DOI:10.20033/j.1003-7241.(2025)06-0005-07

# 基于完整型带宽矩阵的均值偏移算法收敛性分析

朱晓光[1]，乔立永[2]，徐立新[3]，孟胜皓[1]

（1.火箭军士官学校，山东 潍坊 262500；
2.河北师范大学 河北省信息融合与智能控制重点实验室，河北 石家庄 050024；
3.北京理工大学 机电学院，北京 100081）

**摘 要**：均值偏移算法广泛应用于模式识别、图像处理和计算机视觉等领域。对采用具有一般形式的完整型带宽矩阵均值偏移算法进行推导，证明当且仅当带宽矩阵是对称正定矩阵时，均值偏移向量指向密度增加最大的方向。完整型带宽矩阵在数据空间的任意方向具有任意的平滑尺度。模拟实验结果表明，当数据密度在不同于坐标轴方向具有高曲率时，完整型带宽矩阵核密度估计相比于对角型带宽矩阵核密度估计可以获得显著的性能改进。针对目前均值偏移算法收敛性证明存在的问题，依据实数数列有界收敛定理和泛函分析赋范空间柯西数列收敛定理，分别对采用一般化形式完整型带宽矩阵核密度估计数列和位置向量数列的收敛性进行严格的数学证明，论据充分，判定方便，具有一般化普遍意义，为均值偏移算法的研究和应用提供理论基础。

**关键词**：均值偏移；核密度估计；带宽矩阵；收敛性；自适应
中图分类号：TP301.6；TP391　　文献标识码：A　　文章编号：1003−7241(2025)06−0005−07

# Convergence Analysis of Mean Shift Algorithm Based on Full Bandwidth Matrix

**ZHU Xiao-guang[1], QIAO Li-yong[2], XU Li-xin[3], MENG Sheng-hao[1]**

( 1. PLA Rocket Force Nco College, Weifang 262500, China;
2. Hebei Provincial Key Laboratory of Information Fusion and Intelligent Control, Hebei Normal University, Shijiazhuang 050024, China;
3. School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China )

**Abstract:** Mean shift algorithm is widely used is pattern recognition, image processing and computer vision etc. Mean shift algorithm with general full bandwidth matrix is deduced. That the mean shift vector aligns with the direction of density increasing maximally if and only if the bandwidth matrix is symmetric positive definite matrix is proved. Full bandwidth matrix has the flexibility to smooth by different amounts in any directions in the data space. Simulated experiments indicate that, for densities with high amounts of curvature in directions different to those of the coordinate axes, substantial gains can be made by the full bandwidth matrix compared to diagonal bandwidth matrix. Because of the problems in the convergence proof of mean shift algorithm, the convergence of the kernel density estimation series with general full bandwidth matrix and location vectors are proved rigorously in accordance with real sequences bounded convergence theorem and Cauchy's convergence theorem in normed space of functional analysis, the evidences are sufficient, the deductions are easy to understand, it is easy to judge, possessing general significance, and these results contribute substantial theoretical foundation for this algorithm's further research and applications.

**Keywords:** mean shift; kernel density estimation; bandwidth matrix; convergence; adaptive

## 0 引言

均值偏移算法是一种迭代非参数概率密度估计方法，用于复杂多模态特征空间分析和特征聚类识别，由Fukunaga等[1]于1975年首次提出。在被忽略20年后，Cheng对均值偏移算法的基本概念重新进行了定义[2]，提出一种更加一般化的表达形式，指出该算法可用于聚类分析和全局优化，为该算法的推广应用和深入研究奠定基础。Comaniciu等将均值偏移算法引入图像处理和计算机视觉任务，并做了大量开创性基础工作[3−6]。目前，均值偏移算法在图像处理、计算机视觉和模式识别等领域的应用主要包括：图像滤波[3]、图像分割[3]、目标跟踪[4]、目标检测[7]、语音检测[8]和聚类分析[9]等。

