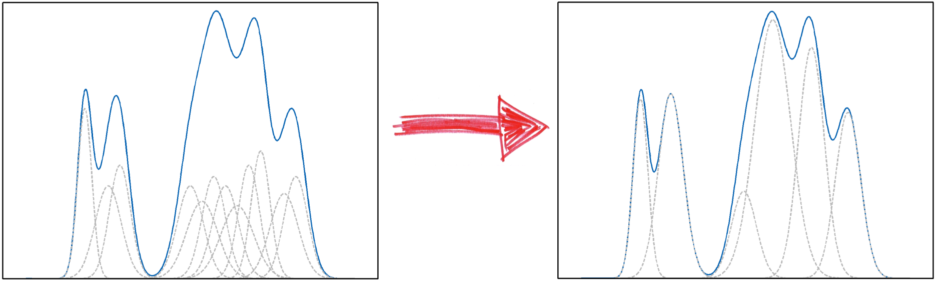
Karlsruhe Institute of Technology

Clustering-based Gaussian Mixture Reduction

Dennis Schieferdecker - 09.06.09

GRK 1194: Self-organizing Sensor-Actuator-Networks



Introduction

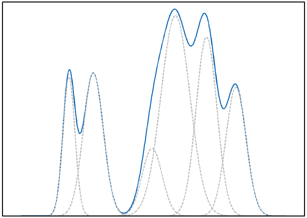
Gaussian Mixtures

Gaussian Mixture Density

- weighted sum of Gaussians

$$f(x; \underline{\eta}) = \sum_{i=1}^{N} \omega_i \cdot \mathcal{N}(x; \mu_i, \sigma_i^2)$$

- universal function approximator
- possible applications
  - target tracking,
  - density estimation,
  - ...



Problems in Application

- recursive multiplication of Gaussian mixtures
- number of components grows rapidely (exponential growth)

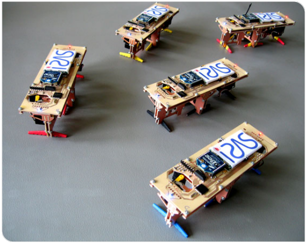
Introduction
Gaussian Mixtures

Gaussian Mixture Density

- weighted sum of Gaussians

$$f(x;\underline{\eta}) = \sum_{i=1}^{N} \omega_i \cdot \mathcal{N}(x;\mu_i,\sigma_i^2)$$

- universal function approximator
- possible applications
  - target tracking,
  - density estimation,
  - ...



Problems in Application

- recursive multiplication of Gaussian mixtures
- number of components grows rapidely (exponential growth)

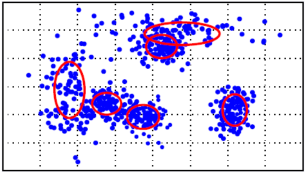
Introduction

Gaussian Mixtures

Gaussian Mixture Density

- weighted sum of Gaussians

$$f(x; \underline{\eta}) = \sum_{i=1}^{N} \omega_i \cdot \mathcal{N}(x; \mu_i, \sigma_i^2)$$



- universal function approximator
- possible applications
  - target tracking,
  - density estimation,
  - ...

Problems in Application

- recursive multiplication of Gaussian mixtures
- number of components grows rapidely (exponential growth)

Introduction

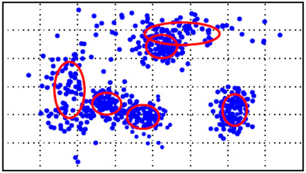
Gaussian Mixtures

Gaussian Mixture Density

- weighted sum of Gaussians

$$f(x;\underline{\eta}) = \sum_{i=1}^{N} \omega_i \cdot \mathcal{N}(x;\mu_i,\sigma_i^2)$$

- universal function approximator
- possible applications
  - target tracking,
  - density estimation,
  - ...



Problems in Application

- recursive multiplication of Gaussian mixtures
- number of components grows rapidly (exponential growth)

Problem Description

Gaussian Mixture Reduction

Goal

- given a mixture $\underline{\tilde{\eta}}$ with N components (true/original mixture),
- find a mixture $\underline{\eta}$ with $K < N$ components (reduced mixture),
- so that a deviation measure $d(\underline{\tilde{\eta}}, \underline{\eta})$ is minimized.

