} \\ \vdots & \vdots & \vdots \\ \mathbf{K_{n1}} & \cdots & \mathbf{K_{nn}} \end{bmatrix} \tag{2}$$

  - input-output correlations
  - correlations between outputs
- No additional computations of inference and predictions.

PSO for Convolved GP Models

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}\left(\theta\right) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}}\ \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-\mathbf{1}}\ \mathbf{y} - \frac{1}{2}\mathbf{log}\left|\ \mathbf{K}_{\mathbf{y},\mathbf{y}}\ \right| - \frac{\mathbf{MJ}}{2}\log\mathbf{2}\pi \quad (3)$$

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{K}_{\mathbf{y},\mathbf{y}}^{-\mathbf{1}}\mathbf{y} - \frac{1}{2}\mathbf{log}\left|\mathbf{K}_{\mathbf{y},\mathbf{y}}\right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}} \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{log}\left| \mathbf{K}_{\mathbf{y},\mathbf{y}} \right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}} \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{log}\left|\mathbf{K}_{\mathbf{y},\mathbf{y}}\right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}} \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{log}\left| \mathbf{K}_{\mathbf{y},\mathbf{y}} \right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

Questions?

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1}\mathbf{y} - \frac{1}{2}\mathbf{log}\left|\mathbf{K}_{\mathbf{y},\mathbf{y}}\right| - \frac{\mathbf{MJ}}{2}\log\mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

Questions?

- Sensitive to initial values;

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}} \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{log}\left| \mathbf{K}_{\mathbf{y},\mathbf{y}} \right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

Questions?

- Sensitive to initial values;
- Unclear indication of model quality;

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}} \mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1} \mathbf{y} - \frac{1}{2}\mathbf{log}\left| \mathbf{K}_{\mathbf{y},\mathbf{y}} \right| - \frac{\mathbf{MJ}}{2}\log \mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

Questions?

- Sensitive to initial values;
- Unclear indication of model quality;

	Search Range [0,100]		Search Range [0,1]	
	Fitness 1	Fitness 2	Fitness 1	Fitness 2
NLL	≈ 51	≈ 46	≈ 242269	≈ 212314

PSO for Convolved GP Models

GP Model – Model Learning

The log-likelihood function:

$$\mathbf{L}(\theta) = -\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{K}_{\mathbf{y},\mathbf{y}}^{-1}\mathbf{y} - \frac{1}{2}\log\left|\mathbf{K}_{\mathbf{y},\mathbf{y}}\right| - \frac{\mathbf{MJ}}{2}\log\mathbf{2}\pi \quad (3)$$

- Maximizing the Log-likelihood
- Commonly used technique – CG

Questions?

- Sensitive to initial values;
- Unclear indication of model quality;

	Search Range [0,100]		Search Range [0,1]	
	Fitness 1	Fitness 2	Fitness 1	Fitness 2
NLL	≈ 51	≈ 46	≈ 242269	≈ 212314
MSE	0.5313	0.2199	0.0101	0.0032

GP Model – CGP learning using PSO

PSO for CGP Model Learning

GP Model – CGP learning using PSO

PSO for CGP Model Learning

- Better global search

GP Model – CGP learning using PSO

PSO for CGP Model Learning

- Better global search
- Easy-to-use

GP Model – CGP learning using PSO

PSO for CGP Model Learning

- Better global search
- Easy-to-use

Works in this paper:

GP Model – CGP learning using PSO

PSO for CGP Model Learning

- Better global search
- Easy-to-use

Works in this paper:

- Explore CGP learning using PSO

GP Model – CGP learning using PSO

PSO for CGP Model Learning

- Better global search
- Easy-to-use

Works in this paper:

- Explore CGP learning using PSO
- Use MSE as fitness;

Simulation Results

- Simulation Design
- Simulation Results

Simlation Results – Design

Numerical Example:

- MISO Modelling Problem

$$\begin{aligned} y(k) &= 0.893y(k-1) + 0.037y^2(k-1) \\ &\quad -0.05y(k-2) + 0.157u(k-1) \\ &\quad -0.05u(k-1)y(k-1) \end{aligned} \tag{4}$$

- $y_1 = y$, $y_2 = -y_1$ or $y_2 = exp(y_1)$
- Generate 1000 observations
- Each simulation repeats 10 times

Experimental Parameters:

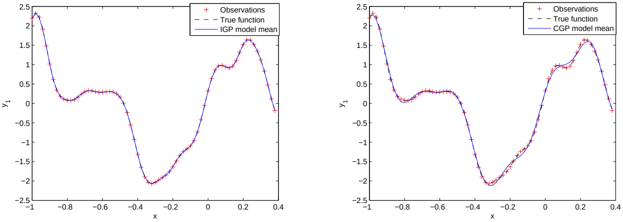
Symbol	Description	Quantity
$\mathbf{T}_{\max}$	Maximum Iterations	2000
c_1,c_2	Acceleration Factors	1.49445
ω_{start}	Start Inertial Factor	0.4
ω_{end}	End Inertial Factor	0.9
k	Shape Control Factor	0.8
CG Restarts	Restart Times	25×2000
$\|\Delta\xi\|$	Minimum Fitness Variation	$1e-5$
$\nu_{d,i}, \upsilon_q$	Coefficients Search Range	$[1, 2]$
α_i, β_j	$\mathbf{P_d}$, $\mathbf{P_q}$Elements Search Range	$[0, 1]$

Simlation Results – Results

Simulation 1 – Modelling with Missing Data

Simlation Results – Results

Simulation 1 – Modelling with Missing Data

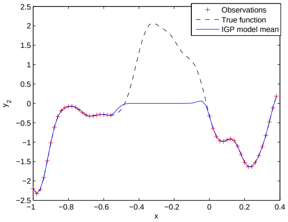


(c) y_1 – IGP with full data

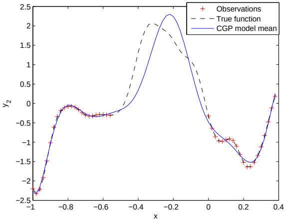
(d) y_1 – CGP with full data

Simlation Results – Results

Simulation 1 – Modelling with Missing Data



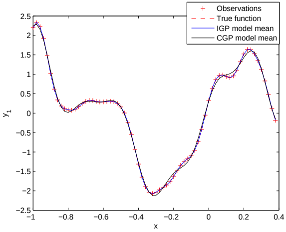
(e) y_2 – IGP with miss data



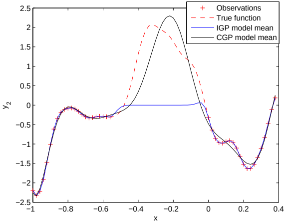
(f) y_2 – CGP with miss data

Simlation Results – Results

Simulation 1 – Modelling with Missing Data



(g) y_1 with full data

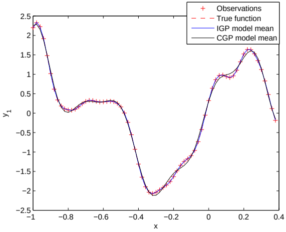


(h) y_2 with miss data

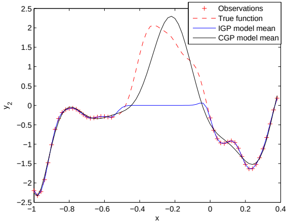
- Similar performance with full data

Simlation Results – Results

Simulation 1 – Modelling with Missing Data



(i) y_1 with full data

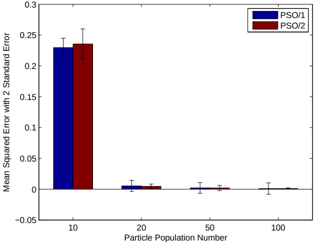


(j) y_2 with miss data

- Similar performance with full data
- CGP has better performance with miss data

Simlation Results – Results

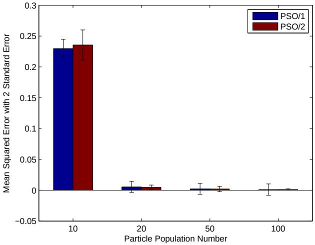
Simulation 2 – MISO System Modelling



Results of MISO modelling problem.
PSO/2 denotes PSO with MSE fitness, while PSO/2 is PSO with LL fitness.

Simlation Results – Results

Simulation 2 – MISO System Modelling

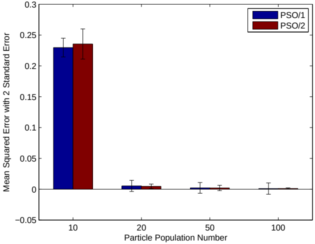


Results of MISO modelling problem.
PSO/2 denotes PSO with MSE fitness, while PSO/2 is PSO with LL fitness.

- 2 PSOs have similar performance

Simlation Results – Results

Simulation 2 – MISO System Modelling

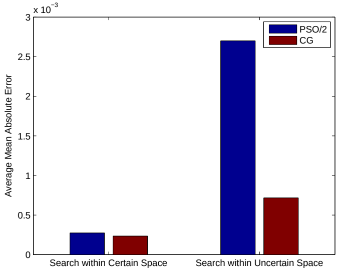


Results of MISO modelling problem.
PSO/2 denotes PSO with MSE fitness, while PSO/2 is PSO with LL fitness.

