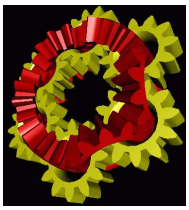
Forschungsschwerpunkt S92

Industrial Geometry

http://www.ig.jku.at



FSP Report No. 36

Bivariate density estimation using BV regularisation

Andreas Obereder, Otmar Scherzer, Arne Kovac

December 2006

FWF
Der Wissenschaftsfonds.

universität innsbruck

Bivariate density estimation using BV regularisation

Andreas Obereder [a], Otmar Scherzer [a], Arne Kovac [b,*]

[a] *Institute of Computer Science, University of Innsbruck, Austria*

[b] *Department of Mathematics, University of Bristol, Bristol, UK*

Abstract

In this paper we study the problem of bivariate density estimation. The aim is to find a density function with the smallest number of local extreme values which is adequate with the given data. Adequacy is defined via Kuiper metrics. The concept of the taut-string algorithm which provides adequate data with a small number of local extrema is generalised for analysing high dimensional data, thereby using Delaunay triangulation and diffusion filtering. Our results are based on equivalence relations in space dimension one of the taut string algorithm with the total variation minimisation and the method of solving the discrete total variation flow equation. The generalisation and some modifications (for instance based on the Fisher information contents) are developed and the performance for density estimation is shown.

Key words: Density estimation, modality, regularisation.

1 Introduction

In this paper we consider the problem of density estimation. Given a sample $x_1, \ldots, x_n \in \mathbb{R}^m$ the task is to specify a simple density function u and hence a distribution function U such that the data look like a typical sample from U.

Figure 1 shows observations from the *Old Faithful Geyser* in the Yellowstone National Park. Each observation consists of two measurements: the duration of the eruption and waiting time to the next eruption (both in minutes), which

* Corresponding author. Tel.: +44 117 3317450; fax: +44 118 9287999.
Email addresses: andreas.obereder@uibk.ac.at (Andreas Obereder), otmar.scherzer@uibk.ac.at (Otmar Scherzer), a.kovac@bristol.ac.uk (Arne Kovac).

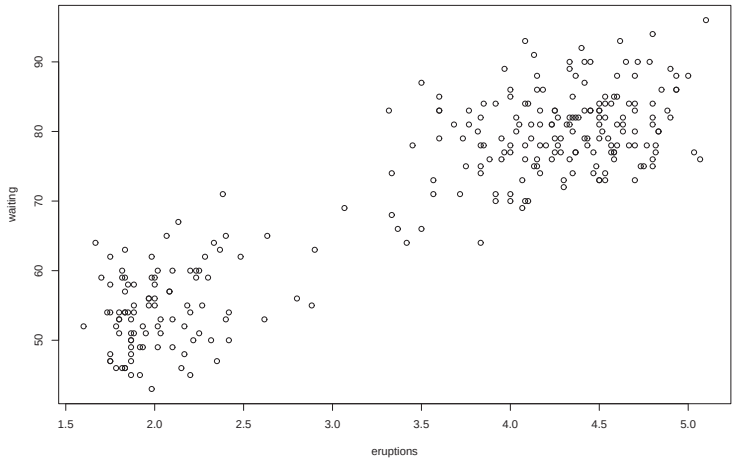


Fig. 1. 'Old Faithful' geyser data.

are plotted against each other. The data indicate a bimodal distribution which is rare for the behaviour of a geyser and so physicists and geologists were interested in investigating possible reasons for this distribution (Azzalini and Bowman, 1990).

Existing literature on density estimation is vast. In particular the one-dimensional situation has been studied extensively and numerous methods have been proposed. Of the different approaches the most popular one is kernel estimation. We refer to Nadaraya (1964), Watson (1964), Silverman (1986), Sheather and Jones (1991), Wand and Jones (1995), Sain and Scott (1996) and Simonoff (1996) and the references given there. Another approach is based on wavelets. We refer to Donoho, Johnstone, Kerkyacharian and Picard (1996), Herrick, Nason and Silverman (2000) and to Chapter 7 of Vidakovic (1999). Mixtures of densities have been considered in the Bayesian framework by Richardson and Green (1997) and Roeder and Wasserman (1997). Other Bayesian methods are to be found in Verdinelli and Wasserman (1998). Multivariate density estimation has been analysed mainly in the context of kernel estimators (Scott and Sain, 2005).

Most of the developed methodology focuses on smoothness of the regression function rather than on simplicity like a small number of local extreme values. In spite of this methods are often judged by their ability to identify peaks in the data as in Loader (1999) and Herrick et al (2000). Work directly concerned with modality has been done by Davies and Kovac (2001, 2004) who use the taut-string method to provide approximations with asymptotically consistent modality. Further work is provided by Müller and Sawitzki (1991) using their concept of excess mass. Their ideas have been extended to multidimensional distributions by Polonik (1995a, 1995b, 1999). Hengartner and Stark (1995) use the Kolmogoroff ball centred at the empirical distribution function to ob-

tain nonparametric confidence bounds for shape restricted densities. Another way of controlling modality is that of mode testing. We refer to Good and Gaskins (1980), Silverman (1986), Hartigan and Hartigan (1985) and Fisher, Mammen and Marron (1994).