虽然均值偏移算法在工程项目中获得广泛应用，但是针对该算法本质特性的理论研究工作很少，本文主要针对采用一般化形式完整型带宽矩阵的均值偏移算法收敛性进行分析。

*基金项目：河北师范大学科研基金资助项目（L2021B31）
收稿日期：2024−03−07

Cheng在严格限定条件下，针对模糊过程均值偏移算法的收敛速率、收敛条件以及局部最大值点进行讨论，不具有普遍意义[2]。Comaniciu等[3]证明当核函数的剖面函数是凸的、单调减函数时，基于均值偏移算法得到的核密度估计数列收敛。针对Comaniciu等[3,5]证明位置向量$\{x_{(m)}\}_{m=1,2...}$收敛性存在的问题，李乡儒等[10]基于模式分类理论，判定当核密度估计数列$\{\hat{f}_{\mathbf{H}}(x_{(m)})\}_{m=1,2...}$的稳定点数量有限时，位置向量数列$\{x_{(m)}\}_{m=1,2...}$收敛，但是没有证明怎样的核函数形式能够使核密度估计数列在数据空间中的局部极值点数量有限，没有给出收敛的最终依据，在实际应用中，依据给定的数据集和核函数，难以直接判定是否满足收敛条件。另外，上述学者均是针对采用形式最简单的单参数型带宽矩阵均值偏移算法的收敛性进行证明，并且忽略了关于位置向量数列$\{x_{(m)}\}_{m=1,2...}$的收敛性证明应当依据泛函分析赋范空间数列收敛理论。如果针对采用一般化形式完整型带宽矩阵的均值偏移算法的收敛性进行证明分析，虽然面临许多新的问题，但是具有更加一般化的意义。

本文重新梳理了Comaniciu等[3]给出的核密度估计数列$\{\hat{f}_{\mathbf{H}}(x_{(m)})\}_{m=1,2...}$的收敛性证明和Guo等[11]给出的位置向量数列$\{x_{(m)}\}_{m=1,2...}$的收敛性证明，分别依据实数数列有界收敛定理和泛函分析赋范空间柯西数列收敛定理，对采用一般化形式完整型带宽矩阵的核密度估计数列$\{\hat{f}_{\mathbf{H}}(x_{(m)})\}_{m=1,2...}$和位置向量数列$\{x_{(m)}\}_{m=1,2...}$的收敛性进行严格的数学证明，论据充分，判定方便，具有一般化普遍意义，为均值偏移算法的研究和应用提供理论基础。

## 1 均值偏移算法

### 1.1 多变量核密度估计

核密度估计是目前最常用的非参数密度估计方法之一。假定$x_i, i=1,2,\cdots,n$是$d$维空间$R^d$中的$n$个数据点，即$x_i=(x_{i1},\cdots,x_{id})^{\mathrm{T}}$，点$x=(x_1,\cdots,x_d)^{\mathrm{T}}$处基于核函数$K_{\mathrm{H}}(x)$的加权多变量核密度估计$\hat{f}_{\mathbf{H}}(x)$的表达式为[12]

$$\hat{f}_{\mathbf{H},K}(x)=\frac{c_{k,d}}{n}\sum_{i=1}^{n}\omega_i\bullet K_{\mathbf{H}}(x-x_i) \tag{1}$$

式中，$K_{\mathrm{H}}(x)$为$d$维变量球状对称核函数，$K_{\mathrm{H}}(x)=|H|^{-1/2}K(H^{-1/2}x)$，在图像处理和计算机视觉应用中，通常$d=2$，并采用放射对称状多变量核函数，即$K(x)=c_{\mathrm{k},d}k(\|x\|^2)$，常量$c_{\mathrm{k},d}$满足$c_{\mathrm{k},d}^{-1}=\int k(\|x\|^2)\mathrm{d}x$，确保$K(x)$的积分为1。单变量核函数$k(x)$称为多变量核函数$K(x)$的剖面函数或轮廓函数，并满足以下性质[2]：

(1) 定义域$x\in[0,\infty)$，值域是实数集R；

(2) $k(x)$是非负的；

(3) $k(x)$是非增的，即如果$a<b$，则$k(a)\geqslant k(b)$；

(4) $k(x)$是分段连续的，且$\int_0^{\infty}k(x)\mathrm{d}x<\infty$。

$\boldsymbol{H}$是$d$维对称正定带宽矩阵，定义了用于估计概率密度的核函数的方向和尺度；权值$\omega_i$控制数据点$x_i$的重要性。基于核函数$K(x)$的加权核密度估计$\hat{f}_{\boldsymbol{H},K}(x)$的完整表达式如下：

$$\hat{f}_{\boldsymbol{H},K}(x)=\frac{c_{k,d}}{n}\sum_{i=1}^{n}\omega_i\bullet|\boldsymbol{H}|^{-1/2}\ k\left\{\left\|\boldsymbol{H}^{-1/2}(x-x_i)\right\|^2\right\} \tag{2}$$

### 1.2 均值偏移算法

核密度估计$\hat{f}_{\boldsymbol{H},K}(x)$表示数据集在$x$点的概率密度值，同时也可以认为是以$x$点作为核函数中心时，其他数据采样点与核函数中心属于同类区域的似然度之和。其他数据采样点与核函数中心的距离越小，它们与核函数中心属于同类区域的可能性越大[13−14]。采样数据集概率密度的局部最大值，即特征空间中的模态位于密度梯度为0的数据点处。与极大似然估计类似，通过求取核密度估计$\hat{f}_{\boldsymbol{H},K}(x)$关于位置向量$x$的偏导数，作为概率密度梯度估计，由此得到采样数据集局部最大分布密度的位置向量，此即为均值偏移算法获得特征空间模态分布位置的本质原理。

