Preserving data moments in density estimation via diffusion using the finite element method

Preservación de momentos de datos en estimación de densidades vía difusión usando el método de elementos finitos

Keith Y. Patarroyo[1,a], **Juan Galvis**[2,b], **Francisco Gómez**[2,c]

Abstract. We design a two-dimensional density estimation scheme via diffusion that conserves the first order moments and the total mass in the estimation process. In order to conserve the first order moments and the total mass throughout the time iteration, a non-local boundary condition is imposed to the diffusion operator. A discrete method is realized by using the finite element method where the boundary condition is weakly imposed using Lagrange multipliers that leads to the solution of a saddle point problem. We show some numerical examples in different geometries using FeniCS.

Keywords: KDE, Diffusion equation, Moments preserving evolution, FEM, Lagrange Multipliers.

Resumen. Se diseña un esquema de estimación de densidades vía difusión que conserva los momentos de primer orden y la masa total en el proceso de estimación. Para poder conservar los momentos de primer orden y la masa total a través del tiempo, se impone una condición de frontera no local al operador de difusión. Un método discreto es propuesto usando el método de elementos finitos donde las condiciones de frontera son impuestas débilmente usando multiplicadores de Lagrange que llevan a un problema de punto de silla. Mostramos algunos experimentos numéricos con distintas geometrías usando FEniCS.

Palabras claves: KDE, Ecuación de Difusión, evolución de preservación de momentos, FEM, Multiplicadores de Lagrange.

Mathematics Subject Classification: 74S05, 62G07, 76R50.

Recibido: octubre de 2018 Aceptado: marzo de 2019

[1]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada
[2]Departamento de Matemática, Universidad Nacional de Colombia, Bogotá, Colombia
[a]keith.patarroyo@umontreal.ca
[b]jcgalvisa@unal.edu.co
[c]fagomezj@unal.edu.co

1. Introduction and Motivation

In this paper, we deal with the estimation of a probability density function from data samples. Our main motivation is the application of such tools to design efficient security policies. Crime and security administration are important aspects of modern society, in particular, big cities around the world have very serious crime issues, robbery, assault and drug-related crimes are of major concern in the daily life of all citizens [11]. Many mathematical models have been proposed to help understand crime-related issues, some are discrete stochastic models [14, 18] other are continuous crime probability density estimations [12]. A popular continuous density estimation method is KDE (Kernel Density Estimation), a well studied statistical tool used in many areas of knowledge [17, 8, 10]. It is also used by many private companies that are currently offering to map crime spreading in city zones [6, 5]. See [12] for an application of these methodologies to study crime in the city of Bogotá (Colombia).

Roughly speaking, if we are in a scenario where we can apply the KDE method, the main ingredients to have into account related to the implementation of the methods are

1. Initial approximation: The data generate and initial approximation of the probability density function associated to the data

2. Initial-boundary value problem: A parabolic equation (such as the heat equation or heterogeneous heat equation) should be posed in a domain enclosing the data

3. Stopping time: The final approximation of the probability density function results from time evolving the initial approximation in 1. until a certain specific time T^* using the parabolic differential equation 2.

In this paper we concern mainly with item 2. above and we refer to [8, 10, 17] and references therein for further details on 1. and 3. An important point is that, when enough data samples are available, the empirical distribution used in 1. above has the right first low order moments such as the directional averages and the variance. Therefore, and important aspect of the KDE method, to our knowledge not studied before, is whenever the time evolution does conserve some important quantities such the x and y averages. In such a case, where low-order moments are preserved, we believe the final probability density function is more specific to the initial data set. We design low order moments conserving KDE scheme. In [8, 10, 17] the authors use, in step 2. above, a heat equation on a bounded domain with homogeneous Neumann boundary condition since imposing homogeneous boundary condition ensures preservations of the total mass of the probability density function. In this work we develop new boundary conditions in order to ensure the conservation of low order moments thorough the time evolution. In particular we detail the boundary condition for the conservation of the components average of the initial approximation used in item 1. above. For the discrete realization of our method we employ

the Finite Element Method (FEM) applied to the two-dimensional diffusion equation with appropriate boundary conditions. In th discrete problem, the new boundary condition is imposed by using the Lagrange multiplier formulation which leads to the solution of a saddle point problem that can be solved efficiently using classical iteration for these systems and classical solvers for finite element stiffness matrices.

The rest of the paper is organized as follows: In Section 2 we present the mathematical model and the conserving moments density estimation. In Section 3 we present the finite element discretization where we use a Lagrange multiplier to impose the conserving moments restrictions. In Section 3 we present some numerical examples with evidence that we indeed can compute density estimation that preserve the total mass and the first order moments of the inital data.

2. Mathematical Model

Given samples $\{\mathbf{x}_i\}_{i=1}^N$ in a domain $\Omega \subset \mathbb{R}^n$ the goal is to compute and approximation of a probability density functions associated to this data set. In order to fix ideas and simplify the presentation we consider the case $n = 2$.

In order to model this problem we use the method of non-parametric statistical estimation, KDE (Kernel Density Estimation) [17]. Heuristically this method approximates the wanted density by an adjusted sum of *kernel* functions centered on the location of the data points, hence with this method the estimated density is fully determined by *all* the data. As is widely known the quality of KDE is strongly dependent on the bandwidth parameter h, this determines the *smothness* of the estimation. In this work we do not consider algorithms for bandwidth selection, we consider mainly the conservation of some moments at all values of h of the estimated density. For more information on the bandwidth selection see [8].

As it was first noted in [9] and expanded in [8], solving the diffusion equation with a discrete data sample $\{\mathbf{s}_i\}_{i=1}^N$ as initial condition in all $\mathbb{R}^n$ gives an estimate of a continuous probability density function, where bandwidth parameter becomes proportional to the time, $h \propto t$. This also holds for a bounded domain $\Omega \subset \mathbb{R}^2$ and with a discrete data sample $\{\mathbf{x}_i\}_{i=1}^N$ inside this domain as initial condition. Therefore, by solving the diffusion equation [16],

$$\frac{\partial u(\mathbf{x},t)}{\partial t} - \Delta u(\mathbf{x},t) = 0 \qquad \mathbf{x} = (x_1, x_2) \in \Omega \subset \mathbb{R}^2,\ t > 0, \tag{1}$$

with initial condition,

$$u(\mathbf{x},0) = \frac{1}{N}\sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}_i) \qquad \mathbf{x}, \mathbf{x}_i \in \Omega \subset \mathbb{R}^2, \tag{2}$$

and some boundary condition, we are able to estimate a continuous probability density function which corresponds to the solution of the equation above

stopped at certain positive time.