Deviation Measures

- Integrated Squared Distance (ISD):
  $d(f_1(x), f_2(x)) = \int_{\mathbb{R}} (f_1(x) - f_2(x))^2 \, dx$
- Kullback-Leibler divergence (KLD):
  $d(f_1(x), f_2(x)) = \int_{\mathbb{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$
- normalized variants

Problem Description

Gaussian Mixture Reduction

Karlsruhe Institute of Technology

Goal

- given a mixture $\underline{\tilde{\eta}}$ with N components (true/original mixture),
- find a mixture $\underline{\eta}$ with $K < N$ components (reduced mixture),
- so that a deviation measure $d(\underline{\tilde{\eta}}, \underline{\eta})$ is minimized.

Deviation Measures

- Integrated Squared Distance (ISD):
  $d(f_1(x), f_2(x)) = \int_{\mathbb{R}} (f_1(x) - f_2(x))^2 \, \mathrm{d}x$
- Kullback-Leibler divergence (KLD):
  $d(f_1(x), f_2(x)) = \int_{\mathbb{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} \mathrm{d}x$
- normalized variants

Reduction Methods

Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR

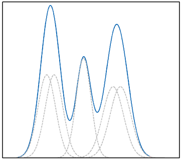
Reduction Methods
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

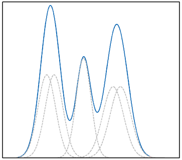
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

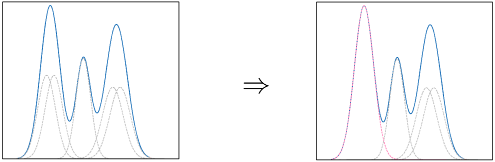
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

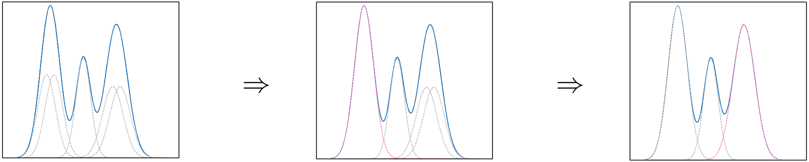
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

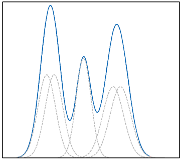
Overview

top-down approaches

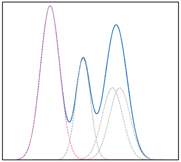
- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (local, global, hybrid)

bottom-up approach

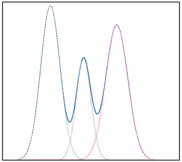
- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



⇒



⇒



Reduction Methods

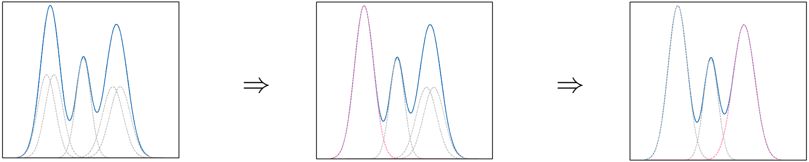
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR

Reduction Methods

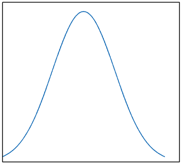
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

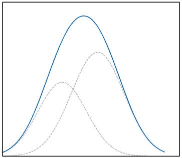
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

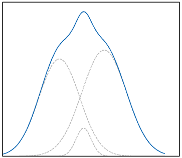
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



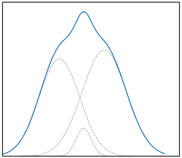
Reduction Methods
Overview

top-down approaches

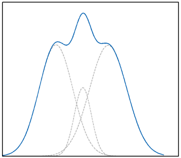
- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



$\Rightarrow$



Reduction Methods

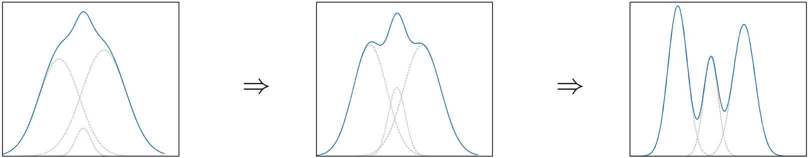
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR



Reduction Methods

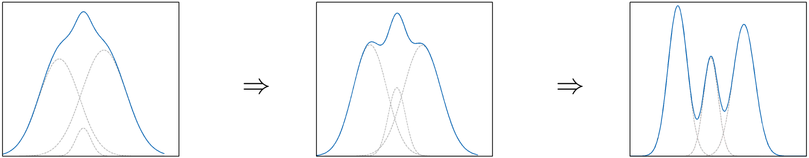
Overview

top-down approaches

- greedy methods
- iteratively replace two Gaussians with one
- chosen according to a deviation measure (West, Williams, Runnalls)

bottom-up approach

- constructive method
- starts with one Gaussian
- adds components as neccessary
- progressive approximation
- PGMR - state-of-the-art

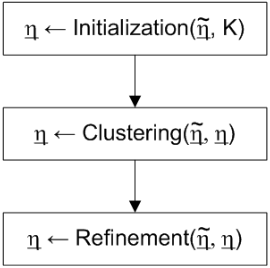


Clustering Method

Overview

Gaussian Mixture Reduction via Clustering (GMRC)

- modular three-step algorithm
- input:
  $\tilde{\underline{\eta}}$ (parameter vector of the original mixture)
  K (number of reduced components)
- output:
  $\underline{\eta}$ (parameter vector of the reduced mixture)



Conception

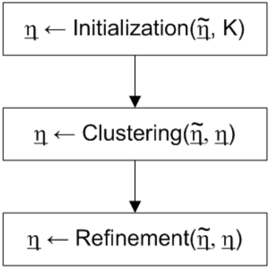
- each component $\underline{\eta}_i$ is interpreted as point (site) in a space with an underlying deviation measure

Clustering Method

Overview

Gaussian Mixture Reduction via Clustering (GMRC)

- modular three-step algorithm
- input:
  $\underline{\tilde{\eta}}$ (parameter vector of the original mixture)
  K (number of reduced components)
- output:
  $\underline{\eta}$ (parameter vector of the reduced mixture)



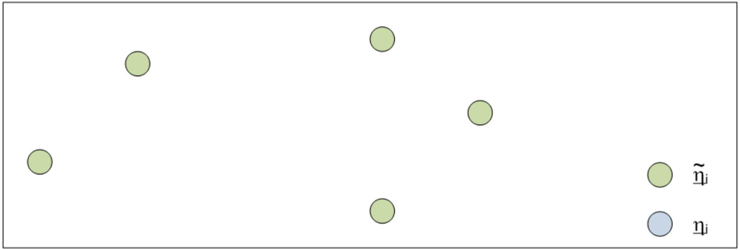
Conception

- each component $\underline{\eta}_i$ is interpreted as point (site) in a space with an underlying deviation measure

Clustering Method

Initialization Step

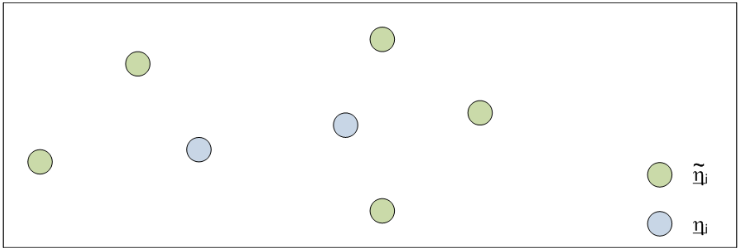
- compute a preliminary solution $\underline{\eta}$ (i.e. using West, Runnalls, ...) $\rightarrow$ initial cluster centers
- associate each original component (site) $\underline{\tilde{\eta}}_i$ with the nearest component of the reduced mixture
- replace each cluster center with a new one, retaining mean and variance of the associated sites



