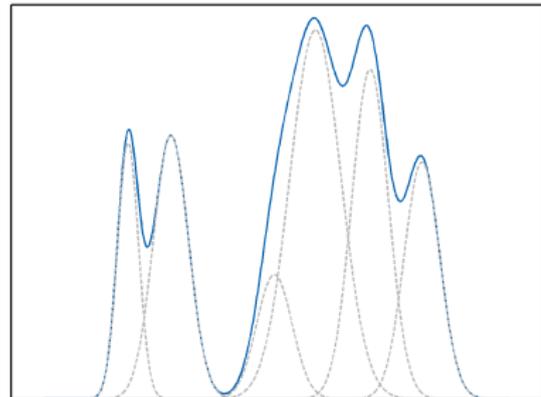
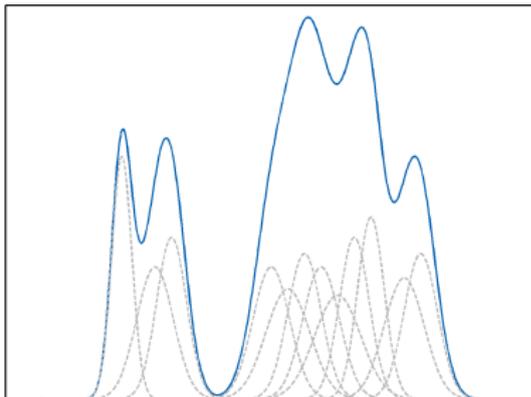


Gaussian Mixture Reduction via Clustering

Dennis Schieferdecker – schiefer@ira.uka.de
Marco Huber – marco.huber@ieee.org

GRK 1194: Self-organizing Sensor-Actuator-Networks

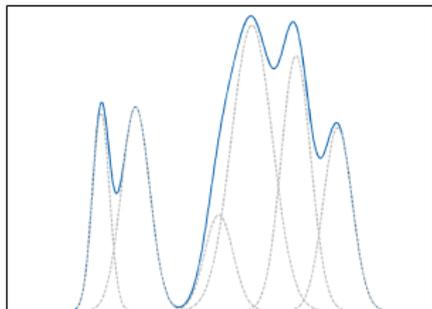


Gaussian Mixture Density

- 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
- used to model probability density functions in estimation algorithms
 - target tracking,
 - machine learning,
 - speaker recognition,
 - ...

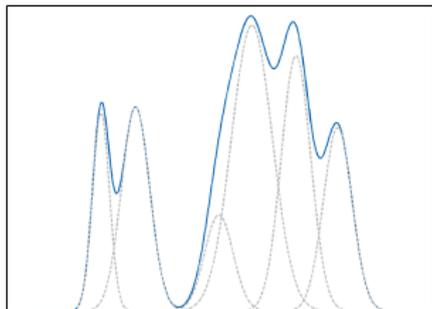


Gaussian Mixture Density

- 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
- used to model probability density functions in estimation algorithms
 - target tracking,
 - machine learning,
 - speaker recognition,
 - ...

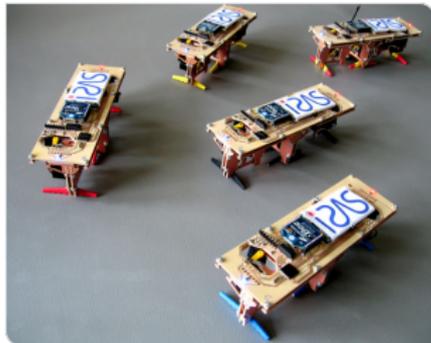


Gaussian Mixture Density

- 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
- used to model probability density functions in estimation algorithms
 - target tracking,
 - machine learning,
 - speaker recognition,
 - ...



Introduction

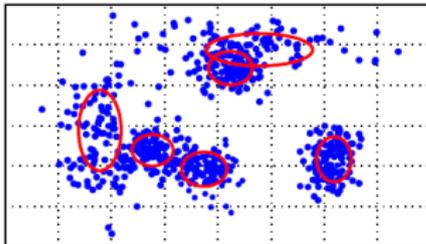
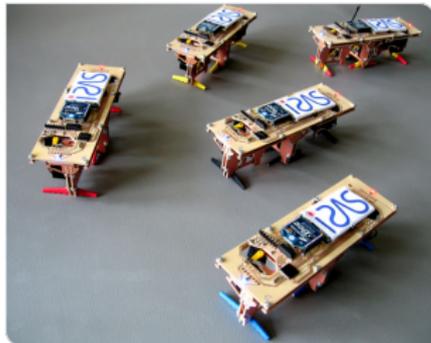
Gaussian Mixtures

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
- used to model probability density functions in estimation algorithms
 - target tracking,
 - machine learning,
 - speaker recognition,
 -

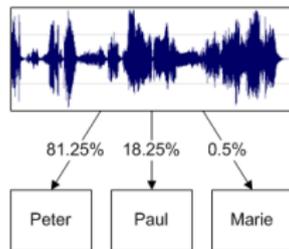
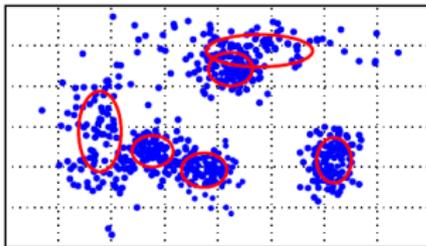
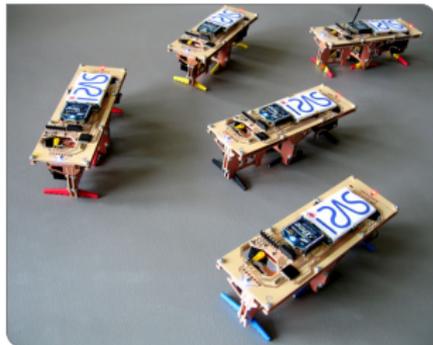


Gaussian Mixture Density

- 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
- used to model probability density functions in estimation algorithms
 - target tracking,
 - machine learning,
 - speaker recognition,
 - ...



Problem Description

Gaussian Mixture Reduction

Problem in Applications

- recursive processing
 - multiplication of Gaussian mixtures
 - convolution of Gaussian mixtures
- number of components grows **exponentially**

Solution

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

Problem Description

Gaussian Mixture Reduction

Problem in Applications

- recursive processing
 - multiplication of Gaussian mixtures
 - convolution of Gaussian mixtures
- number of components grows **exponentially**

Solution

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

Reduction Methods

Existing Algorithms

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (local, global, hybrid)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

Reduction Methods

Existing Algorithms

top-down approaches

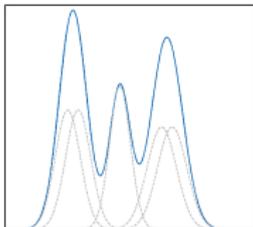
- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (local, global, hybrid)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- **start with full mixture**
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (local, global, hybrid)



bottom-up approaches

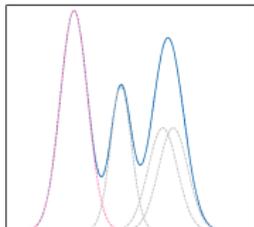
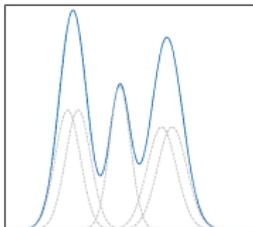
- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (local, global, hybrid)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

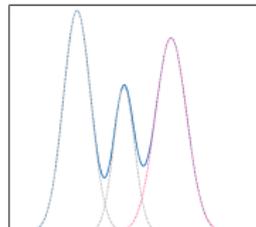
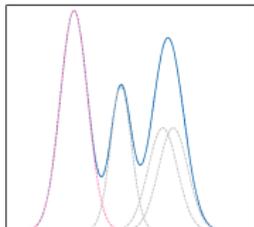
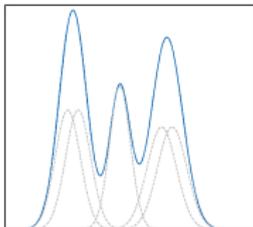


top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (local, global, hybrid)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

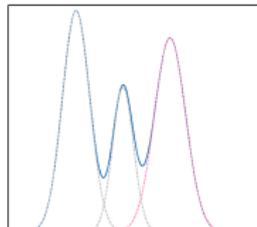
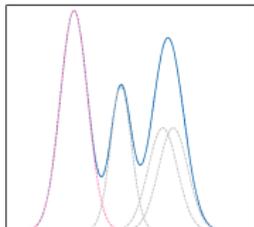
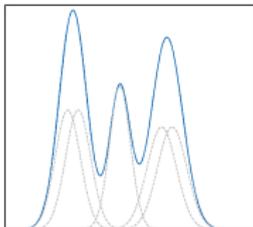


top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (**local**, **global**, **hybrid**)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