假定除有限个点外，核函数剖面函数$k(x)$的导数$k'(x)$在整个定义域$x\in[0,\infty)$上都存在。带宽矩阵$\boldsymbol{H}$是对称正定矩阵，$H^{-1}$亦是对称正定矩阵[15]，$\boldsymbol{H}^{-1/2}$同样是对称正定矩阵[16]，根据向量$l_2$范数的定义，得到：

$$\begin{aligned}\left\|\boldsymbol{H}^{-1/2}(x-x_i)\right\|^2&=\left[\boldsymbol{H}^{-1/2}(x-x_i)\right]^T\left[\boldsymbol{H}^{-1/2}(x-x_i)\right]\\&=(x-x_i)^T\boldsymbol{H}^{-1}(x-x_i)\end{aligned} \tag{3}$$

根据向量求导公式$\frac{\partial x^T\mathbf{A}x}{\partial x}=2\boldsymbol{A}x$(其中A是对称正定矩阵)[15]，得到：

$$\frac{\partial\left\{(x-x_i)^T\boldsymbol{H}^{-1}(x-x_i)\right\}}{\partial x}=2\boldsymbol{H}^{-1}(x-x_i) \tag{4}$$

由此可以得到核密度估计梯度的表达式

$$\begin{aligned}&\nabla\hat{f}_{H,K}(x)=\\&\frac{\partial\hat{f}_{H,K}(x)}{\partial x}=\\&\frac{2c_{k,d}}{n}\sum_{i=1}^{n}\omega_iH^{-1/2}k'\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}\boldsymbol{H}^{-1}(x-x_i)\end{aligned} \tag{5}$$

假定$g(x)=-k'(x)$，并将其作为核函数$G(x)$的剖面函数，则核函数$K(x)$称为核函数$G(x)$的阴影函数[2]。

$$G(x)=c_{g,d}g\left(\left\|x\right\|^2\right) \tag{6}$$

式中，$c_{g,d}$是核函数$G(x)$相应的归一化常量。将$g(x)=-k'(x)$代入式(5)中，依据$c_{k,d}$，$n$，$\omega_i$，$|\boldsymbol{H}|^{-1/2}$，$k'\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\}$均为标量，$\mathrm{H}^{-1}$是$d$维对称正定矩阵，并进行变换得到：

$$\begin{aligned}&\nabla\hat{f}_{\boldsymbol{H},K}(x)=\\&\frac{2c_{k,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}k'\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}\boldsymbol{H}^{-1}(x-x_i)=\\&\frac{2c_{k,d}|\boldsymbol{H}|^{-1/2}\boldsymbol{H}^{-1}}{n}\left[\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}\right]\cdot\\&\left[\frac{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}x_i}{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}}-x\right]\end{aligned} \tag{7}$$

式(7)等于右侧的第一项与基于核函数$G(x)$的密度估计$\hat{f}_{\boldsymbol{H},G}(x)$成比例，$\hat{f}_{\boldsymbol{H},G}(x)$表达式如下：

$$\hat{f}_{\mathrm{H},G}(x)=\frac{c_{g,d}}{n}\sum_{i=1}^{n}\omega_i\cdot|\boldsymbol{H}|^{-1/2}g\left\{\left\|\boldsymbol{H}^{-1/2}(x-x_i)\right\|^2\right\} \tag{8}$$

式(7)等于右侧的第二项即是Comaniciu等定义的均值偏移向量[17]，假定表示为$m_{h,G}(x)$：

$$m_{h,G}(x)=\frac{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}x_i}{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}}-x \tag{9}$$

从形式上看，均值偏移向量$m_{h,G}(x)$是采样数据点的加权平均与核函数中心$x$之间的差值，核函数$G(x)$的剖面函数$g(x)$是权值函数。将式(8)−式(9)代入式(7)得到：

$$m_{h,G}(x)=\frac{c_{g,d}\boldsymbol{H}}{2c_{k,d}}\cdot\frac{\nabla\hat{f}_{\boldsymbol{H},K}(x)}{\hat{f}_{\boldsymbol{H},G}(x)}=c\boldsymbol{H}\frac{\nabla\hat{f}_{\boldsymbol{H},K}(x)}{\hat{f}_{\boldsymbol{H},G}(x)} \tag{10}$$

式(10)表明，在位置$x$处，基于核函数$G(x)$计算得到的均值偏移向量与基于核函数$K(x)$计算得到规格化密度梯度估计成比例，规格化因子是在位置x处基于核函数$G(x)$计算得到的密度估计$\hat{f}_{\boldsymbol{H},G}(x)$。