In this paper we study the problem of density estimation on a high dimensional domain Ω. Our approach relies on a suitable definition of adequacy (Davies, 1995). A measure of adequacy gives rise to a set of adequate functions, each of them representing a plausible model for the data in the sense that the data look like a "typical" sample from the model. The measure we employ in this paper is based on projections of the density function in the directions of its coordinates and evaluating the distance to the data with the Kuiper metric.

Having specified the set of adequacy we look for an adequate function u which is as simple as possible and in particular has the *smallest* possible number of local extreme values. In particular we construct a scale of candidate functions $u^{(1)}, u^{(2)}, u^{(3)}, \ldots$ with decreasing complexity, in particular decreasing modality, and choose the smoothest function that is still adequate as an approximation to the data.

The candidate functions are generated by generalisations of the *taut-string* algorithm for analysing data defined on a high dimensional domain. For the purpose of density estimation of high dimensional data we propose a two step algorithm which consists in generating an auxiliary function y from the sampled data x_i to which a filtering technique is applied. For the purpose of data filtering we discuss total variation regularisation, the total variation flow, and some variants (for instance based on the Fisher information), since these three techniques can be considered possible generalisations of the taut-string algorithm. Generalisation to contact problems derived from minimal surface minimisation as in Mammen and van de Geer (1997) have been generalised in Hinterberger et al. (2003) and Scherzer (2005) but are not discussed further in this paper.

The outline of this paper is as follows:

In Section 2 we discuss some possible generalisation of the taut-string algorithm for analysing high dimensional data. In Section 3 an initial solution for the density estimation process is constructed. Moreover, a grid, which can be used in numerical reconstructions is provided. Section 4 discusses different diffusion filtering methods. Section 5 is concerned with the definition of adequacy for bivariate density functions. Finally the results obtained with the different methods are shown and the results for density estimation are compared.

2 Taut strings in one and higher dimensions

One dimensional density functions approximating scattered point data can be calculated with the taut string algorithm (Davies and Kovac, 2001, 2004). It is the goal of this section to discuss generalisations of the taut string algorithm for density estimation on high-dimensional data.

It has been shown by Mammen and van de Geer (1997) that the solution of the taut string algorithm for given sampling data $\mathbf{y} = (y_1, \ldots, y_{n-1})$ sampled at the midpoints of uniformly distributed grid points $x_1 = 0, x_2 = h, x_3 = 2h, \ldots, x_n = (n-1)h$ in $\Omega = [0,1]$ with sampling distance $h = 1/(n-1)$ is equivalent to minimising

$$f_d(\mathbf{u}) := \frac{1}{2}\sum_{i=1}^{n-1} h\,|u_i - y_i|^2 + \alpha \sum_{i=1}^{n-2} h \frac{|u_{i+1} - u_i|}{h} \ .$$

where the u_i are associated with a function u s.t. $u(x) := u_i$ if $x \in (x_i, x_{i+1})$. The equivalence relation of the taut string algorithm and total variation flow regularisation shows that $\mathbf{u}$ is determined by $u_i = v_{i+1} - v_i$, $i = 1, \ldots, n-1$ where $\mathbf{v}$ minimises the constraint optimisation problem

$$\sum_{i=1}^{n-1} \sqrt{1 + \frac{|v_{i+1} - v_i|^2}{h^2}} \text{ subject to } \left| v_j - h\sum_{k=1}^{j} y_k \right| \leq \alpha \quad j = 1, \ldots, n \ . \tag{1}$$

and $v_1 = 0$. The v_i are associated with a function v which is linear in $[x_i, x_{i+1}]$ and $v(x_i) = v_i$.

In Steidl et al. (2004) it has been shown that the minimiser of f_d and the solution of the *space discrete total variation flow* equation at time $t = \alpha$ which solves

$$\begin{aligned} \dot{u}_1 &\in sgn(u_2 - u_1), \\ \dot{u}_i &\in sgn(u_{i+1} - u_i) - sgn(u_i - u_{i-1}) \qquad (i = 2, \ldots, n-2), \\ \dot{u}_{n-1} &\in -sgn(u_{n-1} - u_{n-2}), \\ u(0) &= y. \end{aligned} \tag{2}$$

are identical.

These considerations reveal that for analysing sampling data with sampling points in $(0,1)$ there are at least three equivalent concepts, contact problems as formulated in Mammen and van de Geer (1997), discrete total variation regularisation (i.e. minimisation of f_d), and the discrete total variation flow (2). We show below that for continuous data the adequate continuous formulations of the three concepts can be generalised to high dimensional data. We focus on

generalisations of total variation regularisation and the discrete total variation flow.