Clustering Method

Initialization Step

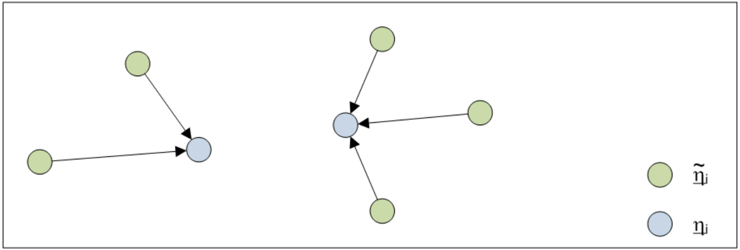
- compute a preliminary solution $\underline{\eta}$ (i.e. using West, Runnalls, ...) $\rightarrow$ initial cluster centers
- associate each original component (site) $\underline{\tilde{\eta}}_i$ with the nearest component of the reduced mixture
- replace each cluster center with a new one, retaining mean and variance of the associated sites



Clustering Method

Initialization Step

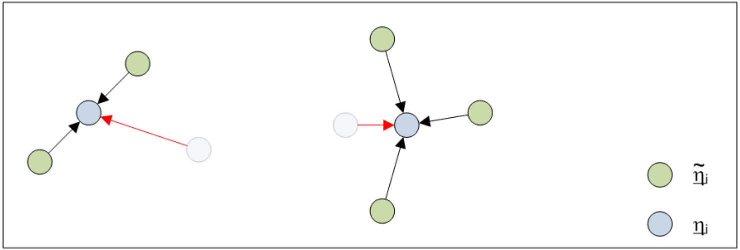
- compute a preliminary solution $\underline{\eta}$ (i.e. using West, Runnalls, . . .)
  $\rightarrow$ initial cluster centers
- associate each original component (site) $\underline{\tilde{\eta}}_i$ with the nearest component of the reduced mixture
- replace each cluster center with a new one, retaining mean and variance of the associated sites



Clustering Method

Initialization Step

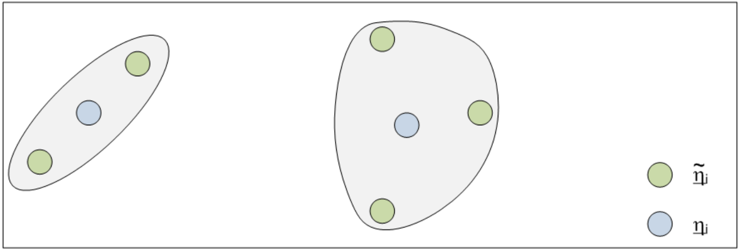
- compute a preliminary solution $\underline{\eta}$ (i.e. using West, Runnalls, . . .)
  $\longrightarrow$ initial cluster centers
- associate each original component (site) $\underline{\tilde{\eta}}_i$ with the nearest component of the reduced mixture
- replace each cluster center with a new one, retaining mean and variance of the associated sites



Clustering Method

Initialization Step

- compute a preliminary solution $\underline{\eta}$ (i.e. using West, Runnalls, ...) $\rightarrow$ initial cluster centers
- associate each original component (site) $\underline{\tilde{\eta}}_i$ with the nearest component of the reduced mixture
- replace each cluster center with a new one, retaining mean and variance of the associated sites



Clustering Method

Clustering Step

- greedy approach
- based on Lloyd's algorihm (k-means algorithm):
  - associate each site $\tilde{\underline{\eta}}_i$ with the 'nearest' center $\underline{\eta}_j$
  - recompute centers according to the current association
  - repeat until the deviation no longer changes or is good enough

determine the 'nearest' center

- associate site $\tilde{\underline{\eta}}_i$ with each center $\underline{\eta}_j$
- temporarily update the affected centers
- compute change in deviation between updated reduced and original mixture (ISD)
- retain association with smallest deviation

Clustering Method

Clustering Step

- greedy approach
- based on Lloyd's algorihm (k-means algorithm):
  - associate each site $\underline{\tilde{\eta}}_i$ with the 'nearest' center $\underline{\eta}_j$
  - recompute centers according to the current association
  - repeat until the deviation no longer changes or is good enough

determine the 'nearest' center

- associate site $\underline{\tilde{\eta}}_i$ with each center $\underline{\eta}_j$
- temporarily update the affected centers
- compute change in deviation between updated reduced and original mixture (ISD)
- retain association with smallest deviation