In the case of having bounded domains, applying boundary conditions to this problem is key to obtain an accurate results. Botev et.al. in [8] propose to use homogeneous Neumann boundary conditions,

$$\frac{\partial u(\mathbf{x},t)}{\partial \boldsymbol{\nu}} = 0, \qquad \mathbf{x} \in \partial\Omega. \tag{3}$$

They claim that this boundary condition accurately estimate probability densities. As it was shown in [15] for a one dimensional data set, homogeneous Neumann boundary conditions are insufficient to conserve both the mean and the mass of the initial data over distinct values of t. In the case of higher dimensions, two dimensions for instance, in order to preserve the averages and the total mass we impose the following to restrictions

$$\int_{\partial\Omega} x_k \frac{\partial u(\mathbf{x},t)}{\partial \boldsymbol{\nu}} d\sigma = \int_{\Omega} \nabla_k u(\mathbf{x},t) d\mathbf{x}, \quad k = 1, 2, \tag{4}$$

and

$$\int_{\partial\Omega} \frac{\partial u(\mathbf{x},t)}{\partial \boldsymbol{\nu}} d\sigma = 0, \tag{5}$$

where $\nabla_i u(\mathbf{x},t) = \mathbf{e}_i \cdot \nabla u(\mathbf{x},t)$ and $\nabla u \cdot \boldsymbol{\nu} = \frac{\partial u(\mathbf{x},t)}{\partial \boldsymbol{\nu}}$. Here, $\boldsymbol{\nu}$ is the normal vector to Ω. The density estimation scheme that satisfies the three functional restrictions above conserves the x and y average and the total mass of the sample. Similar additional restrictions can be designed in order to conserve additional moments such as $\mathbf{E}[u^n]$. This is object of future research for some of the authors. We can also note that these new boundary conditions are non-local and therefore the mathematical analysis is not straightforward when we impose (4)-(5). In order to write a practical estimation, what we compute is, in each time step calculation, the solution with least energy among all functions (in a particular Hilbert space) that satisfy the restrictions (4)-(5). This minimization can be implemented using Lagrange multipliers formalism as we show in the next section.

3. Numerical Solution

In order to approximate the solution to problem (1)-(2) with the restrictions (4)-(5), we employ the Finite Element Method(FEM) to discretize the space variable and the Euler method to discretize the time. Any other time discretization can be used as well. The weak formulation of the problem (1) can be written as,

$$\int_{\Omega} \frac{\partial u(\mathbf{x},t)}{\partial t} v(x) d\mathbf{x} + \int_{\Omega} \nabla u(\mathbf{x},t) \nabla v(\mathbf{x}) d\mathbf{x} = \int_{\partial\Omega} \frac{\partial u(\mathbf{x},t)}{\partial \boldsymbol{\nu}} v(\mathbf{x}) d\sigma \tag{6}$$

where all $v \in H^1(\Omega)$ (The Sobolev space of square integrable functions with square integrable partial derivatives).

Now for the discretization of time we use the implicit Euler method. Given u^0, we compute $u^1, u^2, u^3, \ldots$, such that

$$\int_\Omega [u^{n+1}(\mathbf{x}) - u^n(\mathbf{x})]v(\mathbf{x})d\mathbf{x} + \Delta t \int_\Omega \nabla u^{n+1}(\mathbf{x})\nabla v(\mathbf{x})d\mathbf{x} = \Delta t \int_{\partial\Omega} \frac{\partial u^{n+1}(\mathbf{x})}{\partial \boldsymbol{\nu}} v(\mathbf{x})d\sigma \tag{7}$$

for all test function $v \in H^1(\Omega)$.

For the Galerkin formulation, let $\mathcal{K}$ be a triangulation of Ω, and let V_h be a standard finite element space on $\mathcal{K}$. For a given u_h^0 find u_h^1, $u_h^2, \ldots \in V_h$ such that

$$\int_\Omega [u_h^{n+1}(\mathbf{x}) - u_h^n(\mathbf{x})]v(\mathbf{x})d\mathbf{x} + \Delta t \int_\Omega \nabla u_h^{n+1}(\mathbf{x})\nabla v(\mathbf{x})d\mathbf{x} = \Delta t \int_{\partial\Omega} \frac{\partial u_h^{n+1}(\mathbf{x})}{\partial \boldsymbol{\nu}} v(\mathbf{x})d\sigma \tag{8}$$

for all $v \in V_h$.

In order to compute the finite element approximation u_h^n, let $\{\phi_i(\mathbf{x})\}_{i=1}^{m_h}$ be the standard finite element basis for V_h. Then we have the linear combination,

$$u_h^n = \sum_{j=1}^{m_h} \alpha_j^n \phi_j(\mathbf{x}). \tag{9}$$

The previous formulation yield the matrix problem: given $\vec{\alpha}^0$ find $\vec{\alpha}^1, \vec{\alpha}^2, \vec{\alpha}^3, \ldots$ such that,

$$\mathbf{C}\vec{\alpha}^{n+1} = \mathbf{M}\vec{\alpha}^n, \tag{10}$$

where $\mathbf{C} = \mathbf{M} + \Delta t\mathbf{A} - \Delta t\mathbf{D}$ and

$$\vec{\alpha}^n = [\alpha_i^n]^T, \qquad \mathbf{M} = \left[\int_\Omega \phi_i(\mathbf{x})\phi_j(\mathbf{x})d\mathbf{x}\right], \qquad \vec{b} = \mathbf{M}\vec{\alpha}^n, \tag{11}$$

$$\mathbf{A} = \left[\int_\Omega \nabla\phi_i(\mathbf{x})\nabla\phi_j(\mathbf{x})d\mathbf{x}\right], \qquad \mathbf{D} = \left[\int_{\partial\Omega} \frac{\partial\phi_i(\mathbf{x})}{\partial\boldsymbol{\nu}}\phi_j(\mathbf{x})d\sigma\right]. \tag{12}$$

It remains to impose the boundary conditions. By plugging (9) into (4) and (5) we find,

$$\int_{\partial\Omega} \frac{\partial u_h^n(\mathbf{x},t)}{\partial\boldsymbol{\nu}}d\sigma = \sum_{j=1}^{n_i} \alpha_j^n \int_{\partial\Omega} \frac{\partial\phi_i(\mathbf{x})}{\partial\boldsymbol{\nu}}d\sigma = \vec{c}\cdot\vec{\alpha}^n = 0, \tag{13}$$

and also,

$$\begin{aligned} &\int_{\partial\Omega} x_k \frac{\partial u_h^n(\mathbf{x},t)}{\partial\boldsymbol{\nu}}d\sigma - \int_\Omega \nabla_k u_h^n(\mathbf{x},t)d\mathbf{x} \qquad (14)\\ &= \sum_{j=1}^{n_i} \alpha_j^n \left[\int_{\partial\Omega} x_k \frac{\partial\phi_i(\mathbf{x})}{\partial\boldsymbol{\nu}}d\sigma - \int_\Omega \nabla_k\phi_i(\mathbf{x})d\mathbf{x}\right]\\ &= \vec{d}^{(k)}\cdot\vec{\alpha}^n = 0, \quad k = 1,2, \qquad (15)\end{aligned}$$

where $\vec{c}$ and $\vec{d}^{(k)}$ are vectors with coordinates given by,

$$\vec{c} = \left[\int_{\partial\Omega} \frac{\partial \phi_i(\mathbf{x})}{\partial \boldsymbol{\nu}} d\sigma\right]^T, \quad \vec{d}_k = \left[\int_{\partial\Omega} x_k \frac{\partial \phi_i(\mathbf{x})}{\partial \boldsymbol{\nu}} d\sigma - \int_{\Omega} \nabla_k \phi_i(\mathbf{x}) d\mathbf{x}\right]^T, \ k = 1, 2. \tag{16}$$