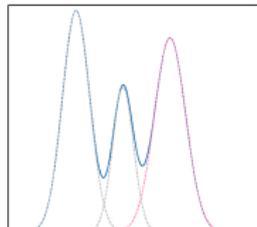
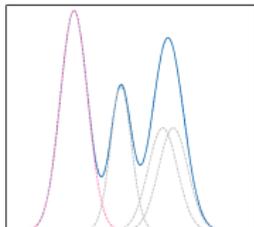
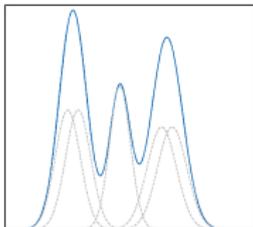


top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR



Reduction Methods

Existing Algorithms

top-down approaches

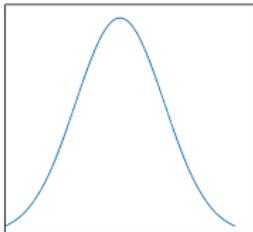
- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

bottom-up approaches

- **constructive methods**
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

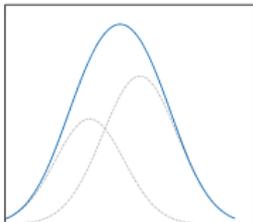


bottom-up approaches

- constructive methods
- **start with one component**
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

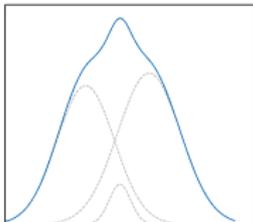


bottom-up approaches

- constructive methods
- start with one component
- **adaptively add/remove components as required**
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)



bottom-up approaches

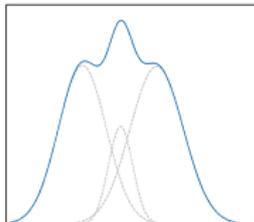
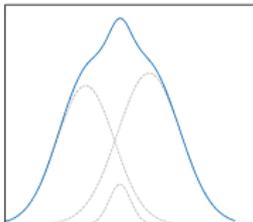
- constructive methods
- start with one component
- **adaptively add/remove components as required**
- progressive convergence towards full mixture
- PGMR

top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

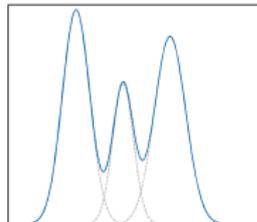
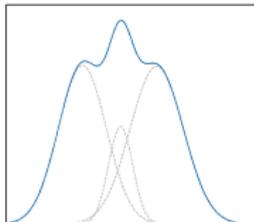
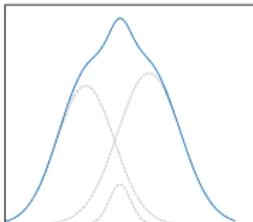


top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- PGMR

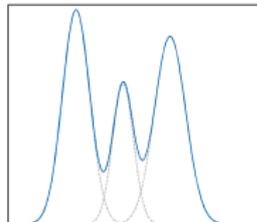
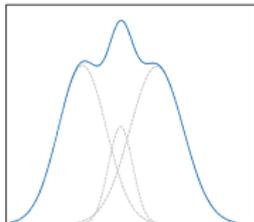
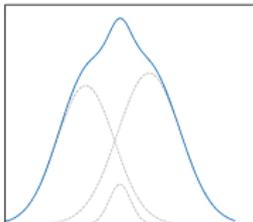


top-down approaches

- greedy methods
- start with full mixture
- iteratively replace a set of Gaussians with a smaller set
- sets can be chosen, using different deviation measures (West, Williams, Runnalls)

bottom-up approaches

- constructive methods
- start with one component
- adaptively add/remove components as required
- progressive convergence towards full mixture
- **PGMR - state-of-the-art**



Gaussian Mixture Reduction via Clustering (GMRC)

- top-down approach using a global deviation measure
- three-step algorithm:



Basic Operation

- quickly determine a rough initial solution
- push solution towards a good local optimum by local search
- refine solution using numerical methods

Gaussian Mixture Reduction via Clustering (GMRC)

- **top-down** approach using a **global** deviation measure
- three-step algorithm:



Basic Operation

- **quickly determine a rough initial solution**
- push solution towards a good local optimum by local search
- refine solution using numerical methods

Gaussian Mixture Reduction via Clustering (GMRC)

- top-down approach using a global deviation measure
- three-step algorithm:



Basic Operation

- quickly determine a rough initial solution
- push solution towards a good local optimum by local search
- refine solution using numerical methods

Gaussian Mixture Reduction via Clustering (GMRC)

- top-down approach using a global deviation measure
- three-step algorithm:



Basic Operation

- quickly determine a rough initial solution
- push solution towards a good local optimum by local search
- refine solution using numerical methods

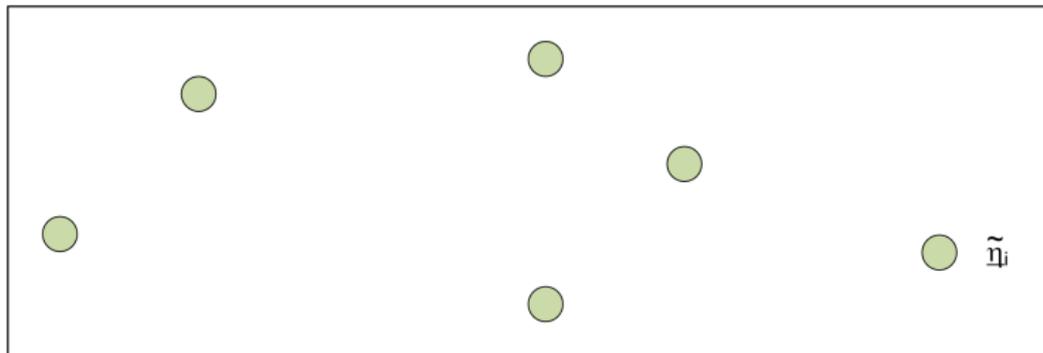
Preprocessing

LocalSearch

Refinement

Conception

- each component $\tilde{\eta}_i$ of a mixture can be mapped to a point (site) in a two-dimensional space
- distances between points correspond to the selected deviation measure

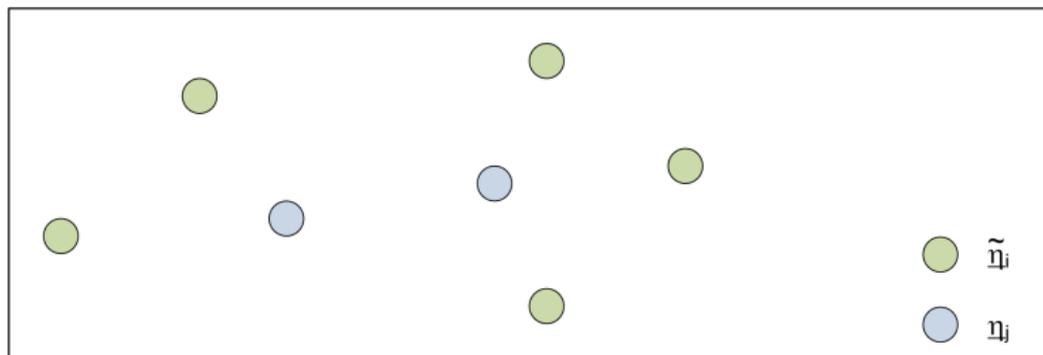
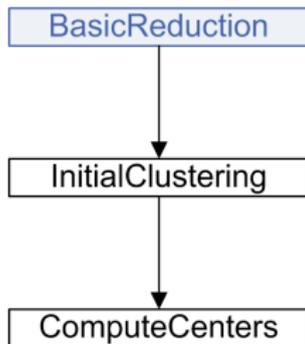


Clustering Method



BasicReduction

- compute an initial solution $\underline{\eta}$ for our problem (i.e. using West's or Runnalls' algorithm)
- the components $\underline{\eta}_j$ of the reduced mixture correspond to **preliminary cluster centers**



Clustering Method

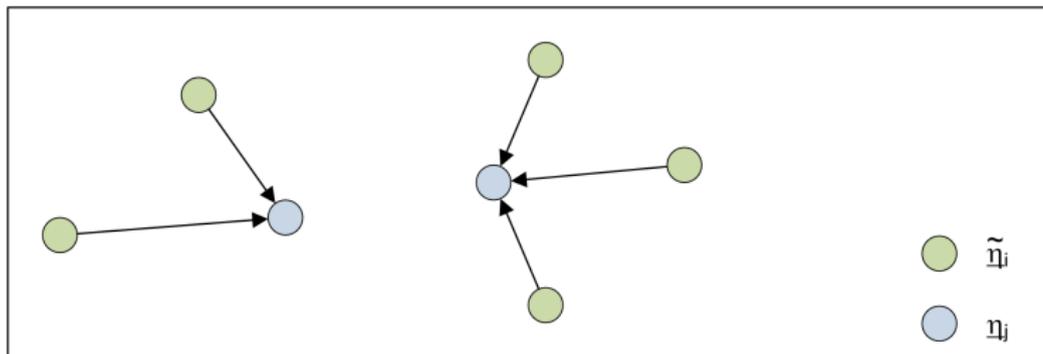
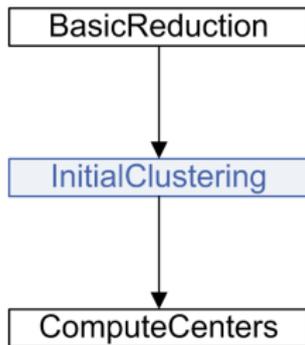


InitialClustering

- associate each original component (site) $\tilde{\eta}_i$ with the nearest one η_j of the reduced mixture (preliminary cluster center),