Clustering Method

Clustering Step

- greedy approach
- based on Lloyd's algorihm (k-means algorithm):
  - associate each site $\underline{\tilde{\eta}}_i$ with the 'nearest' center $\underline{\eta}_j$
  - recompute centers according to the current association
  - repeat until the deviation no longer changes or is good enough

determine the 'nearest' center

- associate site $\underline{\tilde{\eta}}_i$ with each center $\underline{\eta}_j$
- temporarily update the affected centers
- compute change in deviation between updated reduced and original mixture (ISD)
- retain association with smallest deviation

Clustering Method

Refinement Step

Karlsruhe Institute of Technology

Parameter Optimization

- optimize parameter vector $\underline{\eta}$ w.r.t. ISD

$$\min_{\underline{\eta}} \int_{\mathbb{R}} \left(\tilde{f}(x; \underline{\tilde{\eta}}) - f(x; \underline{\eta}) \right)^2 \mathrm{d}x$$

- non-linear optimization problem $\longrightarrow$ Newton approach
- finds local optimum

Weight Optimization

- system of linear equations
- finds global optimum

Clustering Method

Refinement Step

Parameter Optimization

- optimize parameter vector $\underline{\eta}$ w.r.t. ISD

$$\min_{\underline{\eta}} \int_{\mathbb{R}} \left(\tilde{f}(x; \underline{\tilde{\eta}}) - f(x; \underline{\eta}) \right)^2 \mathrm{d}x$$

- non-linear optimization problem $\longrightarrow$ Newton approach
- finds local optimum

Weight Optimization

- system of linear equations
- finds global optimum

Results

Overview

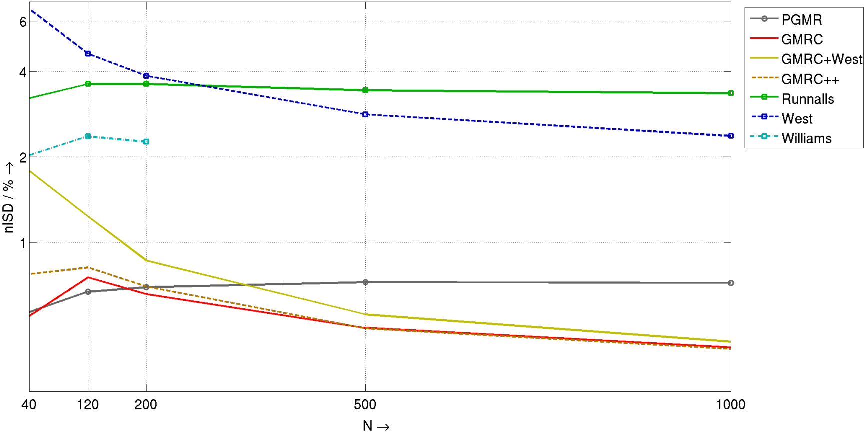
Simulation Setup

- Office PC (Intel Core2 Duo E8400)
- OpenSUSE 11.0
- Matlab 7.7.0 (R2008b)

- reduction of mixtures with $N \in \{40, 120, 200, 500, 1000\}$ components down to $K = 10$
- each evaluated with 1 000 simulation runs