We can impose these constraints on the current time solution by applying the Lagrange multiplier technique [13]. Considering the Lagrangian,

$$\mathcal{L} = \frac{1}{2} \left(\vec{\alpha}^n\right)^T \mathbf{C} \vec{\alpha}^n - \left(\vec{\alpha}^n\right)^T \vec{b} - \left(\vec{\lambda}^n\right)^T \mathbf{B} \vec{\alpha}^n, \tag{17}$$

where

$$\vec{\lambda}^n = [\lambda_i^n]^T, \qquad \mathbf{B} = \begin{bmatrix} \vec{c} \\ \vec{d}_1 \\ \vec{d}_2 \end{bmatrix}. \tag{18}$$

We need to compute the stationary points of the Lagrangian leads to the augmented linear system,

$$\begin{bmatrix} \mathbf{C} & \mathbf{B}^T \\ \mathbf{B} & 0 \end{bmatrix} \begin{bmatrix} \vec{\alpha}^n \\ \vec{\lambda}^n \end{bmatrix} = \begin{bmatrix} \mathbf{M}\vec{\alpha}^n \\ 0 \end{bmatrix}, \tag{19}$$

from which $\vec{\alpha}^n$ can be found.

Hence we obtained the matrix formulation of our problem, given $\vec{\alpha}^0$ find $\vec{\alpha}^1$, $\vec{\alpha}^2$, $\vec{\alpha}^3, \ldots \in \mathbb{R}^{n_i}$ such that they solve (19) where $\mathbf{C}, \mathbf{B}, \vec{b}, \vec{\alpha}^n, \vec{\lambda}^n$ are how they were previously defined.

3.1. Numerical Experiments

We present some numerical results using FeniCS, [7], in different domains. First we consider a simple rectangular domains and then we show the implementation on a different domain.

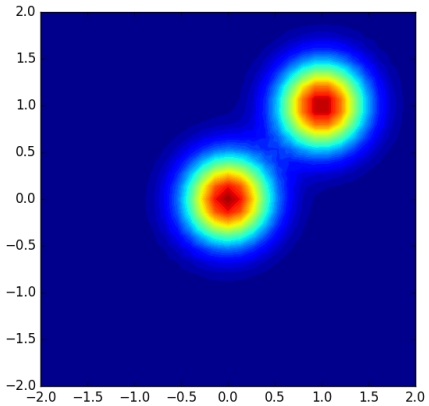
Rectangular Domain-1 $\Omega = [-2, 2] \times [-2, 2]$**:**

For this example we considered a triangulation $\mathcal{K}$ with 30×30 triangles and we used the piecewise linear polynomials as our trial and test function space, that is, $V_h = \mathcal{P}_1(\mathcal{K})$. We use the initial data,

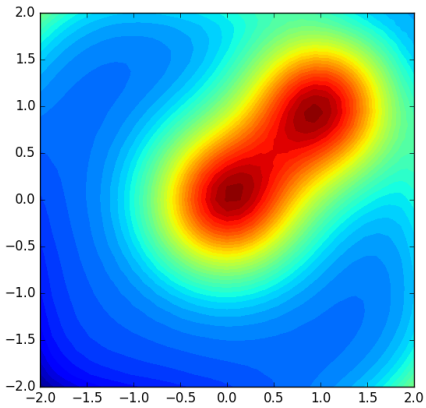
$$u_0^h = e^{-5\left(x^2+y^2\right)} + e^{-5\left((x-1)^2+(y-1)^2\right)}, \tag{20}$$

where we are representing the data events as Gaussian hills interpolated in the space V_h and centered on the location of the events.

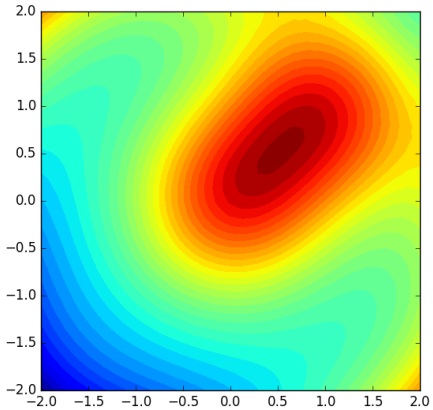
The solution obtained by solving (19), see Appendix A, is depicted in Figures 1(a)-1(d). For comparison we show in Figure 2 the solution at the same time locations of the problem with regular homogeneous Neumann solutions.



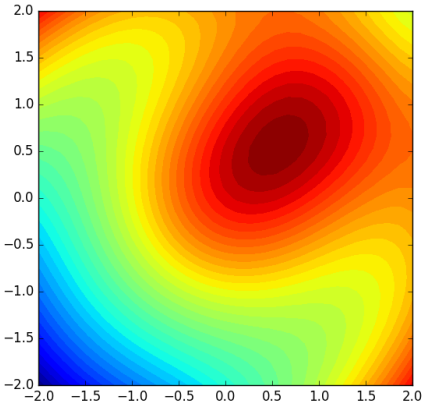
(a) Solution of the problem (19) with parameters in (20) at $t = 0$.



(b) Solution of the problem (19) with parameters in (20) at $t = 0.125$.

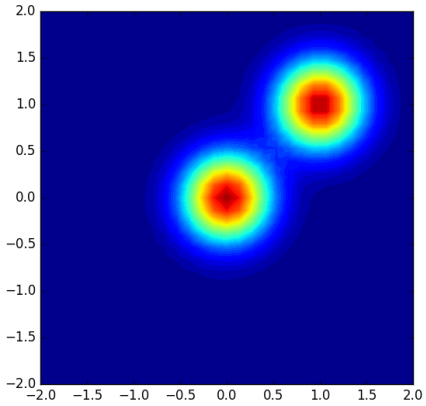


(c) Solution of the problem (19) with parameters in (20) at $t = 0.25$.

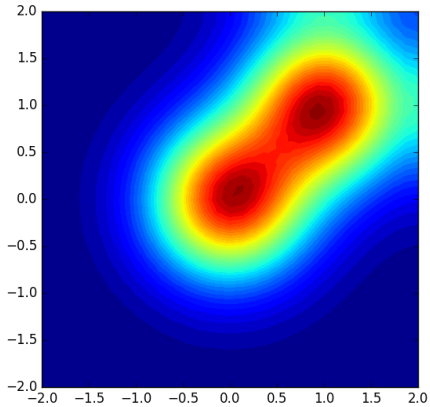


(d) Solution of the problem (19) with parameters in (20) at $t = 0.375$.