- using the Integrated Squared Distance (ISD):

$$d_{ISD}(\tilde{f}(x; \tilde{\eta}_i), f(x; \eta_j)) = \int_{\mathbb{R}} (\tilde{f}(x; \tilde{\eta}_i) - f(x; \eta_j))^2 dx$$

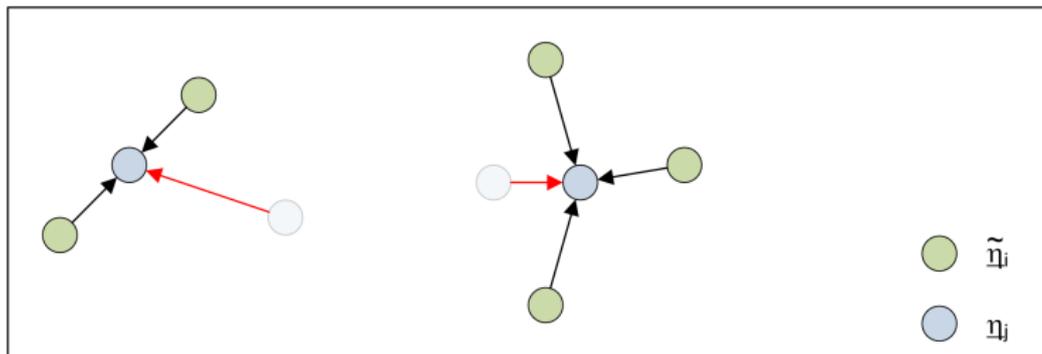
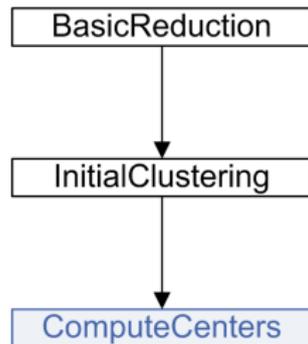


Clustering Method



ComputeCenters

- replace each reduced component η_j with a new one retaining mean and variance of the sum of the associated original components $\tilde{\eta}_i$
- equivalent to computing the **center-of-mass** of the sites associated to each center

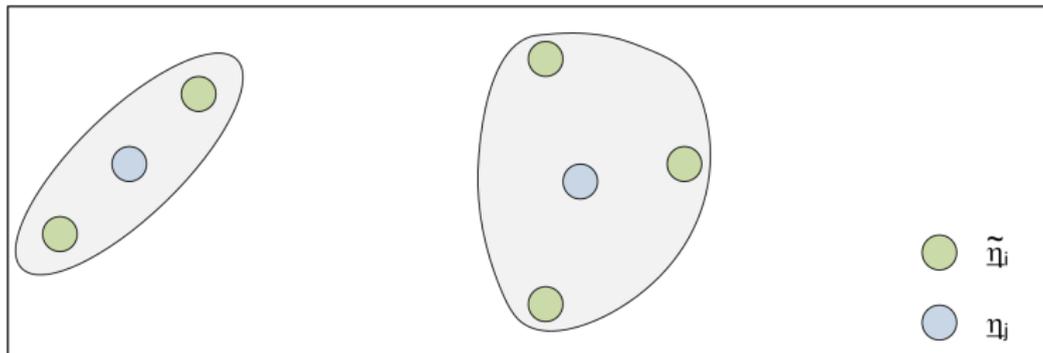
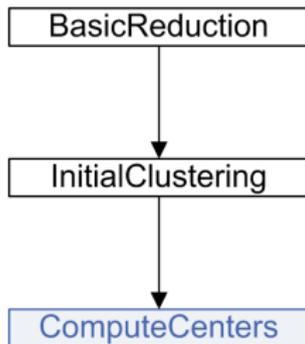


Clustering Method



ComputeCenters

- replace each reduced component η_j with a new one retaining mean and variance of the sum of the associated original components $\tilde{\eta}_i$
- equivalent to computing the **center-of-mass** of the sites associated to each center



Clustering Method

Preprocessing

Local Search

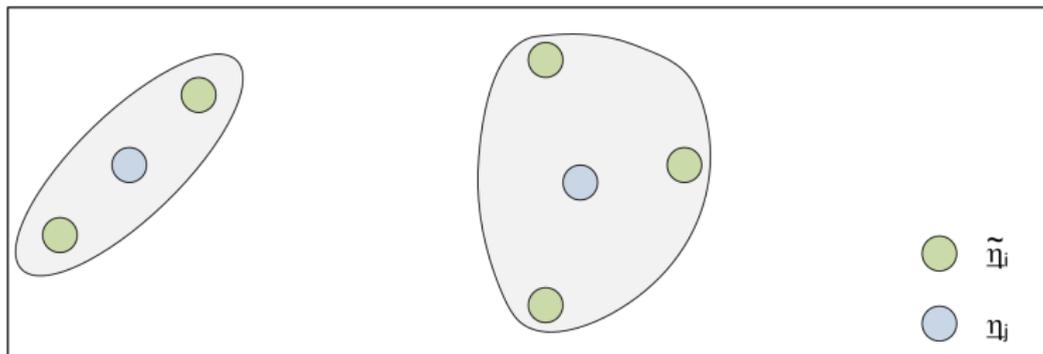
Refinement

Local Search

- greedy approach
- based on Lloyd's algorithm (**k-means**)

Basic Operation

- iteratively find the best association for each site $\tilde{\eta}_i$ to a center η_j ,
- minimizing the selected deviation measure



Clustering Method

Preprocessing

LocalSearch

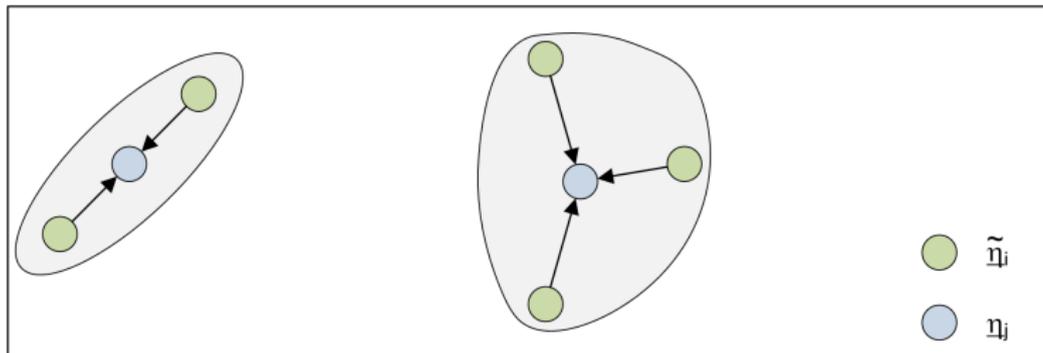
Refinement

Local Search

- greedy approach
- based on Lloyd's algorithm (**k-means**)

Basic Operation

- iteratively find the best association for each site $\tilde{\eta}_i$ to a center η_j ,
- minimizing the selected deviation measure



Clustering Method

Preprocessing

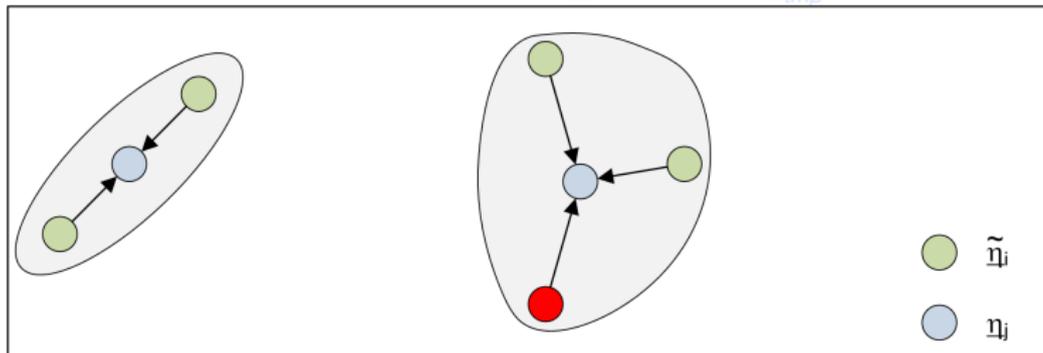
LocalSearch

Refinement

Finding the Best Association

- associate a site $\tilde{\eta}_i$ with one of the current centers η_j
- recompute and temporarily replace the affected centers η_{tmp}
- determine **normalized ISD** (nISD) between the original and the current temporary reduced mixture:

$$d_{nISD}(\tilde{f}(x; \tilde{\eta}), f(x; \eta_{tmp})) = \frac{\int_{\mathbb{R}} (\tilde{f}(x; \tilde{\eta}) - f(x; \eta_{tmp}))^2 dx}{\int_{\mathbb{R}} \tilde{f}(x; \tilde{\eta})^2 dx + \int_{\mathbb{R}} f(x; \eta_{tmp})^2 dx}$$