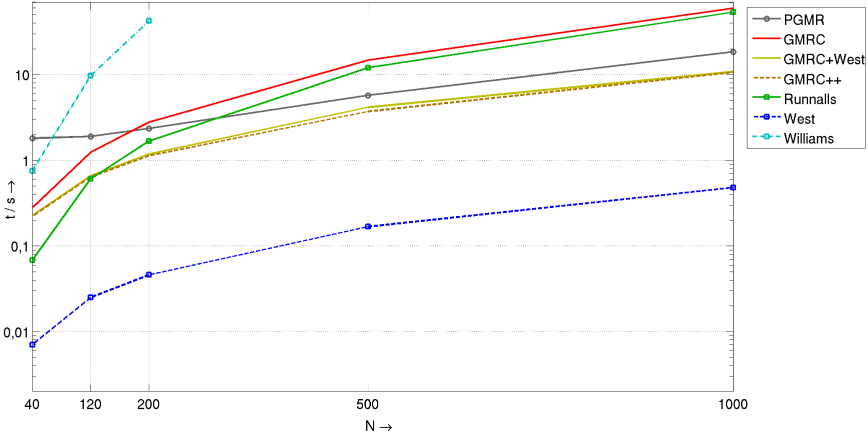
Results

Approximation Quality



Results

Running Time



Results

Impact of Individual Steps

algorithm		running time	norm. ISD
GMRC	complete	2.793 ± 0.052s	0.658 ± 0.494
	w. random init.	1.135 ± 0.045s	1.272 ± 1.561
	w/o clustering	1.742 ± 0.043s	0.774 ± 0.872
	w/o refinement	2.737 ± 0.036s	1.697 ± 0.432
Runnalls		1.678 ± 0.024s	3.606 ± 0.752

(initialization with Runnalls' algorithm; $N = 200$, $K = 10$)

- a good initial solution is mandatory
- clustering step primarily improves variance
- refinement has single-most impact on approximation quality

Results

Impact of Individual Steps

algorithm		running time	norm. ISD
GMRC	complete	2.793 ± 0.052s	0.658 ± 0.494
	w. random init.	1.135 ± 0.045s	1.272 ± 1.561
	w/o clustering	1.742 ± 0.043s	0.774 ± 0.872
	w/o refinement	2.737 ± 0.036s	1.697 ± 0.432
Runnalls		1.678 ± 0.024s	3.606 ± 0.752

(initialization with Runnalls' algorithm; $N = 200$, $K = 10$)

- a good initial solution is mandatory
- clustering step primarily improves variance
- refinement has single-most impact on approximation quality

Results

Impact of Individual Steps

algorithm		running time	norm. ISD
GMRC	complete	2.793 ± 0.052s	0.658 ± 0.494
	w. random init.	1.135 ± 0.045s	1.272 ± 1.561
	w/o clustering	1.742 ± 0.043s	0.774 ± 0.872
	w/o refinement	2.737 ± 0.036s	1.697 ± 0.432
Runnalls		1.678 ± 0.024s	3.606 ± 0.752

(initialization with Runnalls' algorithm; $N = 200$, $K = 10$)

- a good initial solution is mandatory
- clustering step primarily improves variance
- refinement has single-most impact on approximation quality

Results

Impact of Individual Steps

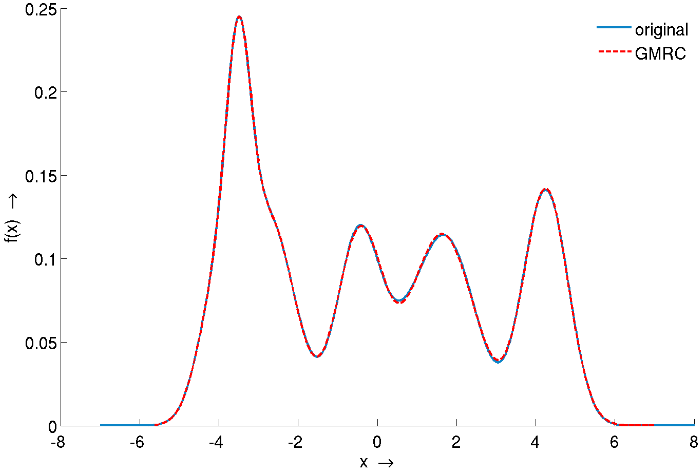
algorithm		running time	norm. ISD
GMRC	complete	2.793 ± 0.052s	0.658 ± 0.494
	w. random init.	1.135 ± 0.045s	1.272 ± 1.561
	w/o clustering	1.742 ± 0.043s	0.774 ± 0.872
	w/o refinement	2.737 ± 0.036s	1.697 ± 0.432
Runnalls		1.678 ± 0.024s	3.606 ± 0.752

(initialization with Runnalls' algorithm; $N = 200$, $K = 10$)