当$d$=2时，$\boldsymbol{H}$是2×2对称正定矩阵，$x$和$x_i$均是2×1列向量，$\nabla\hat{f}_{\boldsymbol{H},K}(x)$和$m_{h,G}(x)$的计算结果是2×1列向量，$\hat{f}_{\boldsymbol{H},G}(x)$的计算结果是标量，根据向量夹角余弦公式，均值偏移向量$m_{h,G}(x)$与核密度估计梯度$\nabla\hat{f}_{\boldsymbol{H},K}(x)$之间的夹角余弦为

$$\begin{aligned}\cos\theta&=\frac{<m_{h,G}(x),\nabla\hat{f}_{\boldsymbol{H},K}(x)>}{\left\|m_{h,G}(x)\right\|\cdot\left\|\nabla\hat{f}_{\boldsymbol{H},K}(x)\right\|}=\frac{m_{h,G}(x)^T\nabla\hat{f}_{\boldsymbol{H},K}(x)}{\left\|m_{h,G}(x)\right\|\cdot\left\|\nabla\hat{f}_{\boldsymbol{H},K}(x)\right\|}\\&=\frac{\nabla\hat{f}_{\boldsymbol{H},K}(x)^T\boldsymbol{H}\nabla\hat{f}_{\boldsymbol{H},K}(x)}{\left\|\boldsymbol{H}\nabla\hat{f}_{\boldsymbol{H},K}(x)\right\|\cdot\left\|\nabla\hat{f}_{\boldsymbol{H},K}(x)\right\|}\end{aligned} \tag{11}$$

根据对称正定矩阵的定义[15]，当且仅当带宽矩阵$H$是对称正定矩阵时，均值偏移向量$m_{h,G}(x)$与核密度估计梯度$\nabla\hat{f}_{\boldsymbol{H},K}(x)$之间的夹角余弦大于0，此时均值偏移向量指向密度增加最大的方向。由于均值偏移向量与局部密度梯度估计的方向一致，因此确定一条通向估计密度静态点的路径[3]。

当核密度估计梯度$\nabla\hat{f}_{\boldsymbol{H},K}(x)$=0，即均值偏移向量$m_{h,G}(x)$=0时，得到位置向量$x$的计算公式：

$$x=\frac{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}x_i}{\sum_{i=1}^{n}\omega_i g\left\{(x-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x-x_i)\right\}} \tag{12}$$

由于位置向量$x$的计算公式是隐式表达式，采用数值计算中常用的非线性方程组不动点迭代法[18]，得到位置向量$x$的迭代求解公式：

$$x_{m+1}=\frac{\sum_{i=1}^{n}\omega_i g\left\{(x_m-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x_m-x_i)\right\}x_i}{\sum_{i=1}^{n}\omega_i g\left\{(x_m-x_i)^{\mathrm{T}}\boldsymbol{H}^{-1}(x_m-x_i)\right\}} \tag{13}$$

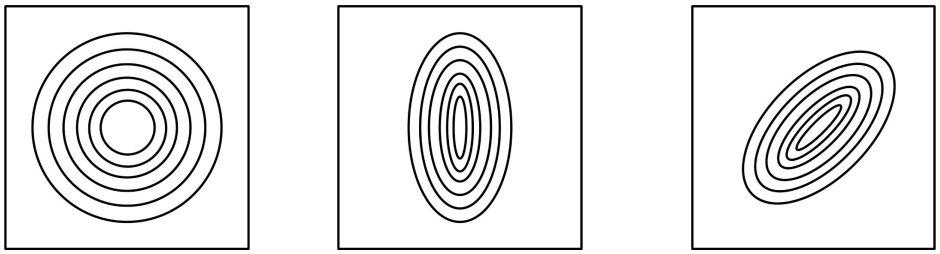
式中，$m$是当前迭代次数。位置向量$x$的迭代更新公式可以看作是自适应加权采样均值，权值依赖于数据点$x_i$和当前数据中心$x_{\mathrm{m}}$之间的Mahalanobis距离。

## 2 核函数带宽矩阵形式的选择

带宽矩阵控制着核平滑的尺度和方向，是影响核密度估计精密程度最重要的因素[19]。带宽矩阵的形式分为3种：单参数型$H_{\mathrm{S}}$、对角型$H_{\mathrm{D}}$和完整型$H_{\mathrm{F}}$。单参数带宽矩阵$\boldsymbol{H}_{\mathrm{S}}=\{\boldsymbol{H}=h^2\boldsymbol{I},h>0\}$，仅需确定1个平滑参数，缺点是在每个坐标轴方向具有相同的平滑尺度；对角型带宽矩阵，$\boldsymbol{H}_{\mathrm{D}}=\{\boldsymbol{H}=\mathrm{diag}[h_1^2,\cdots,h_d^2]\}$，$h_1^2,\cdots,h_d^2>0$，需要确定$d$个平滑参数，但是具有更多灵活性，在$d$维空间的每个坐标轴方向具有不同的平滑尺度；完整型带宽矩阵$H_{\mathrm{F}}$是$d\times d$维对称正定矩阵，需要确定$d(d+1)/2$个平滑参数，相比于前两种形式带宽矩阵，具有一般化形式，计算较复杂，但是具有更好的适应性，在$d$维空间的每个坐标轴方向以及不同于坐标轴方向的任意方向都具有任意的平滑尺度，可以获得更高的核密度估计精度[12]。通过3种形式带宽矩阵的定义可以看出，单参数型$H_{\mathrm{S}}$、对角型$H_{\mathrm{D}}$和完整型$H_{\mathrm{F}}$带宽矩阵集具有以下隶属关系：$H_{\mathrm{S}}\subseteq H_{\mathrm{D}}\subseteq H_{\mathrm{F}}$。