It has been shown by Grasmair (2006) (see also Pöschl and Scherzer, 2006) that for $y \in L^2(0,1)$, the minimiser of *continuous total variation minimisation*, consisting in minimisation of

$$f_c(u) := \frac{1}{2}\int_0^1 (u-y)^2 + \alpha \int_0^1 |u'| \ ,$$

is locally constant or satisfies $u = y$. If $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ is an image or a voxel data $u : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$, then *continuous total variation minimisation* consists in minimisation of

$$\mathcal{F}_c(u) := \frac{1}{2}\int_\Omega (u-y)^2 + \alpha \, |Du| \, ,$$

where $|Du|$ is the total variation of u (see e.g. Evans & Gariepy, 1992). In the image processing community $\mathcal{F}_c$ is called the R(udin)-O(sher)-F(atemi)-functional (Rudin, Osher, Fatemi, 1992).

With a similar argumentation it becomes evident that the total variation inclusion equation

$$\frac{\partial u}{\partial t} \in -\partial \, |Du| \ ,$$

where $\partial \, |Du|$ denotes the subgradient of the total variation $|Du|$ of u, is the continuous formulation of the discrete total variation flow equation. For more background on inclusion equations we refer to Brezis (1973). It is convenient and instructive, but not mathematically rigorous, to write

$$\partial \, |Du| = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) \ . \tag{3}$$

Total variation regularisation and total variation flow can be used if appropriate initial data y has been determined from discrete sampling data. A method to construct the data y which is compatible with the taut-string algorithm is presented in the next section.

3 Initialisation of irregularly sampled data

To calculate a density function on a one-dimensional domain given a sample $\mathbf{x}$ we first define a piecewise constant function y by setting $h_{(i)} := x_{(i+1)} - x_{(i)}$ and $y(x) := 1/((n-1)h_i)$ for all $x \in [x_{(i)}, x_{(i+1)})$ and for $i = 1, \ldots, n-1$. Here $x_{(i)}$ denotes the ordered samples with $x_{(j)} \leq x_{(j+1)}$. Then the taut string algorithm

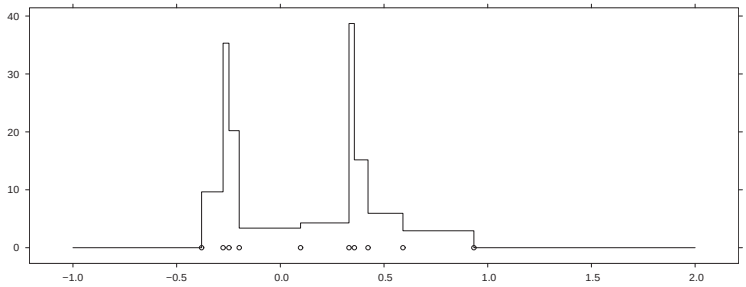


Fig. 2. Top: Randomly distributed data points with piecewise constant initial guess y.

can be used and equivalently be formulated as the problem of minimisation of

$$\begin{aligned} f_e(\mathbf{u}) &:= \frac{1}{2}\sum_{i=1}^{n-1} h_i(u_i - y_i)^2 + \alpha \sum_{i=1}^{n-2} h_i \frac{|u_{i+1} - u_i|}{h_i} \\ &:= \frac{1}{2}\sum_{i=1}^{n-1} h_i(u_i - y_i)^2 + \alpha \sum_{i=1}^{n-2} h_i \left|\nabla_{h_i} u\right|(x_i) \ . \end{aligned}$$

We interpret f_e as a quadrature rule of $\mathcal{F}_c(u)$ with sampling distances h_i and partitioning sampling intervals $I_i := \left[x_{(i)}, x_{(i+1)}\right] \ , \quad i = 1, \ldots, n-1$.

Here $\left|\nabla_{h_i} u\right|(x_{(i)}) = \frac{|u_{i+1} - u_i|}{h_i}$ is the absolute value of the right difference quotient of a function u at x_i with step size h_i. The method of minimising f_e is called *discrete total variation minimisation with irregular samples.*

A typical example of a function y is plotted in Figure 2.

We associate a grid with the nodes (i.e. data points) x_i and the corresponding elements I_i. For high dimensional domains, Delaunay's triangulation can be used to determine partitioning tetrahedrons, thus generalising the concept of sampling intervals I_i in space dimension one. Delaunay's triangulation and the associated Voronoi diagram are well known concepts from computational geometry and used in many applications (Aurenhammer and Klein, 2000). An excellent introduction to this topic is Edelsbrunner (2001). An example of a Delaunay triangulation is shown in Figure 3. The data used for the triangulation were 500 points randomly generated from the distribution shown in Figure Figure 4.

As in the the one-dimensional case we define the value of the initial solution for every grid-element I_i as $1/(marea(I_i)), i = 1, \ldots, n-1$ where m is the number of triangles. Figure 4 shows an initial solution obtained by setting a constant value over each grid cell.