Clustering Method

Preprocessing

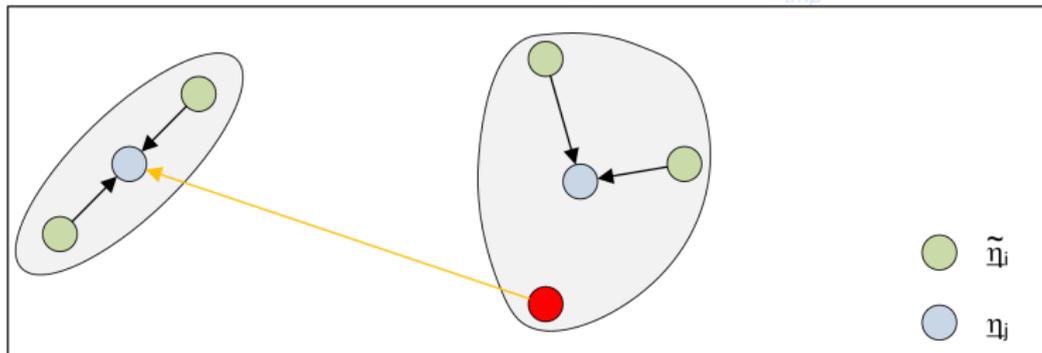
LocalSearch

Refinement

Finding the Best Association

- associate a site $\tilde{\eta}_i$ with one of the current centers η_j
- recompute and temporarily replace the affected centers η_{tmp}
- determine **normalized ISD** (nISD) between the original and the current temporary reduced mixture:

$$d_{nISD}(\tilde{f}(x; \tilde{\eta}), f(x; \eta_{tmp})) = \frac{\int_{\mathbb{R}} (\tilde{f}(x; \tilde{\eta}) - f(x; \eta_{tmp}))^2 dx}{\int_{\mathbb{R}} \tilde{f}(x; \tilde{\eta})^2 dx + \int_{\mathbb{R}} f(x; \eta_{tmp})^2 dx}$$



Clustering Method

Preprocessing

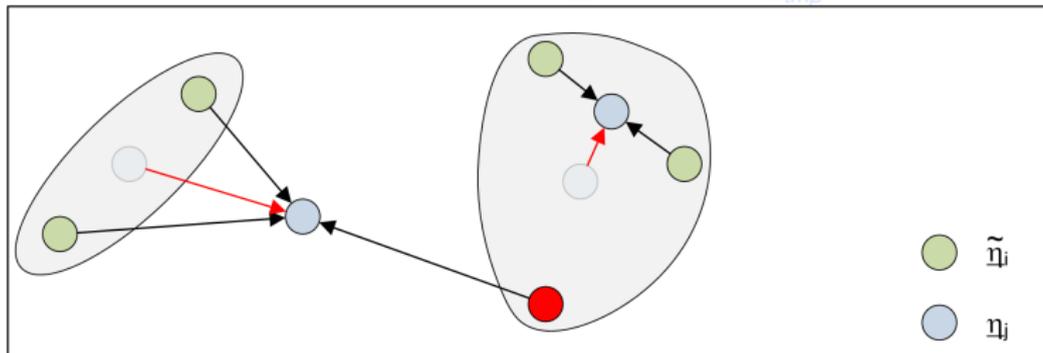
LocalSearch

Refinement

Finding the Best Association

- associate a site $\tilde{\eta}_i$ with one of the current centers η_j
- recompute and temporarily replace the affected centers η_{tmp}
- determine **normalized ISD** (nISD) between the original and the current temporary reduced mixture:

$$d_{nISD}(\tilde{f}(x; \underline{\tilde{\eta}}), f(x; \underline{\eta}_{tmp})) = \frac{\int_{\mathbb{R}} (\tilde{f}(x; \underline{\tilde{\eta}}) - f(x; \underline{\eta}_{tmp}))^2 dx}{\int_{\mathbb{R}} \tilde{f}(x; \underline{\tilde{\eta}})^2 dx + \int_{\mathbb{R}} f(x; \underline{\eta}_{tmp})^2 dx}$$



Clustering Method

Preprocessing

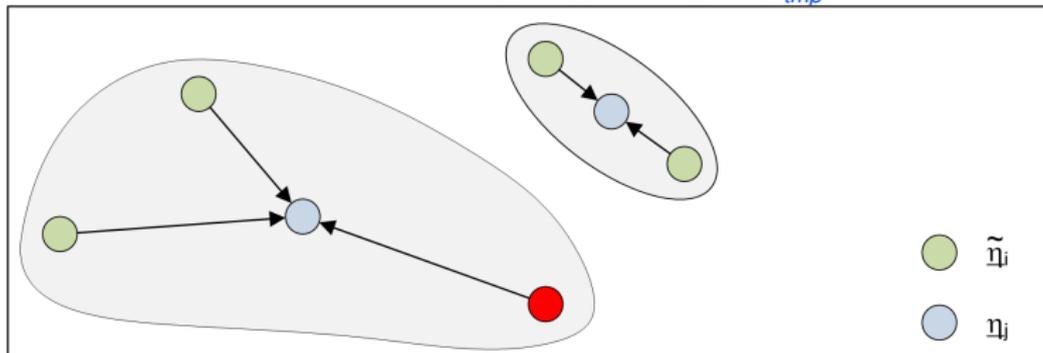
LocalSearch

Refinement

Finding the Best Association

- associate a site $\tilde{\eta}_i$ with one of the current centers η_j
- recompute and temporarily replace the affected centers η_{tmp}
- determine **normalized ISD** (nISD) between the original and the current temporary reduced mixture:

$$d_{nISD}(\tilde{f}(x; \tilde{\eta}), f(x; \eta_{tmp})) = \frac{\int_{\mathbb{R}} (\tilde{f}(x; \tilde{\eta}) - f(x; \eta_{tmp}))^2 dx}{\int_{\mathbb{R}} \tilde{f}(x; \tilde{\eta})^2 dx + \int_{\mathbb{R}} f(x; \eta_{tmp})^2 dx}$$



Clustering Method

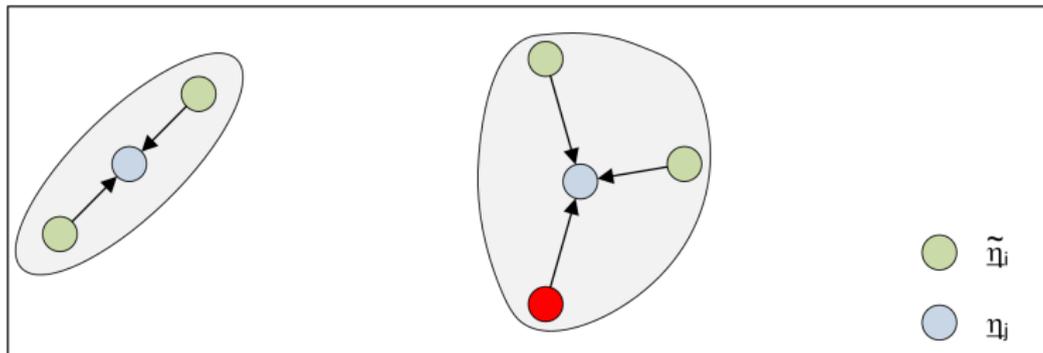
Preprocessing

LocalSearch

Refinement

Finding the Best Association – continued

- revert temporary association
- repeat for all possible associations of site $\tilde{\eta}_i$ to a center
- retain the association with the smallest deviation (i.e. the reduced mixture which is closest to the original one)



Clustering Method

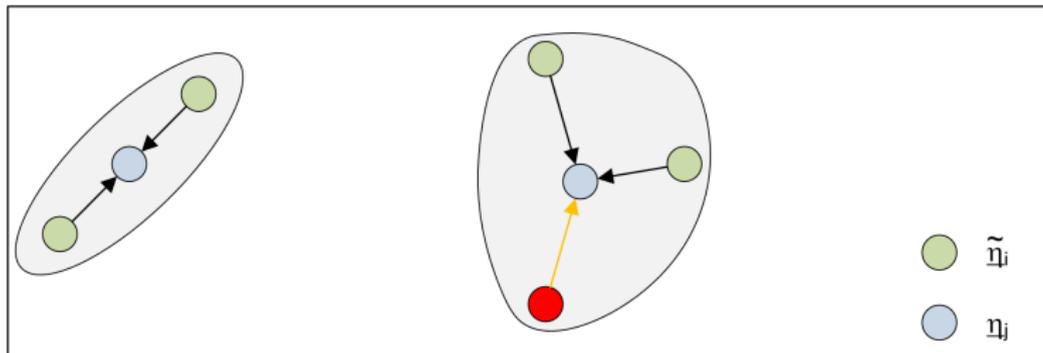
Preprocessing

LocalSearch

Refinement

Finding the Best Association – continued

- revert temporary association
- repeat for all possible associations of site $\tilde{\eta}_i$ to a center
- retain the association with the smallest deviation
(i.e. the reduced mixture which is closest to the original one)



