Accelerated Convergence Using Dynamic Mean Shift

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Kai Zhang,[+] James T. Kwok,[+] and Ming Tang[*]

[+]Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong

[*]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing, China

1 Introduction

1.1 Mean Shift

- A nonparametric mode-seeking algorithm [1] widely used in clustering, segmentation, tracking
- Basic idea: iteratively move x along the gradient of the estimated density $\nabla f(\hat{x})$ until a local maxima is reached
- Connection with bound optimization, Newton method, and EM
- Advantage: simple, robust, can identify arbitrarily shaped clusters
- Disadvantage: complexity is $O(n^2)$, convergence might be slow

1.2 Related Work and Highlights

- Reference [2], combines qusi-Newton analysis with mean shift; complexity becomes cubic of dimension.
- Reference [3], applies over-relaxed bound optimization; overshooting; clustering accuracy may degrade.
- **Our work**, dynamically update the sample set for faster convergence, called *dynamic mean shift* (DMS).
- When the data is locally Gaussian, DMS has
  - the same optimum as the standard mean shift;
  - faster converges (superlinear) than the standard one (linear);
  - the effect of a variable bandwidth mean shift even only with a fixed bandwidth parameter.

1.3 Some preliminaries

Mean shift vector at $\mathbf{x}$:

$$m(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i k'\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)}{\sum_{i=1}^{n} k'\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)} - \mathbf{x},$$

which points towards the direction of the density gradient, with adaptive step length.
Iteration defined by:

$$\mathbf{x}^{(t+1)} = m(\mathbf{x}^{(t)}) + \mathbf{x}^{(t)}, \quad t = 0, 1, 2, \ldots$$

In Gaussian distribution case

$$m(\mathbf{x}) = \mathbf{H}\frac{\hat{\nabla} f_G(\mathbf{x})}{\hat{f}_G(\mathbf{x})} = -\mathbf{H}(\mathbf{H}+\boldsymbol{\Sigma})^{-1}(\mathbf{x}-\boldsymbol{\mu}).$$

Assumption: the data set is (or at least locally) Gaussian. This has been used in the theoretical analysis of mean shift, and can be deemed as a Gaussian mixture model (GMM).

2 DMS Algorithm

1: Initialize $\mathcal{T}^{(0)} = \mathcal{S}^{(0)} = \mathcal{S}$, where S is the original sample set. For $t = 0, 1, \ldots$, do the follows

2: Update the cluster center set according to

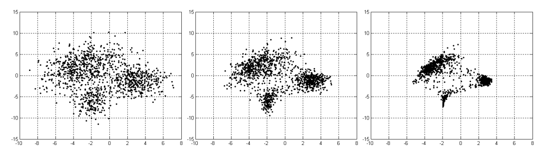
$$\mathbf{y}_i^{(t+1)} = \frac{\sum_{\mathbf{x}_i^{(t)} \in \mathcal{S}^{(t)}} K\left(\mathbf{x}_i^{(t)} - \mathbf{y}_i^{(t)}\right)\mathbf{x}_i^{(t)}}{\sum_{\mathbf{x}_i^{(t)} \in \mathcal{S}^{(t)}} K\left(\mathbf{x}_i^{(t)} - \mathbf{y}_i^{(t)}\right)}, \quad i = 1, 2, \ldots, n.$$

3: Using the shifted cluster center set $\mathcal{T}^{(t+1)} = \{\mathbf{y}_i^{(t+1)}, i = 1, 2, \ldots, n\}$ to replace the sample set

$$\mathcal{S}^{(t+1)} = \mathcal{T}^{(t+1)}.$$

4: The process is repeated until a fixed state $\mathcal{S}^{(t+1)} = \mathcal{S}^{(t)}$, or equivalently, $\mathcal{T}^{(t+1)} = \mathcal{T}^{(t)}$, is reached.

Procedure Illustration



$\mathcal{S}^{(0)}$ $\mathcal{S}^{(1)}$ $\mathcal{S}^{(2)}$

DMS needs $s = 5.1$ steps to converge. Standard mean shift (SMS) needs $s = 14.64$ steps. Clustering accuracy: DMS 95.58%, SMS 95.50%.

3 Properties of DMS Evolution

Proposition 1 *Assume that the sample set $\mathcal{S}^{(t)} = \{\mathbf{x}_i^{(t)}\}$ at the tth iteration is $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{(t)})$. After one DMS iteration, the updated sample set $\mathcal{S}^{(t+1)}$ is still a Gaussian $\mathcal{N}(\boldsymbol{\mu}, \mathbf{P}^{(t)}\boldsymbol{\Sigma}^{(t)}(\mathbf{P}^{(t)})')$, where $\mathbf{P}^{(t)} = \mathbf{I} - \mathbf{H}(\mathbf{H}+\boldsymbol{\Sigma}^{(t)})^{-1}$.*

Remark. After one DMS iteration, the sample set remains a Gaussian. The mean is unchanged, and the covariance is shrunken to $\mathbf{P}^{(t)}$.