如果限定带宽矩阵的形式是单参数型或者对角型，会导致多变量核密度估计的性能明显下降。下面以二维空间为例对比3种形式带宽矩阵对核密度估计的影响。

图1所示为高斯核函数分别采用3种不同形式带宽矩阵的轮廓示意图[12]。在3种情况下，核函数轮廓都是椭圆形；然而，当带宽矩阵是单参数型$H_{\mathrm{S}}$时，这些椭圆被限定为圆形；对于对角型带宽矩阵$H_{\mathrm{D}}$，椭圆的轴对应于坐标方向；只有在完整型带宽矩阵$H_{\mathrm{F}}$的情况下，才能获得



(a) 单参数型$H_S$ (b) 对角型$H_D$ (c) 完整型$H_F$

图1 基于3种不同形式带宽矩阵高斯核函数的轮廓示意图

具有任意方向的椭圆。Wand等[17]针对不同形式带宽矩阵导致双变量核密度估计的性能差异进行研究,结果表明在每个坐标轴方向具有独立平滑参数的密度估计效果较好,对于密度估计的许多情况已经足够,然而,当数据密度在不同于坐标轴方向具有高曲率时,采用完整型带宽矩阵核密度估计可以获得显著的性能改进。

下面采用模拟“哑铃”状密度数据对比分别采用完整型和对角型带宽矩阵时多变量核密度估计的性能差异[20−21]。采用高斯混合函数模拟“哑铃”状密度函数,表达式如下:

$$f=\frac{4}{11}N(-2,2;1,1,0)+\frac{3}{11}N(0,0;0.8,0.8,-0.72)+\frac{4}{11}N(2,-2;1,1,0) \tag{14}$$

根据式(14)绘制的模拟“哑铃”状密度如图2所示,“哑铃”状密度具有以下特点:唯一的模态位于“哑铃”握杆处,“哑铃”两端的数据密度较小。

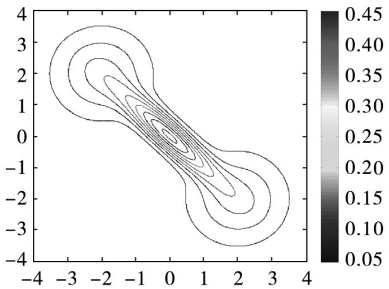
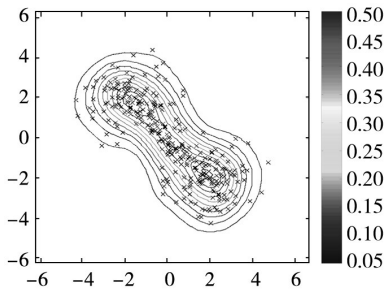
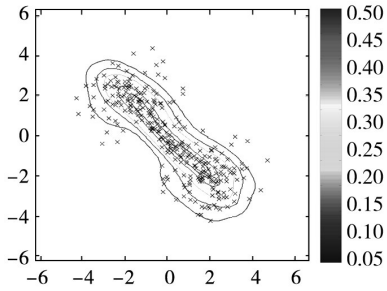


图2 “哑铃”状密度轮廓

依据模拟“哑铃”状密度函数产生300个随机采样数据,分别利用渐进平均积分平方误差法获得对角型带宽矩阵[0.447 7 0;0 0.561 2],利用积分平方偏差和渐进积分方差法获得完整型带宽矩阵[0.564 6 −0.404 3;−0.404 3 0.493 4],并采用高斯核函数估计采样数据密度,结果分别如图3(a)−图3(b)所示。计算得到的对角型和完整型带宽矩阵的对角元素大约都在0.45~0.56之间,因此两者在坐标轴方向的平滑度相当。两者的主要差异在于完整型带宽矩阵远离对角的元素几乎与对角元素的幅值相等,因此在倾斜方向引起大幅平滑。分析图3(a),3(b)分别采用对角型和完整型带宽矩阵高斯核函数密度估计的结果,可以看出,对角型带宽矩阵高斯核函数能够在平行于坐标轴方向产生平滑,但是在中间倾斜部分不能产生适度平滑,导致估计的“哑铃”密度在握杆处过于宽阔平坦,产生双模态;完整型带宽矩阵高斯核函数估计的“哑铃”密度能够适度地平滑中央、有角度的区域,因此能够再现目标密度的单模态。