Clustering Method

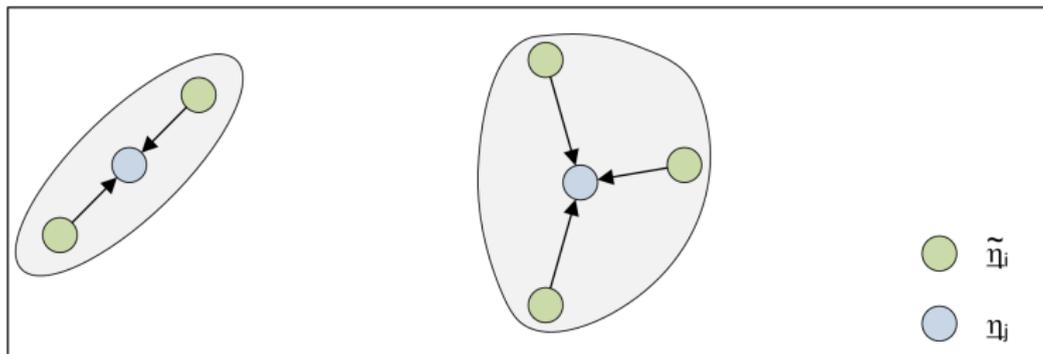
Preprocessing

LocalSearch

Refinement

Finding the Best Association – continued

- revert temporary association
- repeat for all possible associations of site $\tilde{\eta}_i$ to a center
- retain the association with the smallest deviation (i.e. the reduced mixture which is closest to the original one)



Clustering Method

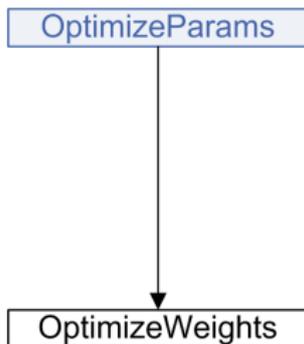


Parameter Optimization

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

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

- non-linear optimization problem
→ **Newton approach**
- finds local optimum



Weight Optimization

- solve system of linear equations

$$\begin{bmatrix} \int_{\mathbb{R}} \mathcal{N}_1^2 dx & & \\ & \ddots & \\ & & \int_{\mathbb{R}} \mathcal{N}_L^2 dx \end{bmatrix} \underline{\omega} = \sum_{j=1}^N \tilde{\omega}_j \begin{bmatrix} \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \mathcal{N}_1 dx \\ \vdots \\ \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \mathcal{N}_L dx \end{bmatrix}$$

- finds global optimum

Clustering Method

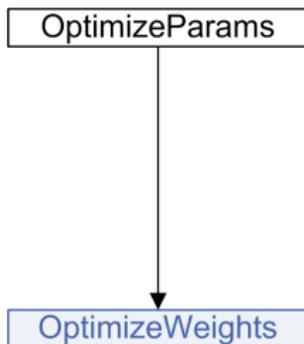


Parameter Optimization

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

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

- non-linear optimization problem
→ **Newton approach**
- finds local optimum



Weight Optimization

- solve system of linear equations

$$\begin{bmatrix} \int_{\mathbb{R}} \mathcal{N}_1^2 dx & & \\ & \ddots & \\ & & \int_{\mathbb{R}} \mathcal{N}_L^2 dx \end{bmatrix} \underline{\omega} = \sum_{j=1}^N \tilde{\omega}_j \begin{bmatrix} \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \mathcal{N}_1 dx \\ \vdots \\ \int_{\mathbb{R}} \tilde{\mathcal{N}}_j \mathcal{N}_L dx \end{bmatrix}$$