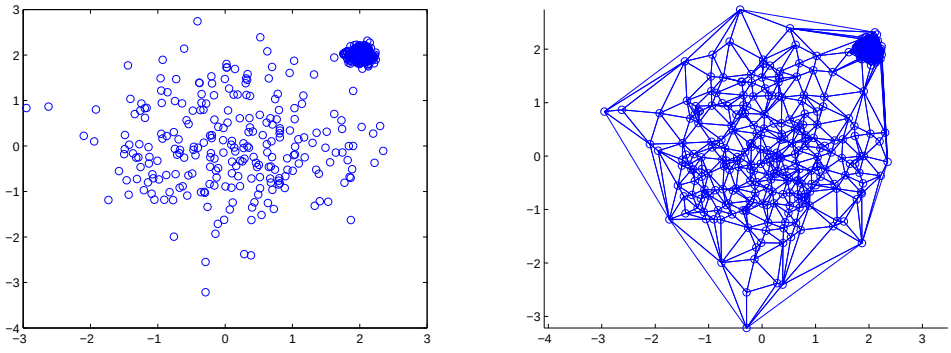


Fig. 3. Scattered data and the Delaunay triangulation

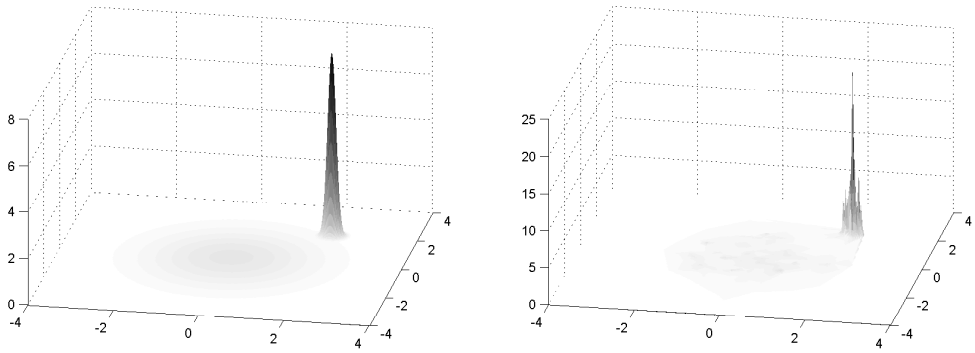


Fig. 4. Underlying distribution and Initial Guess

4 Diffusion filtering after Delaunay's triangulation

For filtering data y derived after Delaunay's triangulation from a discrete sample we use differential equations of the form

$$\begin{aligned} \frac{\partial u}{\partial t} &= \nabla \cdot (d(u, \nabla u)\nabla u) + e(u, \nabla u) \text{ in } \Omega \\ \frac{\partial u}{\partial \mathbf{n}} &= 0 \text{ on } \partial\Omega\,, \\ u(0) &= y\,. \end{aligned} \tag{4}$$

where $d(\cdot,\cdot)$ and $e(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ are appropriate functions.

Particular examples considered in this paper are

- the *total variation flow equation* (3), where $d(u, \nabla u) = \frac{1}{|\nabla u|}$ and $e(u, \nabla u) = 0$,

- the *Fisher information flow equation* with $d(u, \nabla u) = \frac{1}{|u|}$ and $e(u, \nabla u) = \frac{u|\nabla u|^2}{2|u|^3}$, and
- the $\frac{3}{2}$ *Laplacian flow* with $d(u, \nabla u) = \frac{1}{\sqrt{|\nabla u|}}$ and $e(u, \nabla u) = 0$.

The solution of (4) at time T is an approximation of the density to be estimated. We note that the partial differential equation (4) is only formally stated and in general has to be considered an inclusion equation.

4.1 The total variation flow equation

For the solution of (3) we use a standard semi-implicit scheme

$$u^{(n)} - u^{(n-1)} \in \Delta t \nabla \cdot \left(\frac{\nabla u^{(n)}}{|\nabla u^{(n-1)}|} \right) \text{ in } \Omega \text{ and } \frac{\partial u^{(n)}}{\partial \mathbf{n}} = 0 \text{ on } \partial\Omega \,. \tag{5}$$

Hereby we take $u^{(0)} = y$ and consider $u^{(n)}$ an approximation of $u(n\Delta t)$.

Equation (5) is related to adaptive weights smoothing (Polzehl and Spokoiny, 2004) as far as both methods use nonlinear diffusivities depending on the solution and the norm of the gradient of the solution. This brings up the idea of using general smoothing kernels of the form $d = d(u, |\nabla u|)$.