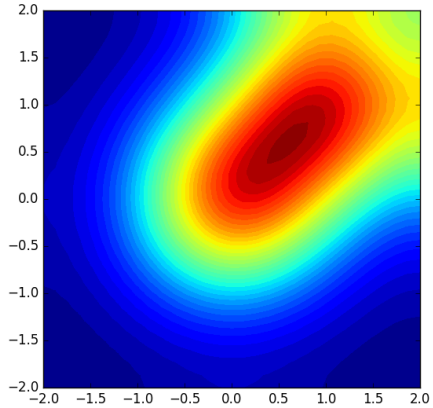
Figure 1: Solution of the problem (19) for the domain $\Omega = [-2, 2] \times [-2, 2]$ and parameters in (20) at distinct times in the interval $t = [0, 0.5]$.



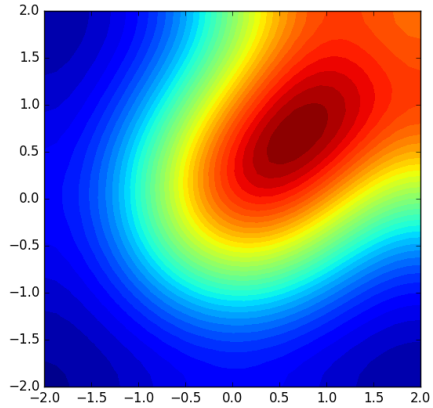
(a) Solution of the Neumann Problem with parameters in (20) at $t = 0$.



(b) Solution with homogeneous Neumann boundary condition and parameters in (20) at $t = 0.125$.



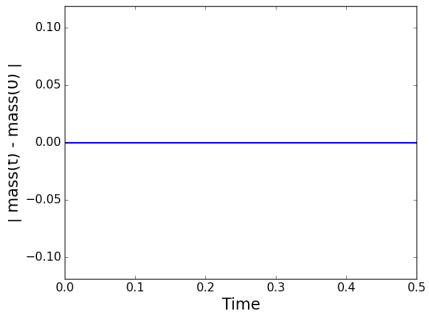
(c) Solution with homogeneous Neumann boundary condition and parameters in (20) at $t = 0.25$.



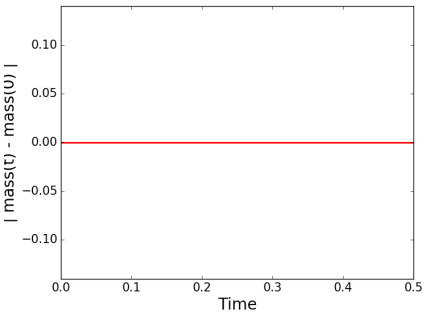
(d) Solution with homogeneous Neumann boundary condition and parameters in (20) at $t = 0.375$.

Figure 2: Solution of the diffusion equation with homogeneous Neumann boundary condition for the domain $\Omega = [-2, 2] \times [-2, 2]$ with parameters in (20) at distinct times in the interval $t = [0, 0.5]$, it can be noted that the solution spreads slower than the mean boundary conditions solution.

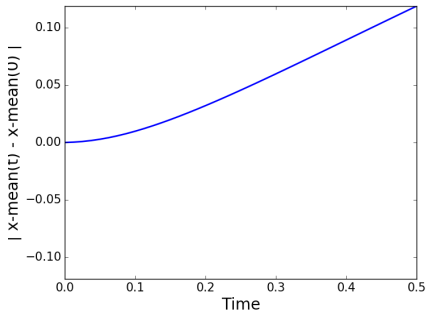
We additionally show the evolutions of the total mass $\Delta m(t) = |m(t) - m(0)|$ and the two averages $\Delta\mu_i(t) = |\mu_i(t) - \mu_i(0)|$ for $i = 1, 2$ over the time evolution of the simulation. See Figure 3.



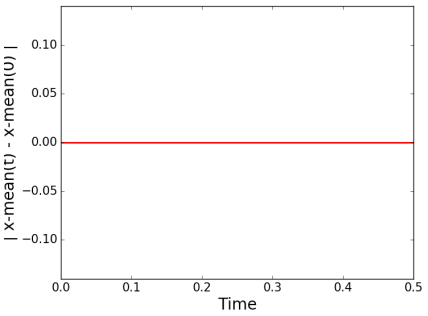
(a) Change in the total mass Δm for the numerical solution with homogeneous Neumann boundary condition.



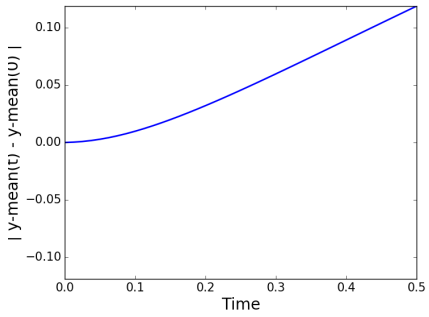
(b) Change in the total mass Δm for the numerical solution with the new boundary conditions.



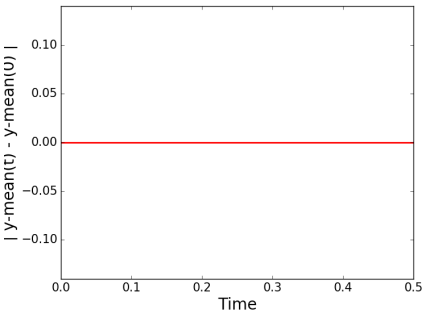
(c) Change in the x-mean $\Delta \mu_x$ for the numerical solution with homogeneous Neumann boundary condition.



(d) Change in the x-mean $\Delta \mu_x$ for the numerical solution with the new boundary conditions.



(e) Change in the y-mean $\Delta\mu_y$ for the numerical solution with homogeneous Neumann boundary condition.



(f) Change in the y-mean $\Delta\mu_y$ for the numerical solution with the new boundary conditions.