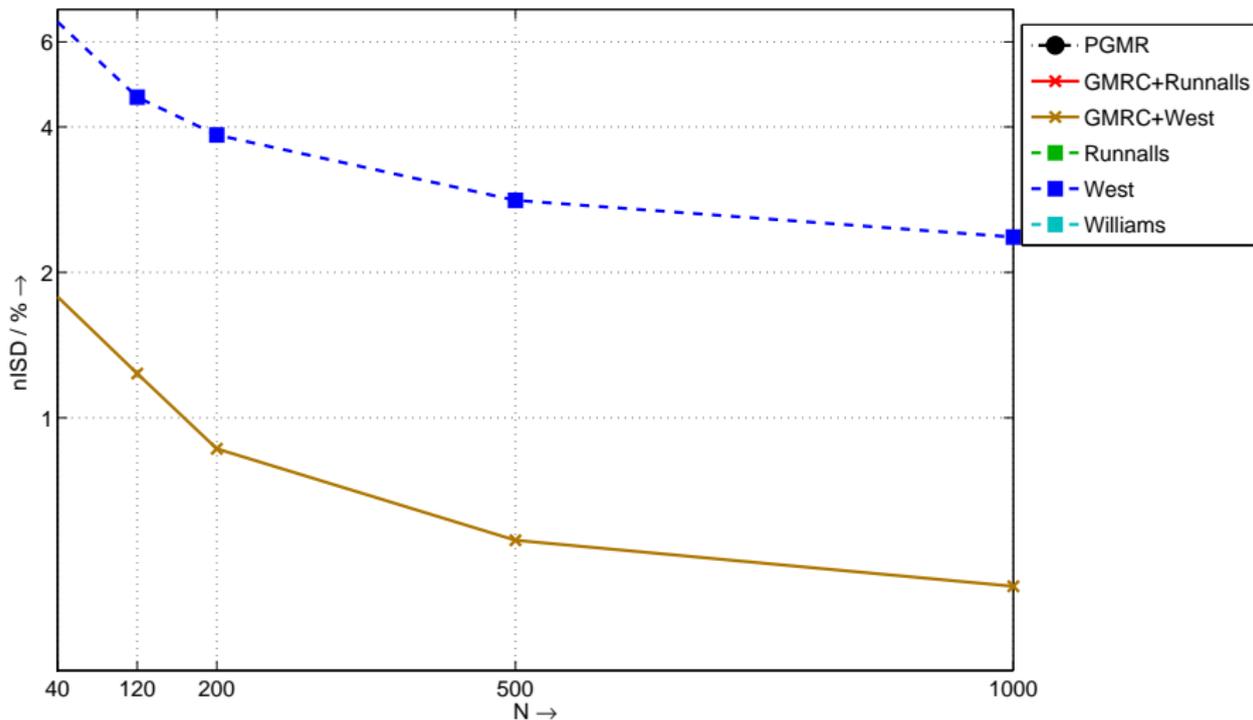
- finds global optimum

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 1000 simulation runs

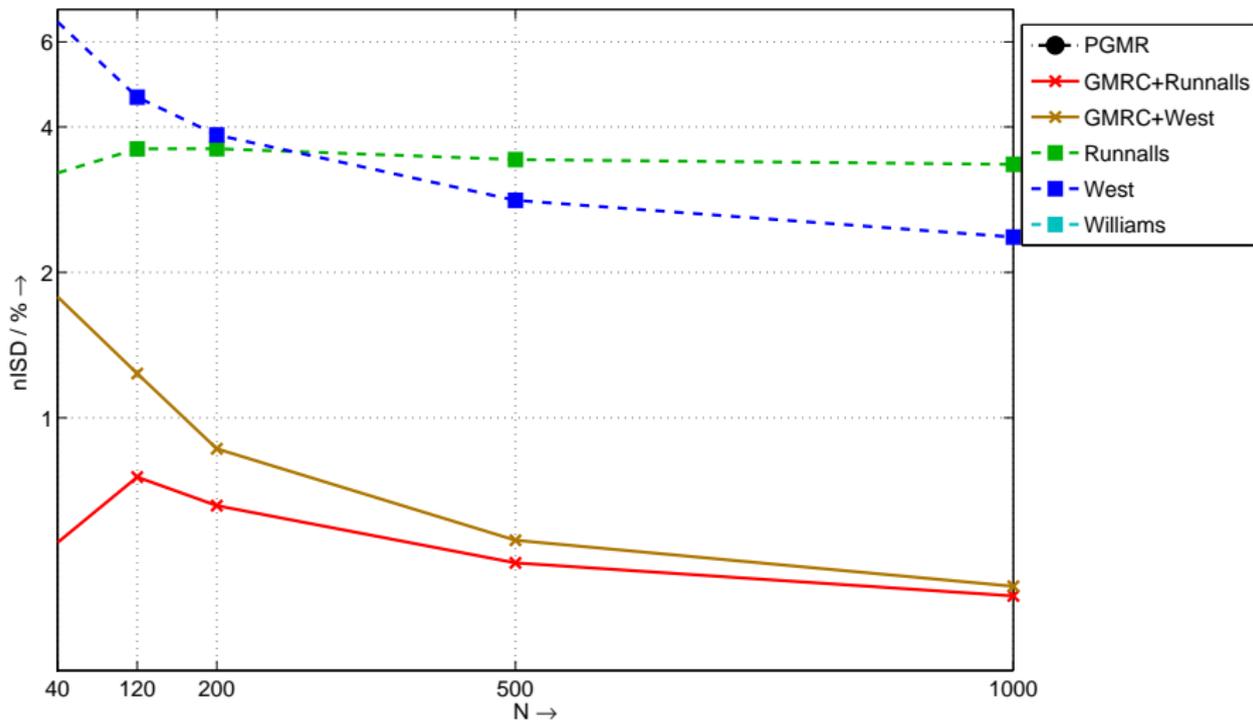
Results

Approximation Quality



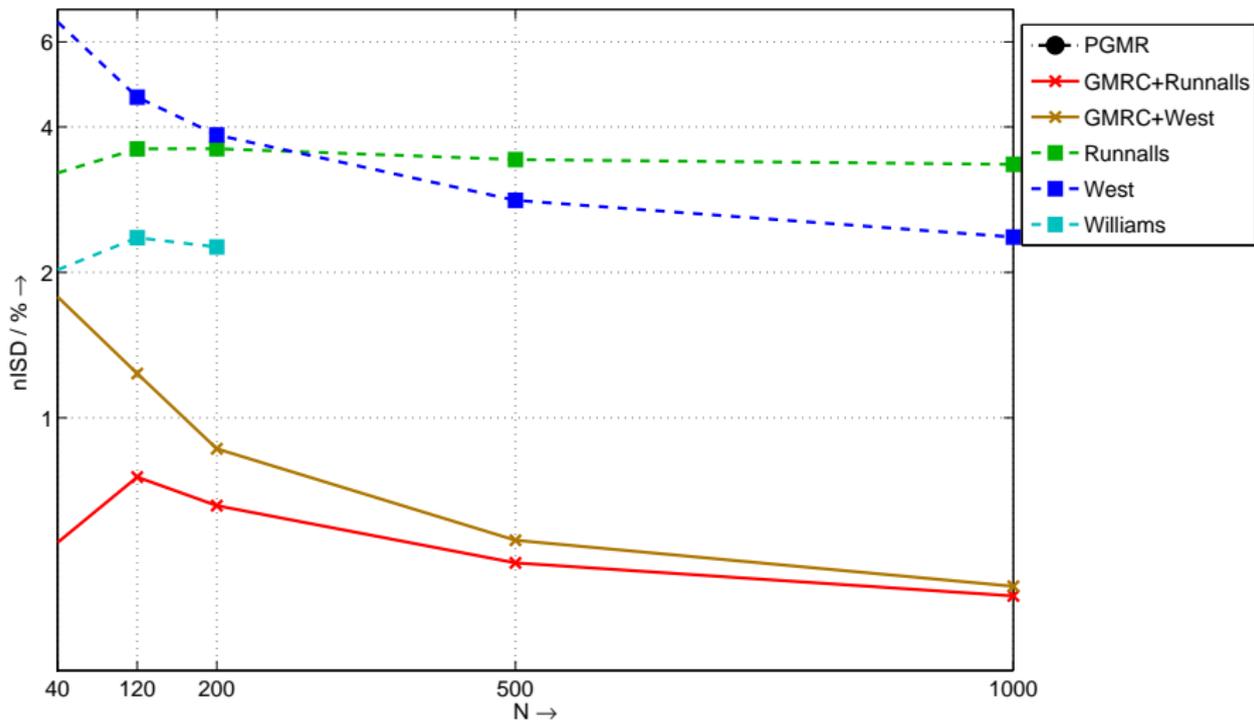
Results

Approximation Quality



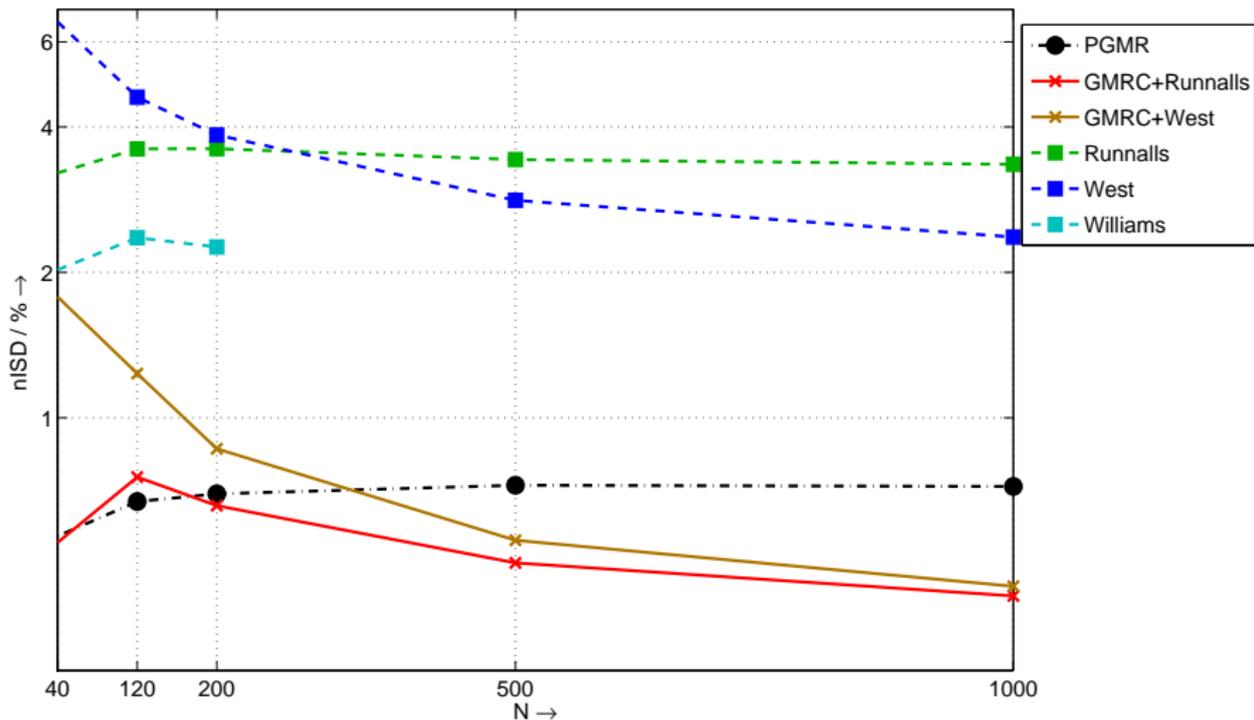
Results

Approximation Quality



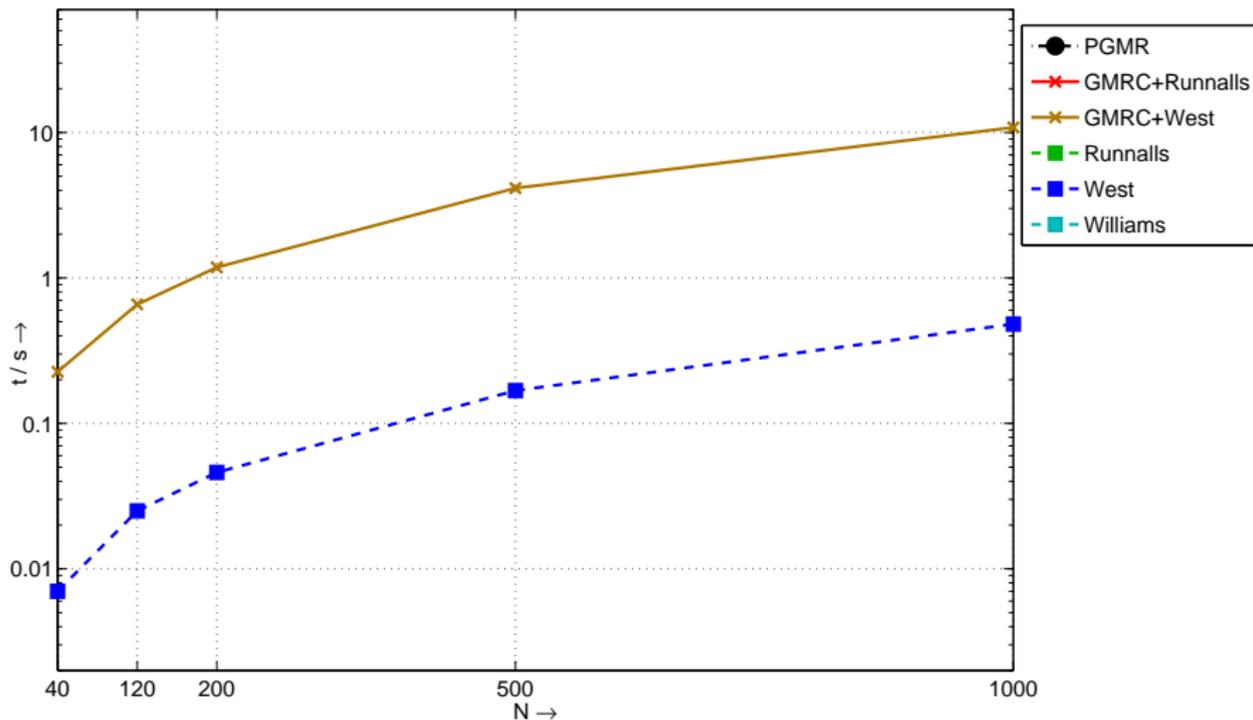
Results

Approximation Quality



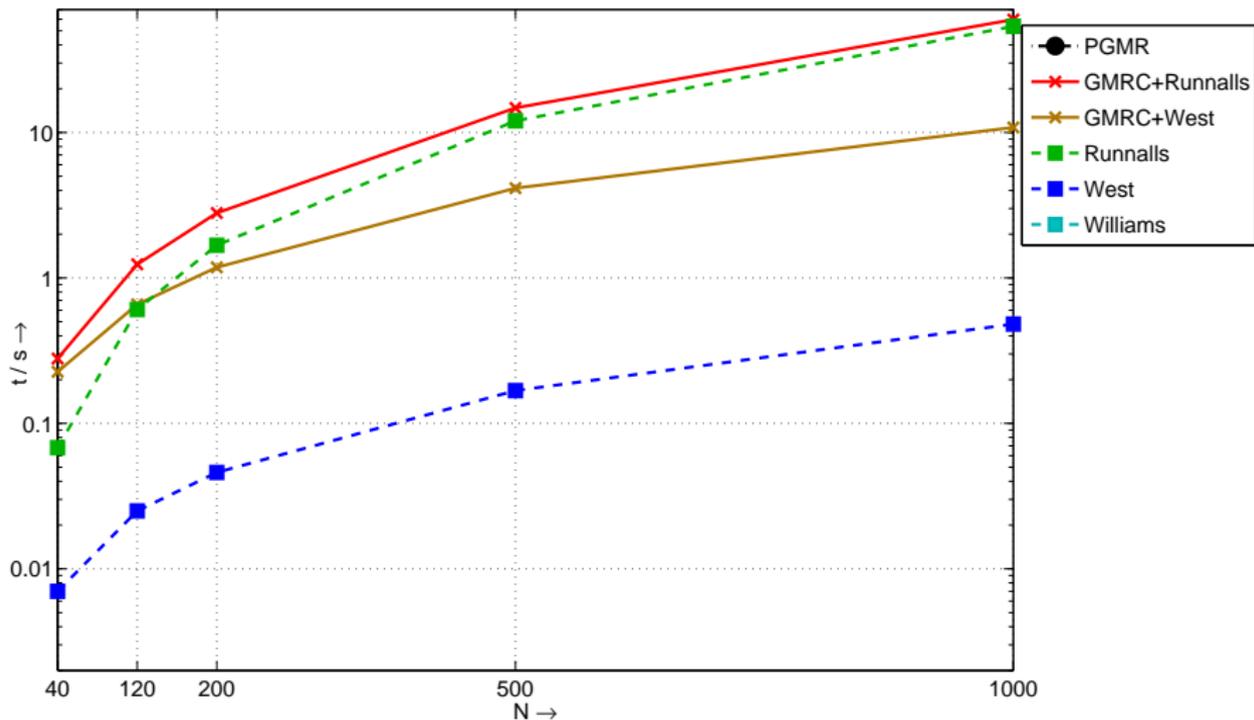
Results

Running Time



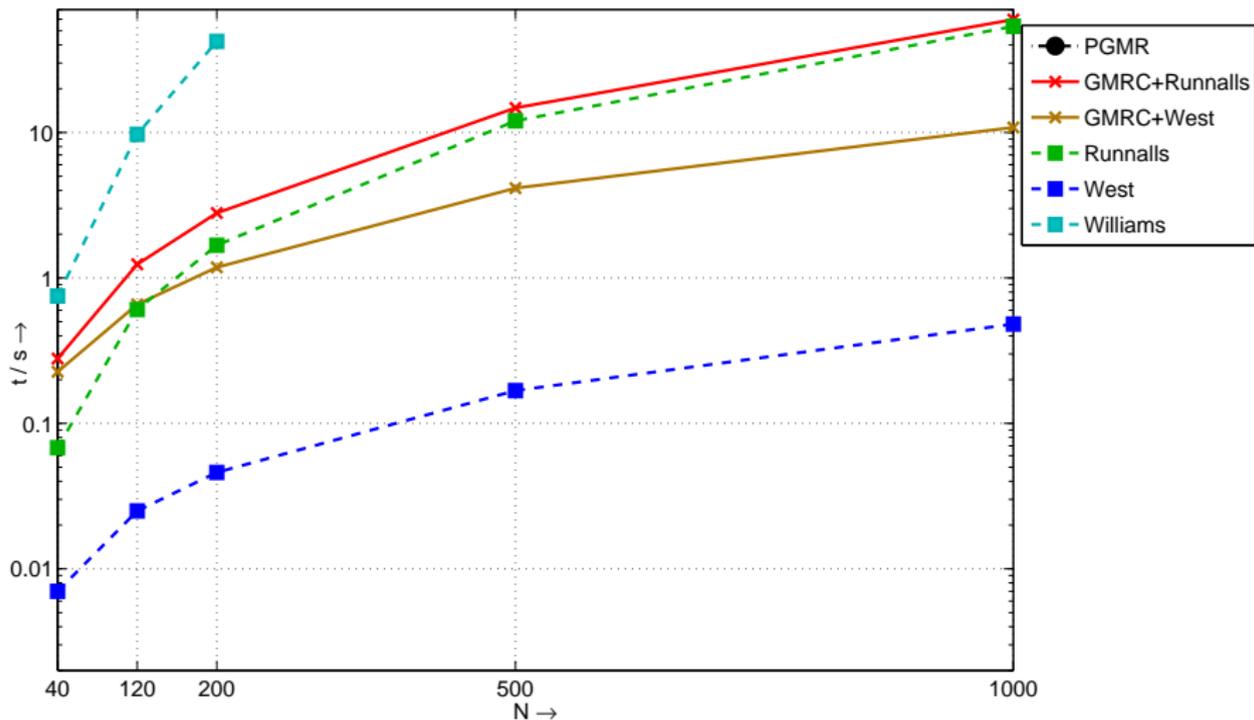
Results

Running Time



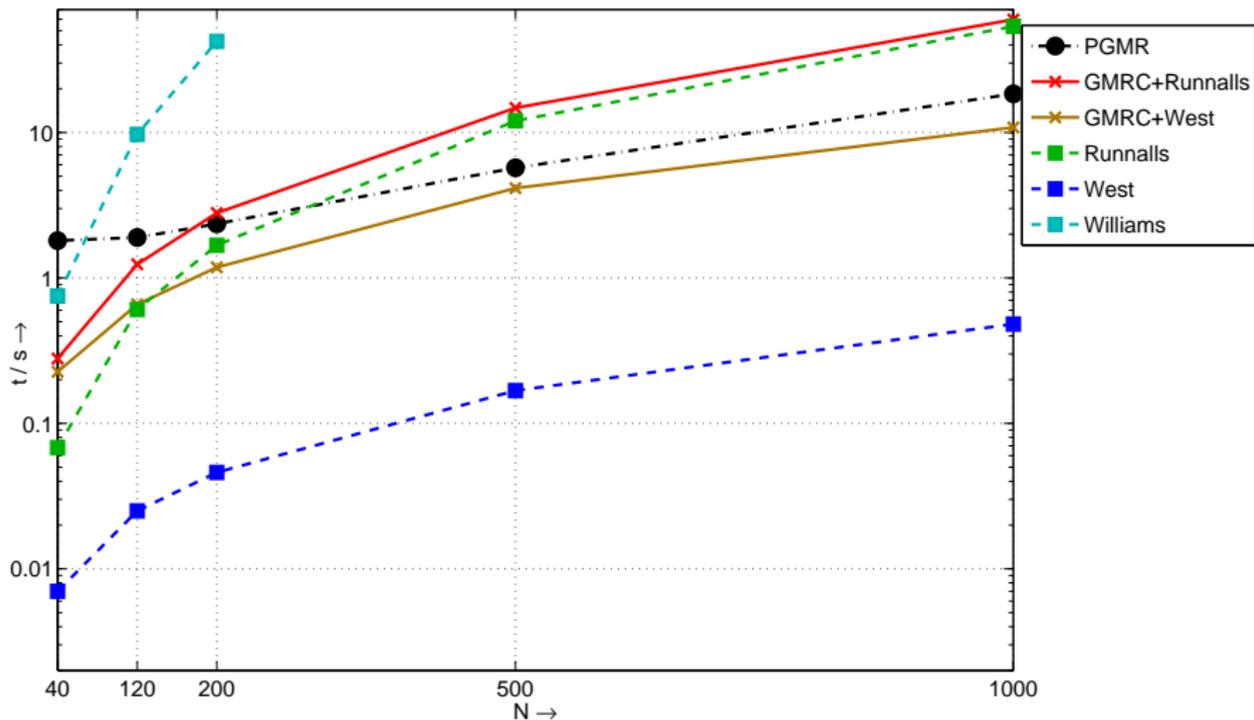
Results

Running Time



Results

Running Time



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

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

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