4.2 Fisher information minimisation

As alternative to total variation regularisation we also consider the a regularisation method (Ambrosio et al., 2005), which consists in minimisation of the functional

$$\mathcal{F}_{Fisher}(u) = \frac{1}{2} \int_\Omega (u - y)^2 + \alpha \frac{1}{2} \int_\Omega \frac{|\nabla u|^2}{|u|}$$

which pronounces high peaks in y. The optimality condition for a minimiser u is

$$\frac{u - y}{\alpha} = \nabla \cdot \left(\frac{\nabla u}{|u|} \right) + \frac{u \cdot |\nabla u|^2}{2|u|^3} \,. \tag{6}$$

Identifying $\alpha = \Delta t$, equation (6) can be interpreted as a fully implicit time step of length α of the following flow equation:

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|u|} \right) + \frac{u \cdot |\nabla u|^2}{2|u|^3} \,. \tag{7}$$

In the numerical experiments we have implemented (7) with the semi-implicit scheme

$$u^{(n)} - u^{(n-1)} = \Delta t \left(\nabla \cdot \frac{\nabla u^{(n)}}{\sqrt{|u^{(n-1)}|^2 + \beta^2}} + \frac{u^{(n-1)} \cdot |\nabla u^{(n-1)}|^2}{\sqrt{4|u^{(n-1)}|^6 + \beta^2}} \right) . \qquad (8)$$

Due to the diffusivity $d(u) = 1/|u|$ pronounced (high) peaks are smeared out little, i.e. they remain significant. The right hand side is, like for the TV-flow, invariant to scaling of u.

4.3 The $\frac{3}{2}$ Laplacian flow equation

In the literature the $\frac{3}{2}$ Laplacian operator is defined as

$$\partial \left(\frac{2}{3} \int_\Omega |\nabla u|^{\frac{3}{2}} \right) = -\nabla \cdot (d(u, |\nabla u|) \nabla u)$$

with

$$d(u, \nabla u) = \frac{1}{\sqrt{|\nabla u|}} .$$

We have used the equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (d(u, |\nabla u|) \nabla u) \qquad (9)$$

for filtering. In the numerical solution we have actually approximated d by

$$d_\beta(u, \nabla u) := \sqrt{1/\sqrt{|\nabla u|^2 + \beta^2}} \approx \frac{1}{\sqrt{|\nabla u|}}$$

and implemented the time steps with a semi-implicit algorithm.

This diffusion filtering approach is the steepest descent flow for the $W^{1,3/2}$-Sobolev semi norm.

5 Approximation and Kuiper metrics

Davies and Kovac (2004) use Kuiper metrics to derive a measure of adequacy for univariate densities. In this section we extend their concept to two dimensions. Given bivariate data $x_1, \ldots, x_n$ and a bivariate density function $f(x^1, x^2)$

the problem is to decide whether the data x_i look like a typical sample from f.

The approach we study here is based on projecting the density in the directions of its coordinates. We consider the marginal distribution functions

$$F^1(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x^1, x^2) dx^2 dx^1 \text{ and } F^2(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x^1, x^2) dx^1 dx^2$$

and define two sets of variables $u_i^1 = F^1(x_i^1)$ and $u_i^2 = F^2(x_i^2)$. If the approximation f is adequate, then the new variables u^1 and u^2 should look like two samples of a uniform distribution on $[0, 1]$. Therefore we consider the empirical distribution of each sample

$$E^1(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{u_i^1 \leq x\}} \text{ and } E^2(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{u_i^2 \leq x\}}$$

and calculate the according distances d_1 and d_2 to a uniform distribution in the Kuiper metric

$$d^j(u^j) = d_{Ku}(E^j, U) = (\sup_{x \in [0,1]} E^j(x) - x) - (\inf_{x \in [0,1]} E^j(x) - x).$$

Let $\mathrm{qu}(n, \alpha)$ be the α-quantile of the maximum of $d^1(U^1)$ and $d^2(U^2)$ so that

$$\mathbb{P}(\max(d^1(U^1), d^2(U^2)) \leq \mathrm{qu}(n, \alpha)) = \alpha$$

where $U_1^1, \ldots, U_n^1, U_1^2, \ldots, U_n^2$ are iid random variables with a uniform distribution on $[0, 1]$. Then a function f is considered adequate with the given data if $\max(d^1(u^1), d^2(u^2)) \leq \mathrm{qu}(n, \alpha)$. In this paper we always use $\alpha = 0.99$.

For small values of n the quantiles $\mathrm{qu}(n, \alpha)$ can be obtained by simulation (see Table 1). For larger values of n the distribution of the Kuiper difference between the uniform distribution and its empirical distribution can be approximated by a Brownian bridge and explicit expressions can be derived (Dudley, 1989) and evaluated to obtain quantiles $\mathrm{qu}(n, \alpha)$.

Given this notion of adequacy we are interested in finding an adequate density which is as simple as possible. Since an exact solution of this problem can not be obtained we consider the sequence of functions generated by the discrete diffusion filtering approaches introduced in the previous section and choose the last adequate function as it has the smallest number of local maxima among them as an approximation to the data.

n	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.999$
100	0.175	0.199	0.228
200	0.125	0.141	0.160
500	0.081	0.091	0.105
1000	0.058	0.065	0.074
2000	0.041	0.046	0.052

Table 1
Quantiles for the Kuiper difference for 5 different sample sizes and 3 different values of α.