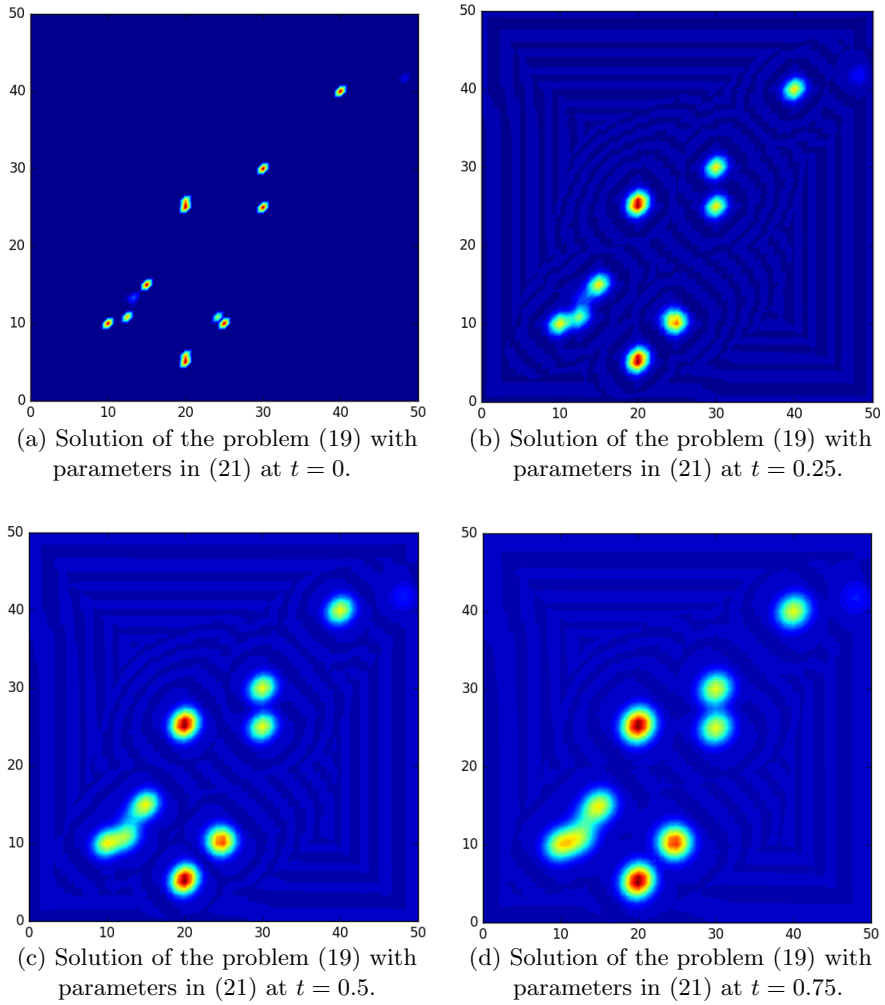
Figure 3: Plots of the evolution of the change in total mas Δm, $x-$ mean $\Delta\mu_x$, and $y-mean\ \Delta\mu_y$, for the density estimation with homogeneous Neumann and the new set of boundary conditions for the problem (19) for the domain $\Omega = [-2,2]\times[-2,2]$ and parameters in (20) in the interval $t \in [0, 0.5]$.

Rectangular Domain-2 $\Omega = [0,50]\times[0,50]$**:**

For this example we considered a triangulation $\mathcal{K}$ formed by 60×60 triangles and we used the linear Lagrange Polynomials as our trial and test function space, $V_h = \mathcal{P}_1(\mathcal{K})$. We use

$$u_0^h = \sum_{i=0}^{15} e^{-5\left((x-x_i)^2+(y-y_i)^2\right)}, \tag{21}$$

The solution obtained are depicted in Figures 7(a)-7(d). For comparison we show in Figure 5 the solution at the same times of the problem with regular homogeneous Neumann solutions.



(a) Solution of the problem (19) with parameters in (21) at $t = 0$.

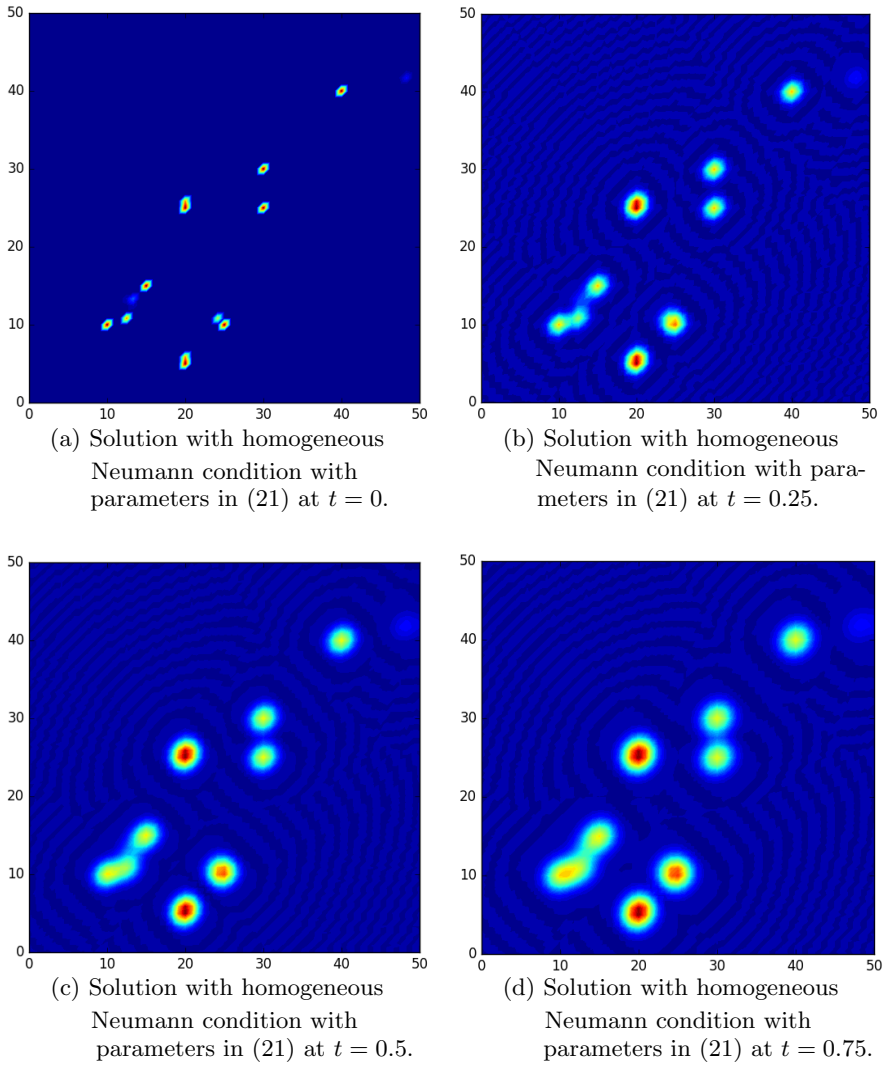
(b) Solution of the problem (19) with parameters in (21) at $t = 0.25$.

(c) Solution of the problem (19) with parameters in (21) at $t = 0.5$.

(d) Solution of the problem (19) with parameters in (21) at $t = 0.75$.

Figure 4: Solution of the problem (19) for the domain $\Omega = [0, 50] \times [0, 50]$ and parameters in (21) at distinct times in the interval $t = [0, 1]$, it can be noted that the solution is spread along all the domain for big values of t.

We can compare the resulting solution with the evolution using homogeneous Neumann boundary condition. See Figure 5.



(a) Solution with homogeneous Neumann condition with parameters in (21) at $t = 0$.

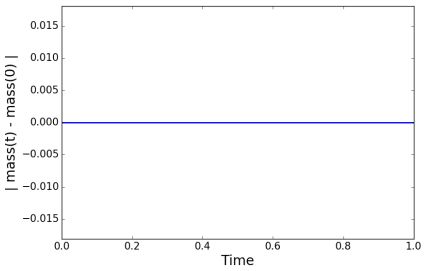
(b) Solution with homogeneous Neumann condition with parameters in (21) at $t = 0.25$.

(c) Solution with homogeneous Neumann condition with parameters in (21) at $t = 0.5$.