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

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

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

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

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

- a good initial solution is mandatory
- **local search primarily improves variance**
- refinement has single-most impact on approximation quality

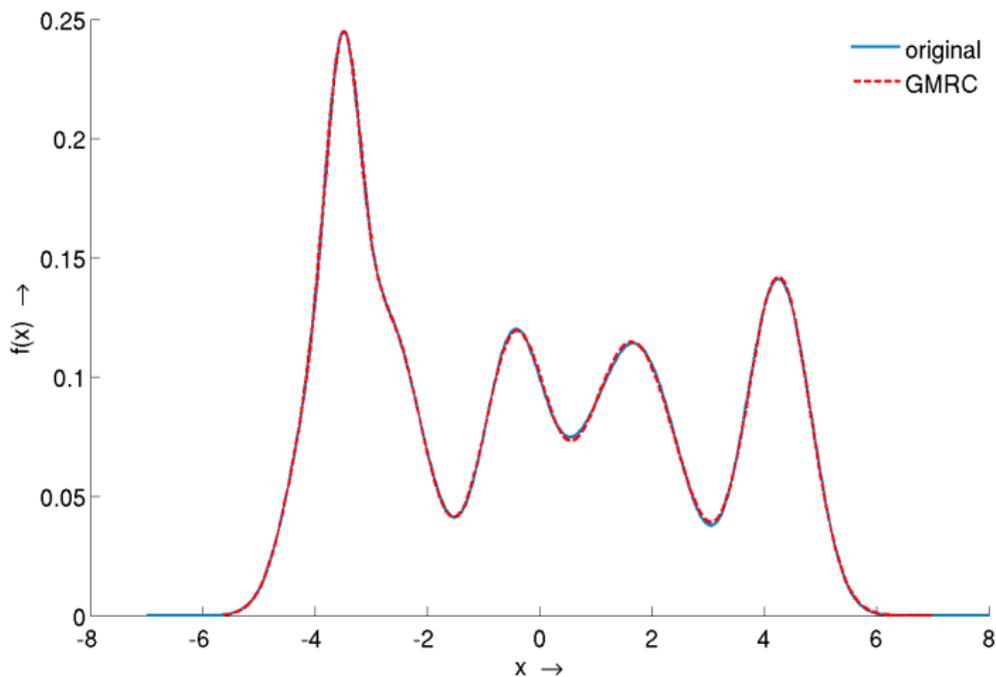
	algorithm	running time	norm. ISD
GMRC	complete	$2.793 \pm 0.052\text{s}$	0.658 ± 0.494
	w. random init.	$1.135 \pm 0.045\text{s}$	1.272 ± 1.561
	w/o local search	$1.742 \pm 0.043\text{s}$	0.774 ± 0.872
	w/o refinement	$2.737 \pm 0.036\text{s}$	1.697 ± 0.432
Runnalls		$1.678 \pm 0.024\text{s}$	3.606 ±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