(a) 对角型带宽矩阵密度估计



(b) 完整型带宽矩阵密度估计

图3 采用不同形式带宽矩阵高斯核函数密度估计结果

## 3 收敛性分析

本节重新梳理Comaniciu等[3]给出的核密度估计数列$\left\{\hat{f}_{\mathbf{H}}(x_{(m)})\right\}_{m=1,2\cdots}$收敛性证明和Guo等[11]给出的位置向量数列$\left\{x_{(m)}\right\}_{m=1,2\cdots}$收敛性证明,分别依据实数数列有界收敛定理和泛函分析赋范空间柯西数列收敛定理,对采用一般化形式完整型带宽矩阵的核密度估计数列$\left\{\hat{f}_{\mathbf{H}}(x_{(m)})\right\}_{m=1,2\cdots}$和位置向量数列$\left\{x_{(m)}\right\}_{m=1,2\cdots}$的收敛性进行严格的数学证明。

**定理1**.如果核函数的剖面函数是凸的、单调减函数,那么通过均值偏移算法得到的密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$收敛,且密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$单调递增。

证明:由于样本空间数据点集$\{x_m\}_{m=1,2\cdots,n}$的数量有限,那么密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$有界,根据实数数列收敛定理[22],只要证明$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$严格单调递增,就可以判断密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$收敛。

根据式(2),加权核密度估计$\hat{f}_{\boldsymbol{H}}(x)$的计算公式为

$$\hat{f}_{\boldsymbol{H}}(x)=\frac{c_{k,d}}{n}\sum_{i=1}^{n}\omega_i\cdot|\boldsymbol{H}|^{-1/2}\,k\left\{\left\|\boldsymbol{H}^{-1/2}(x-x_i)\right\|^2\right\} \tag{15}$$

因此可以得到:

$$\begin{aligned}&\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)=\\&\frac{c_{k,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}\left[k\left(\left\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_i)\right\|^2\right)-\right.\\&\left.k\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\right]\end{aligned} \tag{16}$$

由于核函数的剖面函数$k(x)$是凸函数,根据凸函数的基本性质[21],对于$\forall x_1,x_2\in[0,\infty)$,$x_1\neq x_2$,存在:

$$k(x_2)\geq k(x_1)+k'(x_1)(x_2-x_1) \tag{17}$$

假设$k'(x)=-g(x)$,则:

$$k(x_2)-k(x_1)\geq g(x_1)(x_1-x_2) \tag{18}$$

将式(18)代入式(16)中得到:

$$\begin{aligned}&\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)\geqslant\\&\frac{c_{k,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}\,g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\cdot\\&\left[\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2-\left\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_i)\right\|^2\right]\end{aligned} \tag{19}$$

由于$\boldsymbol{H}^{-1/2}$是$2\times2$对称正定矩阵,$x$和$x_i$均是$2\times1$列向量,$g(x)$的计算结果是标量,并且根据向量范数计算公式[15]:$\|x-y\|^2=\|x\|^2-2\langle x,y\rangle+\|y\|^2$,公式(19)变形为:

$$\begin{aligned}&\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)\\&\geq\frac{c_{f,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}\,g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\cdot\\&\left[\left\|\boldsymbol{H}^{-1/2}x_m\right\|^2-2x_m{}^T\boldsymbol{H}^{-1}x_i-\left\|\boldsymbol{H}^{-1/2}x_{m+1}\right\|^2+2x_{m+1}{}^T\boldsymbol{H}^{-1}x_i\right]\end{aligned} \tag{20}$$

根据式(13),位置向量$\left\{x_{(m)}\right\}_{m=1,2\cdots}$的迭代更新公式为q

$$x_{m+1}=\frac{\sum_{i=1}^{n}\omega_i g\left\{\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right\}x_i}{\sum_{i=1}^{n}\omega_i g\left\{\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right\}} \tag{21}$$

将其代入式(20)中,得到:

$$\begin{aligned}&\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)\geq\\&\frac{c_{f,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}\,g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\left\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\right\|^2\end{aligned} \tag{22}$$

由于剖面函数$k(x)$是单调减函数,对于所有$x\geqslant0$,$k'(x)<0$,$g(x)=-k'(x)>0$,因此$\sum_{i=1}^{n}g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)>0$,由此可以得到$\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)\geqslant0$,即$\hat{f}_{\boldsymbol{H}}(x_{m+1})\geqslant\hat{f}_{\boldsymbol{H}}(x_m)$,当且仅当$x_m=x_{m+1}$时,$\hat{f}_{\boldsymbol{H}}(x_{m+1})=\hat{f}_{\boldsymbol{H}}(x_m)$。又$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$有界,根据实数数列收敛定理[20]:单调有界数列必然收敛,因此核密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$收敛。