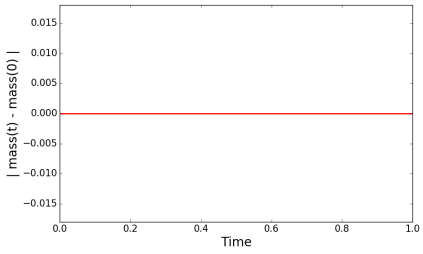
(d) Solution with homogeneous Neumann condition with parameters in (21) at $t = 0.75$.

Figure 5: Solution with homogeneous Neumann boundary condition for the domain $\Omega = [0, 50] \times [0, 50]$ with parameters in (21) at distinct times in the interval $t = [0, 1]$.

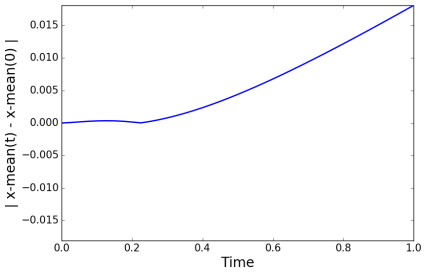
We additionally show the evolutions of the total mass $\Delta m(t) = |m(t) - m(0)|$ and the two averages $\Delta\mu_i(t) = |\mu_i(t) - \mu_i(0)|$ for $i = 1, 2$ over the time evolution of the simulation. See Figure 6.



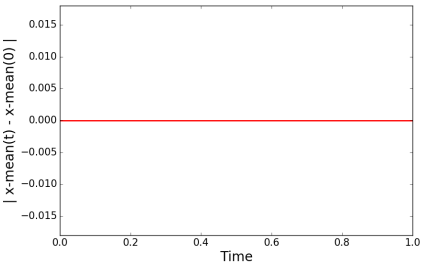
(a) Change in total mass Δm for the numerical solution with homogeneous Neumann boundary conditions.



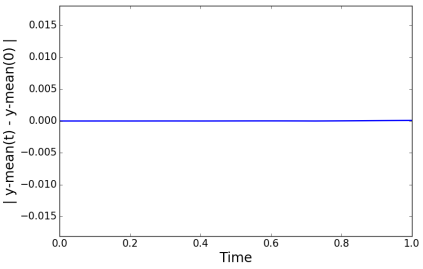
(b) Change in total mass Δm for the numerical solution with the new boundary conditions.



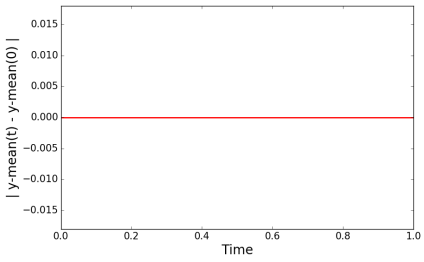
(c) Change in the x-mean $\Delta \mu_x$ for the numerical solution with homogeneous Neumann boundary condition.



(d) Change in x-mean $\Delta \mu_x$ for the numerical solution with the new boundary conditions.



(e) Change in y-mean $\Delta\mu_y$ for the numerical solution with homogeneous Neumann boundary condition.



(f) Change in y-mean $\Delta\mu_y$ for the numerical solution with the new boundary conditions.

Figure 6: Plots of the evolution of the changes in total mass Δm, $x-$mean $\Delta\mu_x$, and $y-$mean $\Delta\mu_y$, for the density estimation with homogeneous Neumann and the new boundary conditions for the problem (19) for the domain $\Omega = [0, 50] \times [0, 50]$ and parameters(21) in the interval $t \in [0, 1]$.

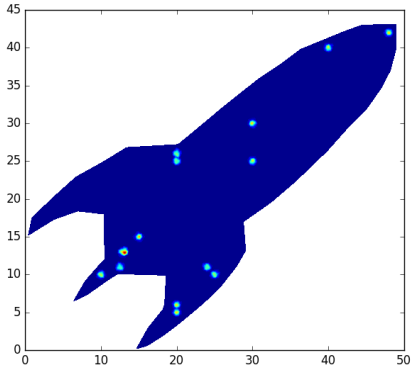
Irregular Domain Ω

For this example we considered a triangulation $\mathcal{K}$ given by the triangulation of an image of a rocket [4] and we used the software Gmsh [3] to generate a mesh. Additionally we used the linear Lagrange Polynomials $\mathcal{P}_1(\mathcal{K}_{im})$ as our trial and test function space $V_h = \mathcal{P}_1(\mathcal{K}_{im})$.

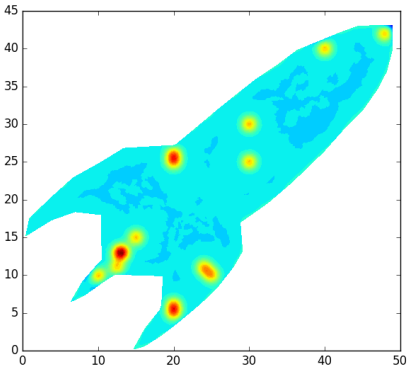
The important parameters of this simulation are,

$$T = 1.0, \qquad steps = 40, \qquad u_0^h = \sum_{i=0}^{15} e^{-5\left((x-x_i)^2+(y-y_i)^2\right)}, \tag{22}$$

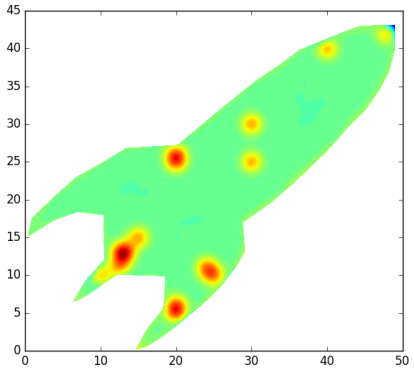
where we are considering a irregular domain and the same initial data points as the previous square. The resulting program with the specification of x_i and y_i is available online in [2]. If we solve (19) with the FeniCS solver we obtain solutions depicted in Figure 7.



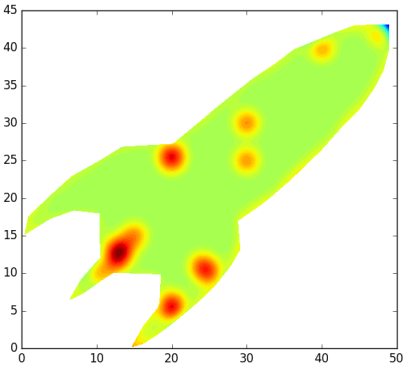
(a) Solution of the problem (19) with parameters in (22) at $t = 0$.



(b) Solution of the problem (19) with parameters in (22) at $t = 0.25$.



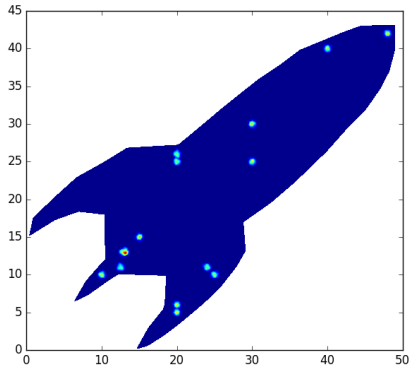
(c) Solution of the problem (19) with parameters in (22) at $t = 0.5$.