6 Numerical examples

In the numerical experiments we compare the diffusion filtering methods (5), (8) and (9) with kernel estimators. To solve the flow equations we have implemented the following algorithm:

(1) For given measurement data x_i we calculate the Delaunay triangulation as described in Section 2 (see figure 3).
(2) Define an initial guess $y_i := 1/(marea(I_i))$ for every element of the Delaunay triangulation where m is the number of triangles (see figure 4).
(3) We make a piecewise linear interpolation y on the Delaunay triangles, such that the integral over each triangle is one; that is the integral equals the integral of the piecewise constant initial guess on each triangle. Then we define a regular rectangular grid covering the Delaunay Triangulation and determine the values y of the interpolated initial guess on the regular grid. In our numerical realisation we used a 281×281 grid.
(4) We use a Finite Difference Method to solve the flow equations on the regular grid with piecewise linear initial guess y with a semi-implicit iteration as shown in equation (5) and (8), respectively.
(5) For each iteration we check whether the current function is still adequate with the data. We stop once this is no longer the case and then choose the function from the previous iteration as an approximation to the data.

6.1 A simulated example

A sample of size 500 was drawn from a mixture of two bivariate normal distributions. The underlying density function was given by

$$f(x, y) = 0.5 \cdot \phi(x; 0, 1) \cdot \phi(y; 0, 1) + 0.5 \cdot \phi(x; 2, 0.1) \cdot \phi(y; 2, 0.1)$$

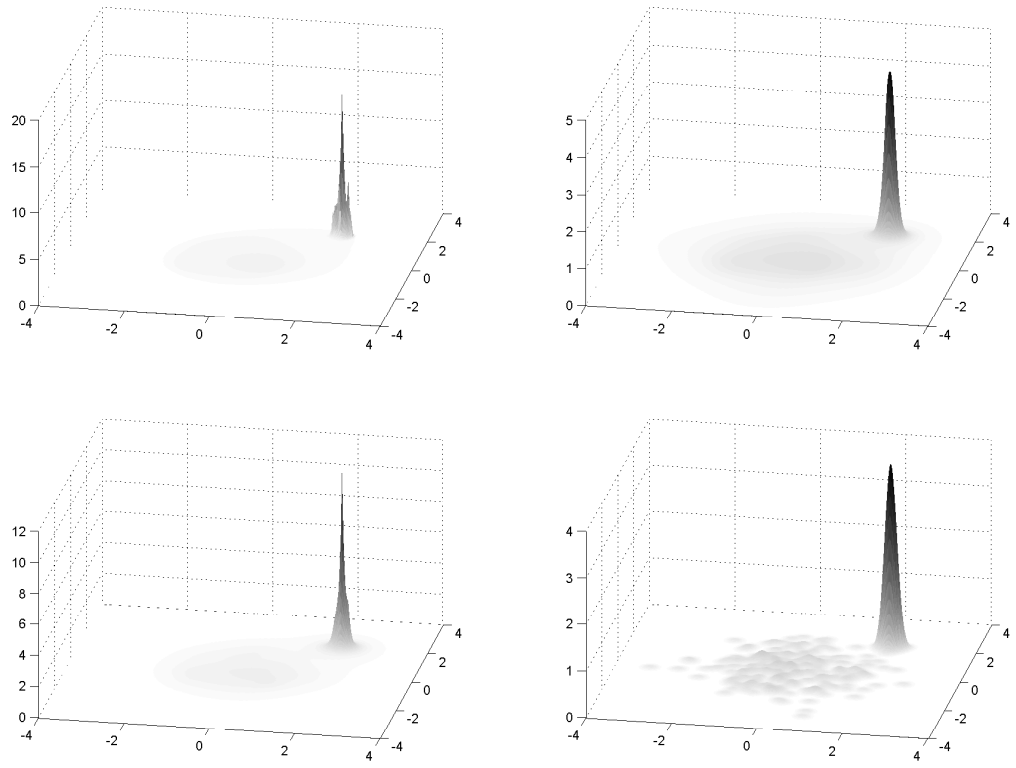


Fig. 5. Top left: TV-flow; Top right: $\frac{3}{2}$ Laplacian flow equation; Bottom left: Fisher information minimisation; Bottom right: Kernel estimator

where $\phi(x; \mu, \sigma)$ denotes the density function of a normal distribution with mean μ and standard deviation σ.

The results are shown in Figure 5. The result top left was calculated using the total variation flow equation where in the numerical realisation the time step size was set to 0.005 and where the flow was stopped by the Kuiper criterion at time $t = 0.33$, i.e. after 66 time steps.