**定理2**.根据泛函分析赋范空间柯西数列收敛理论,如果核密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$收敛,则可以判断位置向量数列$\left\{x_{(m)}\right\}_{m=1,2\cdots}$收敛,并且是$l_2$范数定义的赋范空间中的柯西数列。

证明:由于收敛数列都是柯西数列[20],因此核密度估计数列$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=1,2\cdots}$也是一个柯西数列,对于任何正无穷小数$\varepsilon_m$,存在正整数$N$,对于$m>N$的一切$\left\{\hat{f}_{\boldsymbol{H}}(x_{(m)})\right\}_{m=N+1,N+2,\cdots}$,满足不等式$0\leqslant\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)\leqslant\varepsilon_m$。

根据:$\hat{f}_{\boldsymbol{H}}(x_{m+1})-\hat{f}_{\boldsymbol{H}}(x_m)$

$$\geqslant\frac{c_{f,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\left\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\right\|^2$$

得到:

$$\begin{aligned}&\frac{c_{f,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\left\|\boldsymbol{H}^{-1/2}(x_{m+1}-\right.\\&\left.x_m)\right\|^2<\varepsilon_m\end{aligned} \tag{23}$$

假定$\varepsilon_m\Big/\left(\frac{c_{f,d}}{n}\sum_{i=1}^{n}\omega_i|\boldsymbol{H}|^{-1/2}\,g\left(\left\|\boldsymbol{H}^{-1/2}(x_m-x_i)\right\|^2\right)\right)=\eta_m$。由于常数与无穷小的数积是无穷小[23],因此$\eta_m$依然是正无穷小,则$0<\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\|^2<\eta_m$,根据不等式开方公式[24],得到:$0<\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\|^2<\sqrt{\eta_m}$。正无穷小$\eta_m$的平方根$\sqrt{\eta_m}$依然是正无穷小。依据上述分析,对于$\{\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\|\}_{N<m<M}$,分别存在以下不等式:$\{0<\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\|<\lambda_m\}_{N<m<M}$,其中$\lambda_m=\sqrt{\eta_m}$,$\{\lambda_m\}_{N<m<M}$均为正无穷小。根据不等式求和公式[23],得到:$0<\sum_{m=N+1}^{M-1}\|\boldsymbol{H}^{-1/2}(x_{m+1}-x_m)\|<\lambda_\Sigma$,其中$\lambda_\Sigma=\sum_{m=N+1}^{M-1}\lambda_m$,由于有限个无穷小的和也是无穷小[24],

$\lambda_{\Sigma}$依然是正无穷小。由于$\sum_{m=N+1}^{M-1}\left\|\boldsymbol{H}^{-1/2}(x_{(m+1)}-x_{(m)})\right\|$:$\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-1)})\right\|+\sum_{=N+1}^{f-2}\left\|\boldsymbol{H}^{-1/2}(x_{(m+1)}-x_{(m)})\right\|$,根据向量范数的三角不等式[20]:$\|z-y\|\geqslant\|z\|-\|y\|$,可以得到:

$$\begin{aligned}&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-1)})\right\|\\&=\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-2)})-\boldsymbol{H}^{-1/2}(x_{(M-1)}-x_{(M-2)})\right\|\\&\geq\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-2)})\right\|-\left\|\boldsymbol{H}^{-1/2}(x_{(M-1)}-x_{(M-2)})\right\|\end{aligned}\tag{24}$$

同理可以得到:

$$\begin{aligned}&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-2)})\right\|=\\&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-3)})-\boldsymbol{H}^{-1/2}(x_{(M-2)}-x_{(M-3)})\right\|\geqslant\\&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-3)})\right\|-\left\|\boldsymbol{H}^{-1/2}(x_{(M-2)}-x_{(M-3)})\right\|\end{aligned}\tag{25}$$

依次类推,可以得到:

$$\begin{aligned}&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(M-1)})\right\|\\&\geqslant\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(N+1)})\right\|-\left\|\boldsymbol{H}^{-1/2}(x_{(M-1)}-x_{(M-2)})\right\|-\\&-\left\|\boldsymbol{H}^{-1/2}(x_{(N+2)}-x_{(N+1)})\right\|=\\&\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(N+1)})\right\|-\sum_{m=N+1}^{M-2}\left\|\boldsymbol{H}^{-1/2}(x_{(m+1)}-x_{(m)})\right\|\end{aligned}\tag{26}$$

将式(26)代入$\sum_{m=N+1}^{M-1}\|\mathrm{H}^{-1/2}(x_{(\mathrm{m}+1)}-x_{(\mathrm{m})})\|\}$中得到:

$$\sum_{m=N+1}^{M-1}\left\|\boldsymbol{H}^{-1/2}(x_{(m+1)}-x_{(m)})\right\|\geq\left\|\boldsymbol{H}^{-1/2}(x_{(M)}-x_{(N+1)})\right\|\tag{27}$$