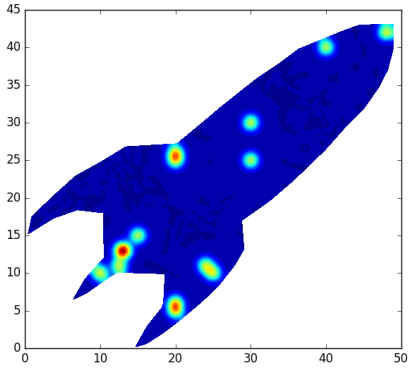
(d) Solution of the problem (19) with parameters in (22) at $t = 0.75$.

Figure 7: Solution of the problem (19) for the domain Ω with the shape of a "Rocket" and parameters in (22) at distinct times in the interval $t = [0, 1]$.

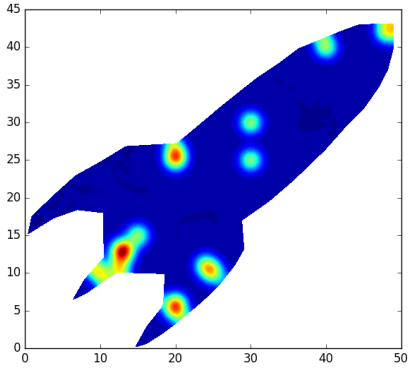
We can compare the resulting solution with the homogeneous Neumann boundary condition. See Figure 8.



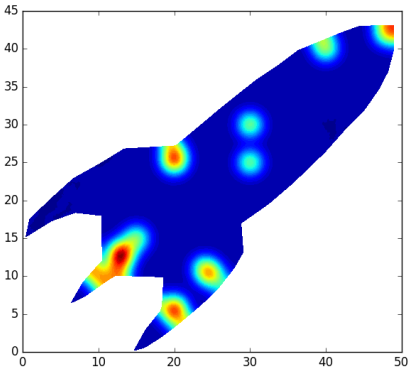
(a) Solution with homogeneous Neumann condition with parameters in (22) at $t = 0$.



(b) Solution with homogeneous Neumann condition with parameters in (22) at $t = 0.25$.



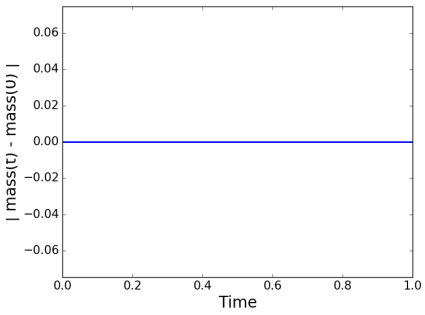
(c) Solution with homogeneous Neumann condition with parameters in (22) at $t = 0.5$.



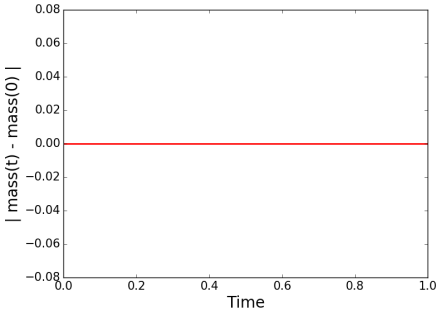
(d) Solution with homogeneous Neumann condition with parameters in (22) at $t = 0.75$.

Figure 8: Solution of the homogeneous Neumann boundary condition problem for the domain Ω with the shape of a "Rocket" with parameters (22) at distinct times in the interval $t = [0, 1]$.

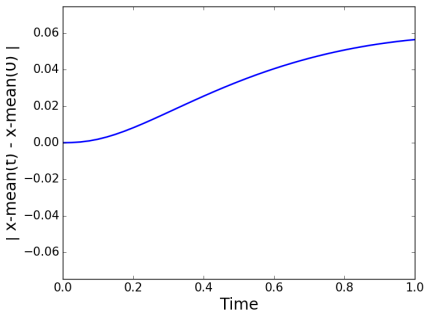
We present the change of mass total mass $\Delta m(t) = |m(t) - m(0)|$ and the averages $\Delta\mu_i(t) = |\mu_i(t) - \mu_i(0)|$ for $i = 1, 2$ over the time evolution of the solution, obtaining the results in Figure 9.



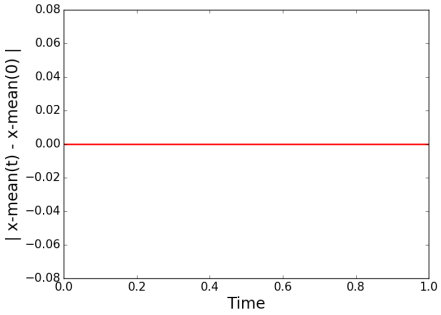
(a) Change in total mass Δm for the numerical solution with homogeneous Neumann boundary condition.



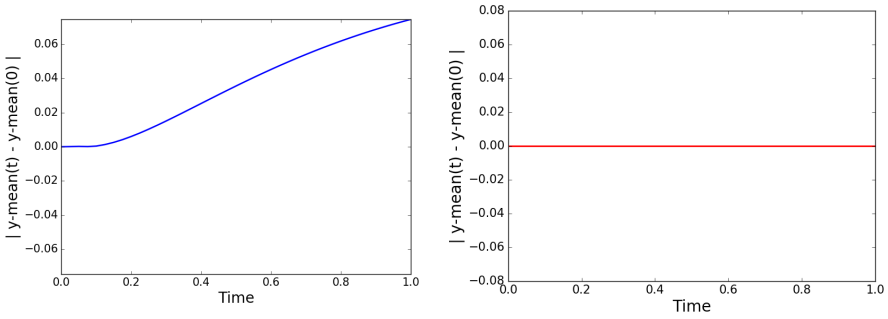
(b) Change in total mass Δm for the numerical solution with the new boundary conditions.



(c) Change in the x-mean $\Delta \mu_x$ for the numerical solution with homogeneous Neumann boundary condition.



(d) Change in the x-mean $\Delta \mu_x$ for the numerical solution with the new set of boundary conditions.



(e) Change in y-mean $\Delta\mu_y$ for the numerical solution with homogeneous Neumann boundary condition.

(f) Change in y-mean $\Delta\mu_y$ for the numerical solution with the new boundary conditions.

Figure 9: Plots of the evolution of the change in total mass Δm, $x-$ mean $\Delta\mu_x$, and $y-$mean $\Delta\mu_y$, for the density estimation with homogeneous Neumann and new set of boundary conditions for the problem (19) for the domain Ω with the shape of a "Rocket" and parameters in (22) in the interval $t \in [0, 1]$.