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

- a good initial solution is mandatory
- local search primarily improves variance
- **refinement has single-most impact on approximation quality**

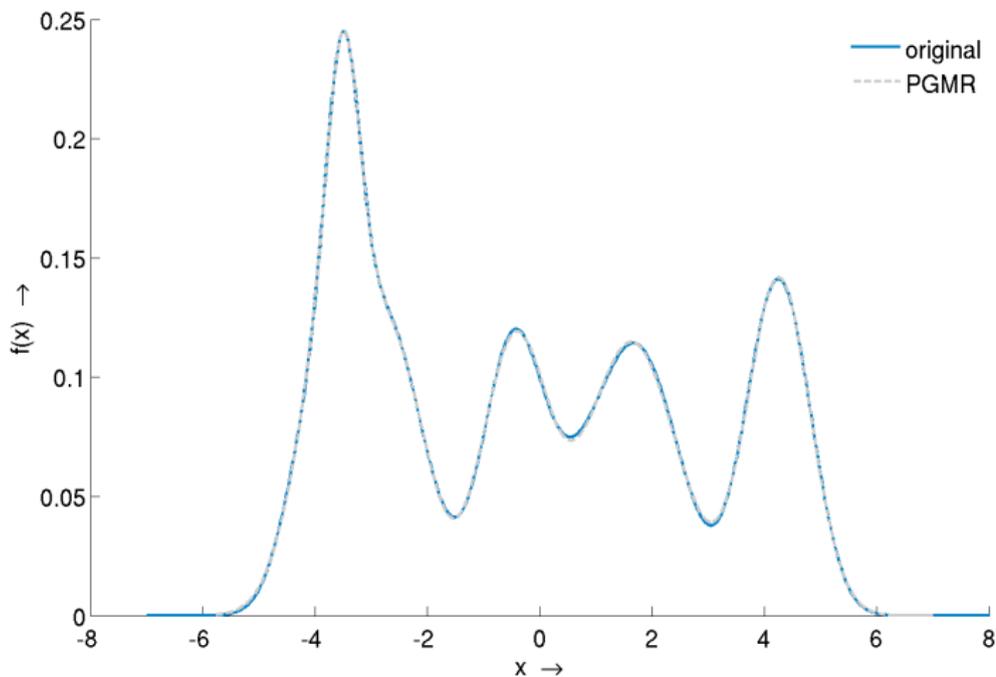
Results

Visualization



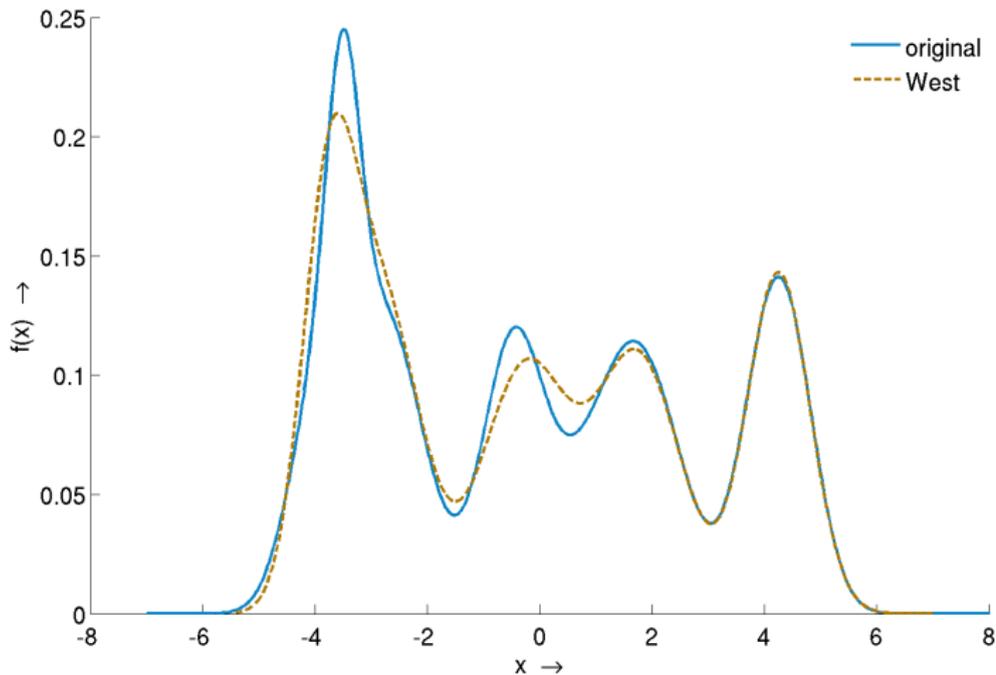
Results

Visualization



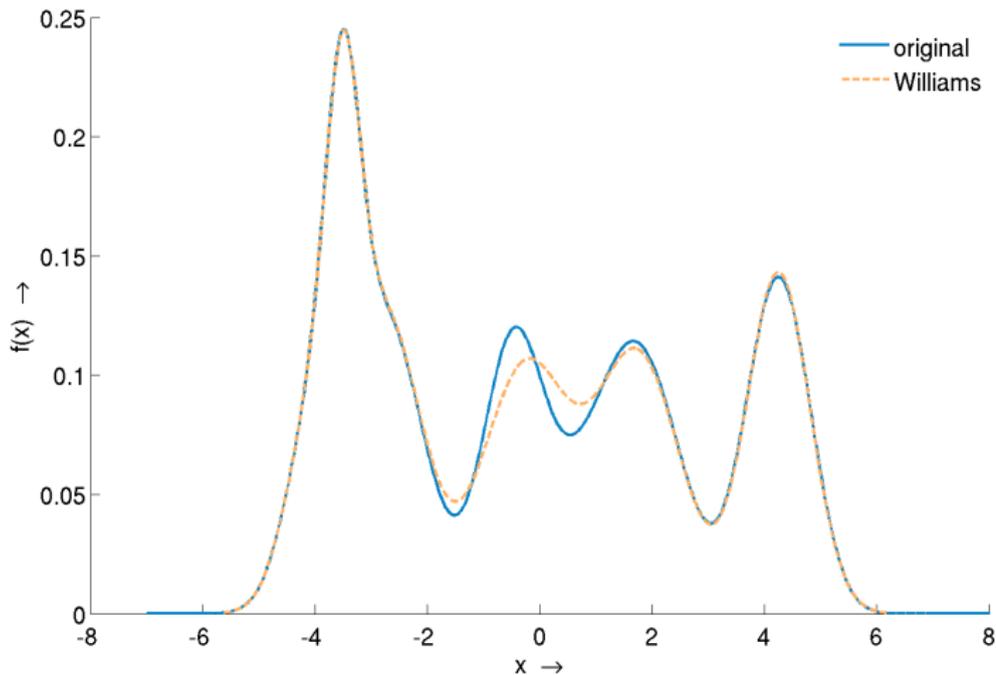
Results

Visualization



Results

Visualization



Summary

- novel mixture reduction algorithm
 - **top-down** approach, using a **global** deviation measure
 - based on **k-means** clustering method
 - combines **discrete** and **continuous** optimization methods
- compared to the current state-of-the-art **PGMR**:
 - faster computation
 - similar approximation quality

Outlook

- extension to **multivariate** Gaussian mixtures
- refine empirical choice of
 - West's and Runnalls' algorithm in the **preprocessing step**
 - k-means as **clustering approach**
- introduce **adaptive reduction** of components

Summary

- novel mixture reduction algorithm
 - **top-down** approach, using a **global** deviation measure
 - based on **k-means** clustering method
 - combines **discrete** and **continuous** optimization methods
- compared to the current state-of-the-art **PGMR**:
 - faster computation
 - similar approximation quality

Outlook

- extension to **multivariate** Gaussian mixtures
- refine empirical choice of
 - West's and Runnalls' algorithm in the **preprocessing step**
 - k-means as **clustering approach**
- introduce **adaptive reduction** of components

Thank you for your attention!



time for questions