根据前述推导:$0<\sum_{m=N+1}^{M-1}\|\boldsymbol{H}^{-1/2}(x_{\mathrm{m}+1}-x_{\mathrm{m}})\|\}<\lambda_{\Sigma}$,其中$\lambda_{\Sigma}$是任意给定的无穷小正数。因此,存在正整数$N$和$M$,对于$M>m>N$的一切$\|\mathrm{H}^{-1/2}(x_{(\mathrm{M})}-x_{(\mathrm{N}+1)})\|$,存在不等式$0<\|\boldsymbol{H}^{-1/2}(x_{(\mathrm{M})}-x_{(\mathrm{N}+1)})\|<\lambda_{\Sigma}$。由于向量$\mathrm{H}^{-1/2}x_{(\mathrm{M})}$和$\mathrm{H}^{-1/2}x_{(\mathrm{N}+1)}$之间的距离依据$l_2$范数定义,根据赋范空间柯西数列定义[24],向量数列$\left\{\boldsymbol{H}^{-1/2}x_{(m)}\right\}_{m=1,2\ldots}$是柯西序列。根据$l_2$范数定义的赋范空间是Banach空间,柯西数列在Banach空间中都收敛[24],因此向量数列$\left\{\boldsymbol{H}^{-1/2}x_{(m)}\right\}_{m=1,2\ldots}$收敛,可以假定$\lim_{m\to\infty}\boldsymbol{H}^{-1/2}x_{(\mathrm{m})}=\boldsymbol{H}^{-1/2}x$。根据前文所述,$\boldsymbol{H}^{-1/2}$是对称正定矩阵,因此是可逆矩阵,得到$\lim_{m\to\infty}x_{(\mathrm{m})}=x$,即位置向量$\left\{x_{(m)}\right\}_{m=1,2\ldots}$收敛。同理,根据赋范空间柯西数列的性质[20],位置向量数列$\left\{x_{(m)}\right\}_{m=1,2\ldots}$是由$l_2$范数定义的赋范空间中的柯西数列。

## 4 结束语

均值偏移算法是一种在模式识别、图像处理和计算机视觉等领域广泛应用的聚类算法。本文对采用具有一般形式的完整型带宽矩阵均值偏移算法进行推导,证明当且仅当带宽矩阵是对称正定矩阵时,均值偏移向量指向密度增加最大的方向。带宽矩阵控制着核平滑的尺度和方向,是最重要的影响核密度估计精度的因素。完整型带宽矩阵具有最大的适应性,在数据空间的任意方向具有任意的平滑尺度。模拟实验表明,当数据密度在不同于坐标轴方向具有高曲率时,完整型带宽矩阵核密度估计相比于对角型带宽矩阵核密度估计可以获得显著的性能改进。针对目前均值偏移算法收敛性证明存在的问题,依据实数数列有界收敛定理和泛函分析赋范空间柯西数列收敛定理,分别对采用一般化形式完整型带宽矩阵的核密度估计数列和位置向量数列的收敛性进行严格的数学证明,论据充分,判定方便,具有一般化普遍意义,为均值偏移算法的研究和应用提供理论基础。

**参考文献**:

[1] Fukunaga K U A,Hostetler L D.Hostetler,L.D.The estimation of the gradient of a density function,with applications in pattern recognition.IEEE Transactions on Information Theory IT-21,32-40[J].IEEE Transactions on Information Theory,1975,21(1):32-40.

[2] Cheng Y.Mean shift,mode seeking,and clustering[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,1995,17(8):790-799.

[3] Comaniciu D,Meer P.Mean shift:a robust approach toward feature space analysis[J].IEEE Trans Pattern Analysis &Machine Intelligence,2002,24(5):603-619.

[4] Comaniciu D,Ramesh V,Meer P.Kernel-Based Object Tracking[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2003(25):564-577.

[5] Comaniciu D,Ramesh V,Meer P.(2000).IEEE Comput.Soc IEEE Conferen ce on Computer Vision and Pattern Recognition.CVPR 2000-Hilton Head Island,SC,USA(13-15 June 2000)-Real-time tracking of non-rigid objects using mean shift.,2(0),142-149.doi:10.1109/cvpr.2000.854761

[6] Comaniciu D,Meer P.Robust analysis of feature spaces:Color image segmentation[C]//Computer Vision and Pattern Recognition,1997.Proceedings.1997 IEEE Computer Society Conference on.IEEE,1997.

[7] Chen,Xiaopeng et al.Adaptive Bandwidth Mean Shift Object Detection[C]//2008 IEEE Conference on Robotics,Automation and Mechatronics,2008:210-215.

[8] Yang J C,He Q H,Li Y X,et al.Speaker Change Detection based on Mean Shift[J].Journal of Computers,2013,8(3):638-644.

(下转第50页)