4. Conclusions

We designed a method to run the diffusion equation for estimation purposes in which we can preserve some moments of the estimated density. This is achieve by designing appropriate boundary conditions or boundary restrictions for the estimated density. We impose this restrictions during the time evolutions by using Lagrange multipliers. We presented numerical encouraging numerical experiment to verify the fitness of the new set of boundary conditions. Our numerical results motivate more theoretical analysis concerning the new set of boundary conditions which are non-local. These and other practical aspects as the selection of the correct stopping time for the new set of boundary restriction is under investigation. We also observed from our numerical experiments that, in the estimation using moments preserving boundary conditions, diffusion patterns are affected and therefore they have some differences when comparing with the estimation by using homogeneous Neumann boundary conditions. This is particularly noticeable with large times t and in certain initial conditions for rectangular domains. The estimation using the new boundary conditions in rectangular geometries gives no significant change in the density compared to estimation using homogeneous Neumann boundary conditions; nevertheless, the initial $x-$mean, $y-$mean and total mass is conserved in the whole estimation process. There is a significant difference in the estimation using the new boundary conditions in non-regular geometries introducing a form of averaging in the density estimation in all the domain compared to the case

of estimation using homogeneous Neumann boundary conditions.

References

[1] *13. Poisson equation with pure Neumann boundary conditions*, The FEniCS Project, https://fenicsproject.org/olddocs/dolfin/2016.2.0/python/demo/documented/neumann-poisson/python/documentation.html.

[2] *2D Density Estimation via Diffusion with FEM-Neumann Boundary Conditions. Patarroyo K.*, Available online, https://goo.gl/Qlpruk.

[3] *C. Geuzaine and J. -F. Remacle*, Gmsh (C), (1997-2017). https://gmsh.info/.

[4] *Rocket Clipart #18923*, WikiClipArt, https://wikiclipart.com/rocket-clipart_18923/.

[5] *Cadcorp Spatial Information System® (Cadcorp SIS®)*, [A spatial statistics program for the analysis of crime incident locations], 2017, https://www.icpsr.umich.edu/CrimeStat/ CrimeStat®.

[6] *National Institute of Justice*, Hot Spot Mapping Using KDE, 2017, https://www.cadcorp.com/about-us/.

[7] M. S. Alnaes, J. Hake J. Blechta, B. Kehlet A. Johansson, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, *The FEniCS Project Version 1.5. Archive of Numerical Software*, https://fenicsproject.org/ **3** (2015).

[8] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, *Kernel density estimation via Diffusion*, Ann. Statist **38** (2010), no. 5, 29162957.

[9] P. Chaudhuri and J. S. Marron, *Scale space view of of curve estimation*, Ann. Statist., 2000, 28 408428. MR1790003.

[10] M. S. Gerber, *Predicting crime using Twitter and kernel density estimation*, Decision Support Systems **61** (2014), no. 115, 125.

[11] E. L. Glaeser and B. Sacerdote, *Why is there more crime in cities?*, National Bureau of Economic Research Working Paper 4530, 1996.

[12] F. Gómez, A. Torres, J. Galvis, J. Camargo, and O. Martinez, *Hotspot mapping for Perception of Security. Smart Cities Conference (ISC2)*, IEEE International, 2016.

[13] M. G. Larson and F. Bengzon, *The Finite Element Method: Theory, Implementation, and Applications*, Springer-Verlag Italia, Milano, 2013.

[14] N. Memon, J. D. Farley, D. L. Hicks, T. Rosenorn, and (Eds.), *Mathematical Methods in Counterterrorism*, Springer, 2009 Edition.

[15] Y. K. Patarroyo, *Mean conservation for density estimation via diffusion using the finite element method*, Submitted to Boletín de Matemáticas (2017), https://arxiv.org/pdf/1702.07962.pdf.

[16] S. Salsa, *Partial differential equations in action from modelling to theory*, Springer-Verlag Italia, Milano., 2008.

[17] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, London., 1986.

[18] S. Tench, H. Fry, and P. Gill, *Spatio-temporal patterns of IED usage by the Provisional Irish Republican Army*, European Journal of Applied Mathematics **27** (2016), no. 3, 377–402.

A. Implementation of the Lagrange multiplier method in FeniCS

We make use of the FeniCS method for Lagrange Multipliers [1], were we consider a mixed (product) space W from two separate spaces V_h and $\mathbb{R}^3$.

$$W = \{(v, \xi_0, \xi_1, \xi_2) \text{ such that } v \in V_h, (\xi_0, \xi_1, \xi_2) \in \mathbb{R}^3\}.$$

In this formulation for the three boundary restrictions (4) and (5) we introduce one new real test number ξ and one Lagrange multiplier λ. Hence our modified Galerkin formulation yields, given $\left(u_h^0, \lambda_0^0, \lambda_1^0, \lambda_2^0\right)$, find $\left(u_h^1, \lambda_0^1, \lambda_1^1, \lambda_2^1\right)$, $\left(u_h^2, \lambda_0^2, \lambda_1^2, \lambda_2^2\right), \ldots \in W$ such that,

$$a\left\{\left(u_h^n, \lambda_0^n, \lambda_1^n, \lambda_2^n\right), (v, \xi_0, \xi_1, \xi_2)\right\} = L\left\{(v, \xi_0, \xi_1, \xi_2)\right\} \quad \forall\, (v, \xi_0, \xi_1, \xi_2) \in W, \tag{23}$$

where,

$$a\left\{\left(u_h^n, \lambda_0^n, \lambda_1^n, \lambda_2^n\right), (v, \xi_0, \xi_1, \xi_2)\right\} = a_1 + a_2 + a_3,$$

$$a_1 = \int_\Omega \left[u_h^{n+1}(\mathbf{x})v(\mathbf{x}) - \nabla_1 u_h^{n+1}(\mathbf{x}, t)\xi_1 - \nabla_2 u_h^{n+1}(\mathbf{x}, t)\xi_2 - \nabla_1 v(\mathbf{x}, t)\lambda_1^n - \nabla_2 v(\mathbf{x}, t)\lambda_2^n\right] d\mathbf{x},$$

$$a_2 = \Delta t \int_\Omega \nabla u_h^{n+1}(\mathbf{x}) \nabla v(\mathbf{x}) d\mathbf{x},$$

$$a_3 = \int_{\partial\Omega} \left\{ \frac{\partial u_h^{n+1}(\mathbf{x})}{\partial \boldsymbol{\nu}} \left[-\Delta t v(\mathbf{x}) + \xi_0 + x_1\xi_1 + x_2\xi_2\right] + \frac{\partial v(\mathbf{x})}{\partial \boldsymbol{\nu}} \left[\lambda_0^n, + x_1\lambda_1^n + x_2\lambda_2^n\right] \right\} d\sigma,$$

$$L(v) = \int_\Omega u_h^n(\mathbf{x}) v(\mathbf{x}) d\mathbf{x}.$$

Hence imposing the three boundary conditions (4) and (5) while solving the Galerkin formulation (23) (this is done automatically by FeniCS), yield an

equivalent linear system as the matrix formulation in the augmented linear system (19).

The resulting FeniCS codes area available at `https://goo.gl/8yW8DK`.