- 2 PSOs have similar performance
- Population number 25 is used in the simulations

Simlation Results – Results

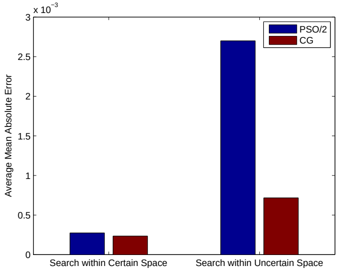
Simulation 3 – MISO System Modelling



Statistic Results of MISO system modelling.
PSO/2 denotes PSO with MSE fitness.

Simlation Results – Results

Simulation 3 – MISO System Modelling

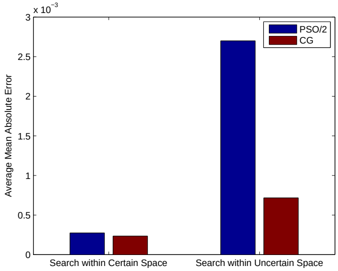


Statistic Results of MISO system modelling.
PSO/2 denotes PSO with MSE fitness.

- Similar performance when search space is well-defined

Simlation Results – Results

Simulation 3 – MISO System Modelling

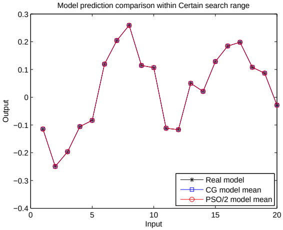


Statistic Results of MISO system modelling.
PSO/2 denotes PSO with MSE fitness.

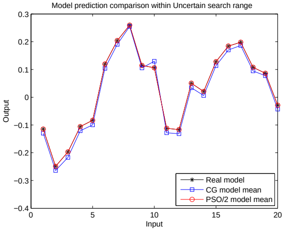
- Similar performance when search space is well-defined
- PSO has better performance with unknown or large search space

Simlation Results – Results

Simulation 3 – MISO System Modelling



(k) Certain Search range



(l) Uncertain Search range

Model prediction comparison
within certain or uncertain search ranges.

- Similar performance when search space is well-defined
- PSO has better performance with unknown or large search space

Simlation Results – Results

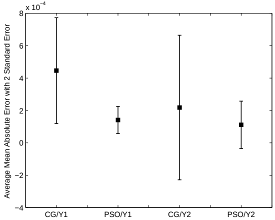
Simulation 4 – MIMO System Modelling

- Two outputs are correlated
- Linear and nonlinear correlations

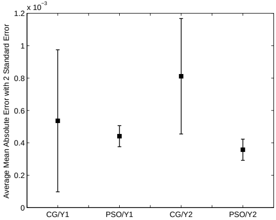
Simlation Results – Results

Simulation 4 – MIMO System Modelling

- Two outputs are correlated
- Linear and nonlinear correlations



(m) $y_2 = -y_1$

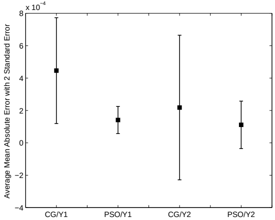


(n) $y_2 = \mathbf{exp}(y_1)$

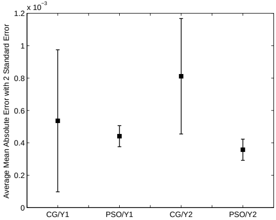
Simlation Results – Results

Simulation 4 – MIMO System Modelling

- Two outputs are correlated
- Linear and nonlinear correlations



(o) $y_2 = -y_1$



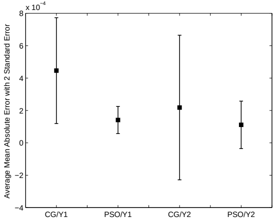
(p) $y_2 = \mathbf{exp}(y_1)$

- Model with PSO produces smaller average MSE and SE within a uncertain search range

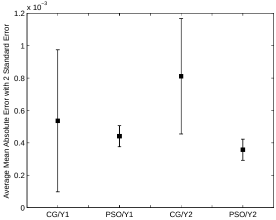
Simlation Results – Results

Simulation 4 – MIMO System Modelling

- Two outputs are correlated
- Linear and nonlinear correlations



(q) $y_2 = -y_1$



(r) $y_2 = \mathbf{exp}(y_1)$

- Model with PSO produces smaller average MSE and SE within a uncertain search range
- CGP model with PSO performs well both for linear and nonlinear problems

Conclusion

- Better modelling ability for missing-data problems

Conclusion

- Better modelling ability for missing-data problems
- Clear indications of MSE than LL fitness

Conclusion

- Better modelling ability for missing-data problems
- Clear indications of MSE than LL fitness
- Better learning ability for uncertain search range

Conclusion

- Better modelling ability for missing-data problems
- Clear indications of MSE than LL fitness
- Better learning ability for uncertain search range
- Work well for linear and nonlinear systems

Future Works

- More complex questions:NLTI to NLTV

Future Works

- More complex questions:NLTI to NLTV
- Use improved PSO methods

Thanks!

Questions?