Proposition 2 $|\boldsymbol{\Sigma}^{(t)}|$ *and* $|\mathbf{P}^{(t)}|$ *decrease with t,* $\lim_{t\to\infty} |\boldsymbol{\Sigma}^{(t)}| = \lim_{t\to\infty} |\mathbf{P}^{(t)}| = 0$.

Remark. $|\boldsymbol{\Sigma}^{(t)}|$ measures the spread of the sample set $\mathcal{S}^{(t)}$. Therefore $\mathcal{S}^{(t)}$ will gradually shrink during the DMS iteration.

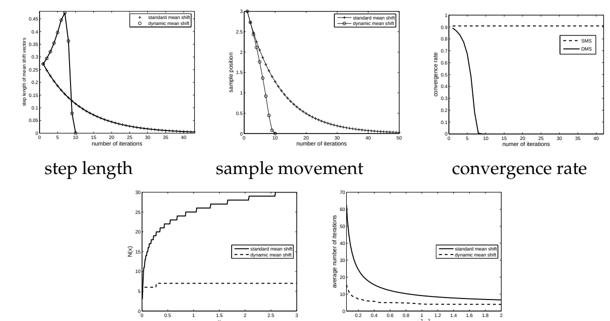
Proposition 3 *SMS converges linearly, while DMS converges superlinearly.*

Definition. The *order of convergence* is defined as p if the error $\mathbf{e}^{(t)} = \mathbf{z}^{(t)} - \mathbf{z}^*$ behaves like $\|\mathbf{e}^{(t+1)}\|_2 / \|\mathbf{e}^{(t)}\|_2^p \to c$ where $c > 0$, and z^* is the optimum. For both DMS and SMS, $p = \|\mathbf{P}^{(t)}\|_2$.

4 Examples

4.1 1-d Gaussian $\mathcal{N}(0, 1)$

Gaussian kernel; bandwidth $h = 0.32$

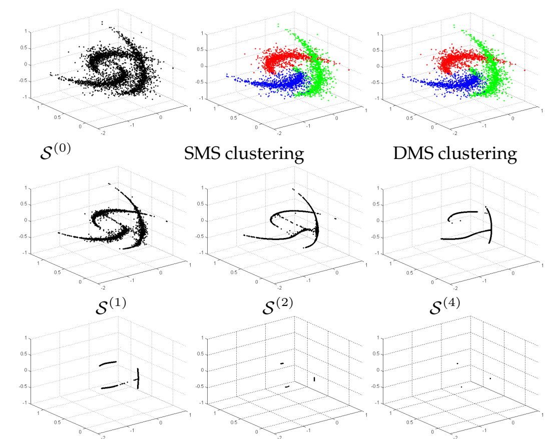


step length / sample movement / convergence rate

number of iterations

4.2 Non-Gaussian data set

Gaussian kernel; bandwidth $h = 0.15$



$\mathcal{S}^{(0)}$ / SMS clustering / DMS clustering

$\mathcal{S}^{(1)}$ $\mathcal{S}^{(2)}$ $\mathcal{S}^{(4)}$

$\mathcal{S}^{(6)}$ $\mathcal{S}^{(8)}$ $\mathcal{S}^{(9)}$

Average number of iterations: DMS takes 7.6 steps; SMS takes 26.6 steps.

5 Complexity

- Complexity of (dynamic) mean shift: $O(dsn^2)$, where d is dimension, n is sample size, and s the average number of iterations.
- DMS converges faster, therefore the complexity is lower. Empirically, the time consumption is 40% ~70% lower.
- The "parallel"-like structure of DMS allows further speedups.

6 Experiments

- Joint spatial-range domain mean shift
- Image from Berkeley image segmentation data base

original / SMS result / DMS result

image	size	SMS time	SMS # iterations	DMS time	DMS # iterations
mountain	321×481	1.05	20.23	0.47	7.01
cottage	321×481	3.315	20.06	1.732	7.81
plane	321×481	0.23	20.12	0.13	8.58
squirrel	209×288	1.552	18.78	0.962	8.45
woman	321×481	3.535	26.39	1.832	9.54
house	192×255	2.483	23.33	1.061	7.61

7 Conclusions and Future Work

MDS can gradually shrink the sample set into compact clusters, and allows a fixed bandwidth procedure to achieve a variable bandwidth effect. It achieves faster convergence in both theory and practice.
In the future, we shall extend DMS to classification problems. Another direction is to study the convergence of DMS using level sets.

References

[1] D. Comaniciu and P. Meer, "Mean shift, A Robust Approach towards Feature Space Analysis," PAMI 2002.

[2] C. Yang, R. Duraiswami, D. DeMenthon and L. Davis, "Mean shift Analysis using Quasi-Newton Methods," ICIP 2003.

[3] C. Shen, M. J. Brooks and A. Hengel, "Fast Global Kernel Density Mode Seeking with Application to Localization and Tracking," ICCV 2005.