Using equation (8) we obtain the solutions represented in the top right of the figure. Here the time step size was set equal to 0.05 and 49 time steps were performed. The bottom left of Figure 5 shows the result using equation (9) where the time step size was set to 0.05 and $t = 3.9$ (i.e. 78 time steps). Finally the bottom right of the figure shows the result of an kernel estimator using a Gaussian kernel and the largest bandwidth such that the Kuiper criterion was satisfied (0.947).

Comparing the results the most obvious observation is that the kernel estimator produces a much more rough approximation than the diffusion filtering methods. A close inspection reveals that the kernel estimator produces 82 lo-

cal maxima whereas TV-flow needs only 6 local maxima and $\frac{3}{2}$-Laplacian and Fisher flow both produce exactly 2 local maxima.

We also note that TV-flow tends to create levels with constant values. The $\frac{3}{2}$-Laplacian flow is a good compromise between smoothing and preserving high peaks with small footpoints (base area). The best results are obtained using the Fisher information minimisation, since the solution u itself is used in the diffusivity ($d(u) = 1/|u|$) and the gradient of u is taken into account in an energy term.

The advantage of using flow equations compared to minimising energy functionals where α is fixed, is that the solution for different times (i.e. different values of α, see the results in Section 2, Steidl et al., 2004) can be calculated efficiently, i.e. no additional iteration is needed.

6.2 The geyser data

It is interesting to see how the diffusion filtering methods performs on the geyser data that we introduced in Section 1. Figure 6 shows perspective plots of the three methods and a contour plot for the $\frac{3}{2}$ Laplacian flow. The bimodality of the data set is clearly shown by all three methods although a close inspection of the output of the Fisher flow reveals an artificial third local extreme value to satisfy the Kuiper criterion. BV flow and Laplacian flow on the other hand both produce only the required two local extreme values.

Conclusions

In this paper we have shown that Delaunay triangulation and diffusion filtering generalise the concept of the taut-string algorithm for analysing high dimensional data. In principle the concept applies in any space dimension, however, so far, the dimensionality is limited by the partial differential equation solver where up to date software can handle space dimension three, but typically not more.

Acknowledgement

The work of OS is supported by the Austrian Science Foundation (FWF) Projects Y-123INF, FSP 9203-N12 and FSP 9207-N12. Part of the work of AO and OS has been supported by the Austrian Ministry for Economy and Labour

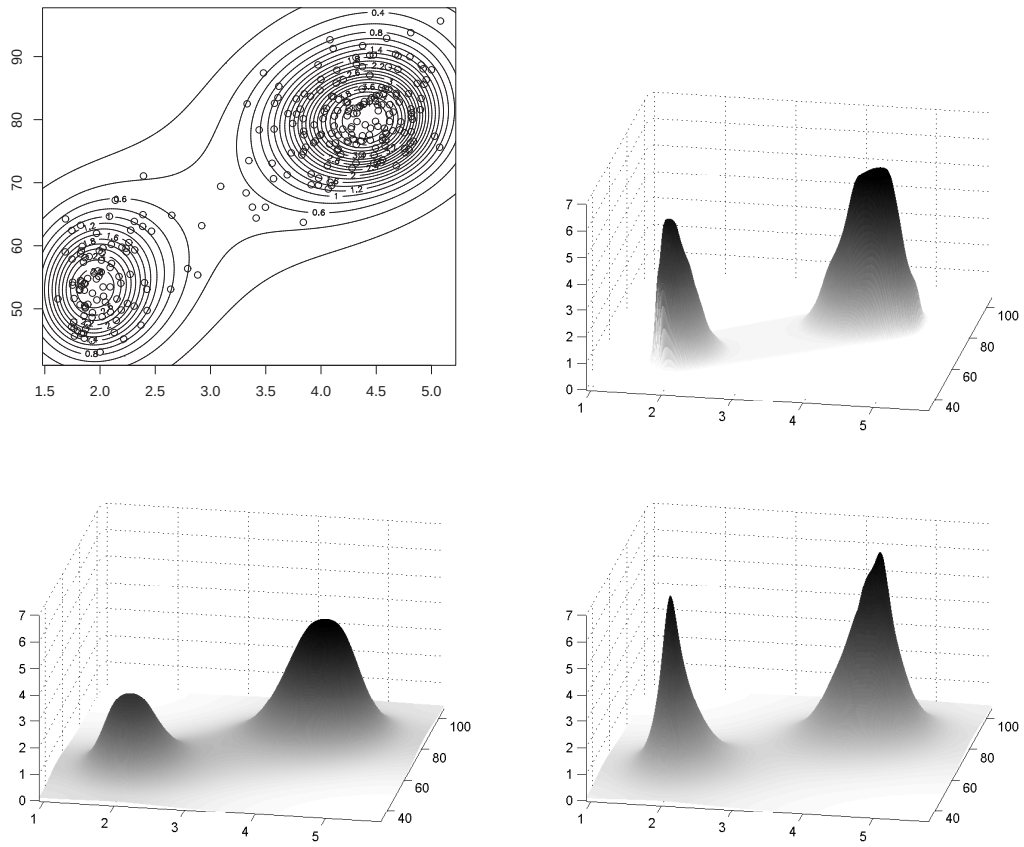


Fig. 6. Top left: Scatter plot of geyser data and contour plot of $\frac{3}{2}$ Laplacian flow; Top right: TV-flow; Bottom left: $\frac{3}{2}$ Laplacian flow; Bottom right: Fisher flow