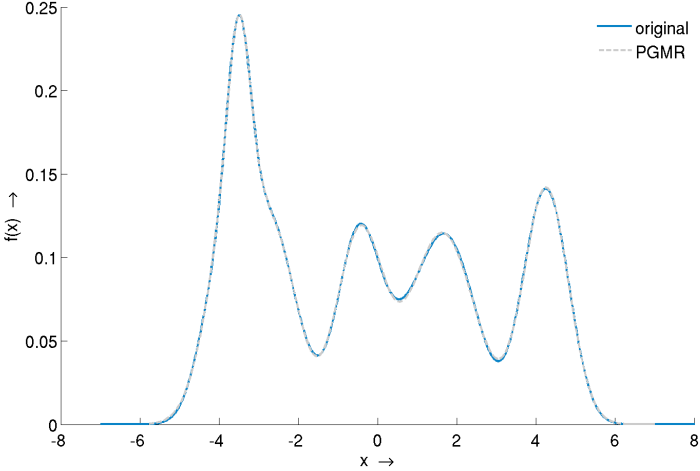
- a good initial solution is mandatory
- clustering step primarily improves variance
- refinement has single-most impact on approximation quality

Results

Visualization

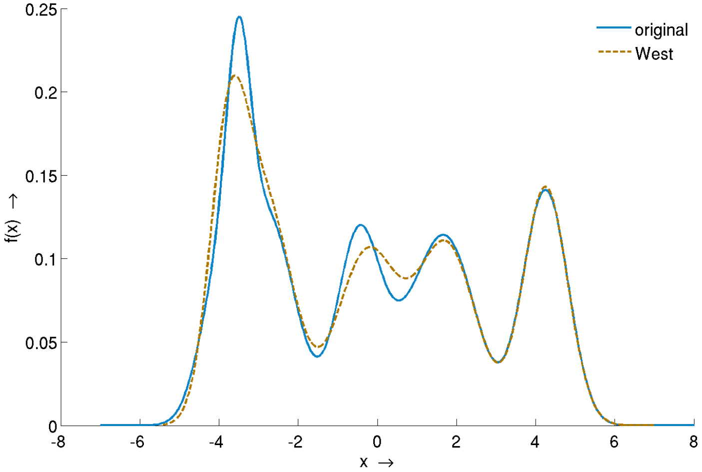


Results
Visualization



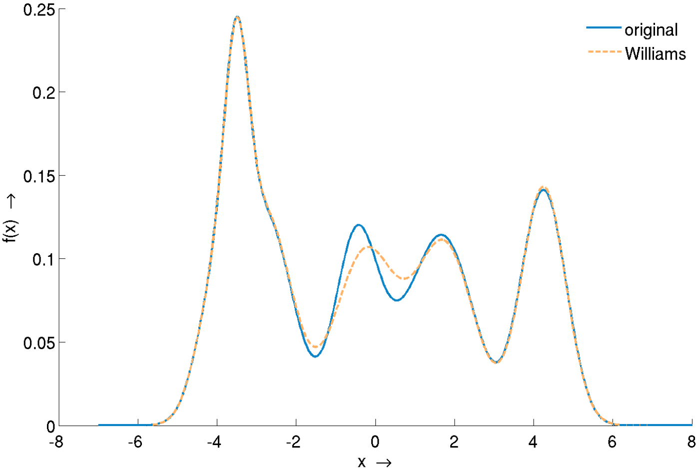
Results

Visualization



Results

Visualization



Conclusion

Karlsruhe Institute of Technology

Summary

- novel top-down, global reduction algorithm
- competitive w.r.t. current state-of-the-art (PGMR)
- combines algorithmic and numerical ideas

Outlook

- extension to multivariate Gaussian mixtures
- impact of different clustering methods and deviation measures
- adaptive reduction of components

Conclusion

Karlsruhe Institute of Technology

Summary

- novel top-down, global reduction algorithm
- competitive w.r.t. current state-of-the-art (PGMR)
- combines algorithmic and numerical ideas

Outlook

- extension to multivariate Gaussian mixtures
- impact of different clustering methods and deviation measures
- adaptive reduction of components

Thank you for your attention!

time for questions