and by the Government of Upper Austria within the framework "Industrial Competence Centers" The work of AK was supported by the DFG within the Sonderforschungsbereich 475 at the University of Dortmund.

OS wants to thank Franz Aurenhammer and Oswin Aichholzer (Graz) for many discussions on Delaunay's triangulations.

References

Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser.

Aurenhammer, F. and Klein, R. (2000). *Handbook of Computational Geometry*, chapter 5, pages 201–290. Elsevier Science Publishing.

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data*

Analysis: the Kernel Approach with S-Plus Illustrations. Oxford University Press, Oxford.

Brézis, H. (1973). *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. Number 5 in North-Holland Mathematics Studies. North-Holland Publishing Co, Amsterdam-London.

Davies, P. (1995). Data features. *Statistica Neerlandica*, 49:185–245.

Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Annals of Statistics*, 29:1–65.

Davies, P. L. and Kovac, A. (2004). Densities, spectral densities and modality. *Annals of Statistics*, 32(3):1093–1136.

Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539.

Dudley, R. (1989). *Real Analysis and Probability*. Wadsworth and Brooks/Cole, California.

Edelsbrunner, H. (2001). *Geometry and Topology for Mesh Generation*. Cambridge University Press.

Evans, L. C. and Gariepy, R. F. (1992). *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL.

Fisher, N. I., Mammen, E., and Marron, J. S. (1994). Testing for multimodality. *Computational Statistics and Data Analysis*, 18:499–512.

Good, I. and Gaskins, R. (1980). Density estimating and bump-hunting by the penalizedlikelihood method exemplified by scattering and meteorite data. *Journal of the American Statistical Association*, 75:42–73.

Grasmair, M. (2006). The equivalence of the taut string algorithm and bv-regularization. *Journal of Mathematical Imaging and Vision*.

Hengartner, N. and Stark, P. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Annals of Statistics*, 23:525–550.

Herrick, D., Nason, G., and Silverman, B. (200). Some new methods for wavelet density estimation. Technical report, Department of Mathematics, University of Bristol,United Kingdom.

Hinterberger et al., W. (2003). Tube methods for bv regularization. *Journal of Mathematical Imaging and Vision*, 19:219–235.

Loader, C. (1999). Bandwidth selection: classical or plug-in? *Annals of Statistics*, 27:415–438.

Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Annals of Statistics*, 25:387–413.

Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.*, 86(415):738–746.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 10:186–190.

Polonik, W. (1995a). Density estimation under qualative assumptions inhigher dimensions. *Journal of Mulitivariate Analysis*, 55:61–81.

Polonik, W. (1995b). Measuring mass concentrations and estimating density-contour clusters–an excess mass approach. *Annals of Statistics*, 23:855–881.

Polonik, W. (1999). Concentration and goodness of fit in higherdimensions:

(aymptotically) distribution-free methods. *Annals of Statistics*, 27:1210–1229.

Polzehl, J. and Spokoiny, V. (2004). Propagation-separation approach for local likelihood estimation. *WIAS-Preprint 787*.

Pöschl, C. and Scherzer, O. (2006). Characterization of minimizers of convex regularization functionals. Technical Report 42, Institute for Computer Science, University Innsbruck.

Richardson, S. and Green, P. (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society*, 59:731–792.

Roeder, K. and Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *J. Amer. Statist. Assoc.*, 92(439):894–902.

Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

Sain, S. and Scott, D. (1996). On locally adaptive density estimation. *Journal of the American Statistical Association*, 91:1525–1534.

Scherzer, O. (2005). Taut-string algorithm and regularization programs with g-norm data fit. *Journal of Mathematical Imaging and Vision*, 23(2):135–143.

Sheather, S. and Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, 53:683–690.

Silverman, B. W. (1986). *Density estimation for statistics amd data analysis*. Chapman and Hall, London.

Simonoff, J. (1996). *Smoothing methods in statistics*. Springer-Verlag, New York.

Steidl, G., Weickert, J., Brox, J., Mrazek, P., and Welk, M. (2004). On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM J. Numer. Anal.*, 42(2):686–713.

Verdinelli, I. and Wasserman, L. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *Ann. Statist.*, 26(4):1215–1241.

Vidakovic, B. (1999). *Statistical modeling based on wavelets*. Wiley, New York.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. Chapman and Hall Ltd., London.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā*, 26:101–116.