HOKKAIDO UNIVERSITY

Title	On behavior of signs for the heat equation and a diffusion method for data separation
Author(s)	Giga, Mi-Ho; GIGA, Yoshikazu; Ohtsuka, Takeshi et al.
Citation	Hokkaido University Preprint Series in Mathematics, 998, 1-23
Issue Date	2012-02-13
DOI	https://doi.org/10.14943/84145
Doc URL	https://hdl.handle.net/2115/69804
Type	departmental bulletin paper
File Information	re998.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

On behavior of signs for the heat equation and a diffusion method for data separation

Mi-Ho Giga

Graduate School of Mathematical Sciences
University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan
mihogiga@ms.u-tokyo.ac.jp

Yoshikazu Giga[1]

Graduate School of Mathematical Sciences
University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan
and
Department of Mathematics, Faculty of Sciences
King Abdulaziz University
P. O. Box 80203
Jeddah 21589, Saudi Arabia
labgiga@ms.u-tokyo.ac.jp

Takeshi Ohtsuka[2]

Division of Mathematical Sciences
Graduate School of Engineering
Gunma University
4-2 Aramaki-machi, Maebashi City, Gunma 371-8510, Japan
ohtsuka106@gmail.com

Noriaki Umeda

Graduate School of Mathematical Sciences
University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan
umeda@ms.u-tokyo.ac.jp

[1]The author is supported by the Japan Society for the Promotion of Science (JSPS) through grant Kiban (S) 21224001.

[2]The author is supported by the Japan Society for the Promotion of Science (JSPS) through Grant-in-Aid for Young Scientists (B) 22740109.

Abstract

Consider the solution $u(x,t)$ of the heat equation with initial data u_0. The diffusive sign $S_D[u_0](x)$ is defined by the limit of sign of $u(x,t)$ as $t \to 0$. A sufficient condition for $x \in \mathbf{R}^d$ and u_0 such that $S_D[u_0](x)$ is well-defined is given. A few examples of u_0 violating and fulfilling this condition are given. It turns out that this diffusive sign is also related to variational problem whose energy is the Dirichlet energy with a fidelty term. If initial data is a difference of characteristic function of two disjoint sets, it turns out that the boundary of the set $S_D[u_0](x) = 1$ (or -1) is roughly an equi-distance hypersurface from A and B and this gives a separation of two data sets.

1 Introduction

We consider a simple Cauchy problem for the heat equation in $\mathbf{R}^d$ ($d \geq 1$) with a real-valued bounded (measurable) initial data u_0 of the form

$$u_t - \Delta u = 0 \quad \text{in} \quad \mathbf{R}^d \times (0,\infty) \tag{1.1}$$

$$u\,|_{t=0} = u_0. \tag{1.2}$$

The unique bounded solution u is known (see e.g. [W]) to be represented by the Gaussian kernel G_t of the form

$$u\,(x,t) = \int_{\mathbf{R}^d} G_t(x-y)\,u_0\,(y)\,dy = (G_t * u_0)\,(x) \tag{1.3}$$

with $G_t(x) = (4\pi t)^{-d/2} \exp\left(-|x|^2/4t\right)$. We are interested in the behavior of sign of u as t tends to zero.

We set

$$S_D[u_0]\,(x) = \lim_{t \downarrow 0} \text{ sgn } u\,(x,t) \tag{1.4}$$

and call it the *diffusive sign* (by the heat equation) of u_0 at x, where we use the convention that

$$\text{sgn } a = \begin{cases} 1, & a > 0, \\ -1, & a < 0, \\ 0, & a = 0. \end{cases}$$

If u_0 is continuous at $\hat{x}$ and sgn $u_0(\hat{x}) \neq 0$, the diffusive sign is well-defined at $\hat{x}$ and agrees with sgn $u_0(\hat{x})$ since $u\,(x,t)$ is continuous at $(\hat{x},0)$; see e.g. [GGS]. However, if $u_0(\hat{x}) = 0$, the diffusive sign may not be well-defined even if u_0 is continuous near $\hat{x}$. We show this phenomenon by giving explicit examples where $u\,(\hat{x},t)$ changes its sign infinitely many times as t tends to zero (Lemma 2.2 and Theorem 2.3).

Our main goal of this paper is to give a sufficient condition for u_0 so that $S_D[u_0](x)$ is well-defined for a given point x. In one-dimensional problem this is related to the *number of changes of sign* which is also called the "number of zeros" in the literature. Let $Z[u_0]$ be the supremum over all k such that there exists $-\infty < x_0 < x_1 < \cdots < x_k < \infty$ with

$$u_0(x_i)\, u_0(x_{i+1}) < 0 \quad (i = 0, 1, \ldots, k-1).$$

(If there is no such k, we set $Z[u_0] = 0$.) If one restricts x_i's in a fixed open interval I we write $Z_I[u_0]$ instead of $Z[u_0]$. The quantity $Z_I(u_0)$ is the number of changes of sign in I. We say $Z[u_0]$ is locally finite if $Z_I(u_0)$ is finite for all bounded open interval I. If u_0 is bounded, piecewise continuous, we shall show that $S_D[u_0](\hat{x})$ exists for $\hat{x}$ when $Z[\bar{u}_0]$ is (locally) finite for

$$\bar{u}_0(x) = \big\{\, u_0(\hat{x}+x) + u_0(\hat{x}-x) \,\big\}/2 \tag{1.5}$$

provided that u_0 is continuous at $\hat{x}$ with $u_0(\hat{x}) = 0$ (Theorem 2.1). For a higher dimensional case one should replace $\bar{u}_0$ by

$$\bar{u}_0(r) = \int\limits_{|\omega|=1} u_0\left(\hat{x} + |r|\,\omega\right) \mathrm{d}\mathcal{H}^{d-1}, \tag{1.6}$$

where $\mathcal{H}^{d-1}$ denotes $d-1$ dimensional Hausdorff measure so that $\mathrm{d}\mathcal{H}^{d-1}$ is the surface element (Theorem 2.4). These assertions can be proved by a simple application of the strong maximum principle [PW]. Under this setting one is able to prove that the set of x when $S_D[u_0](x) = 0$ is a codimension one set, so it is negligible in the sense of the Lebesgue measure. This means the zero set of the diffusive sign is thin even if the original zero set of u_0 has an interior.

The diffusive sign is related to the asymptotic sign for a problem of deblurring images. For a given gray scale image u_0 one way to recover the image is to minimize a strictly convex variational problem

$$\frac{1}{2}\int_{\mathbf{R}^d} |\nabla v|^2 \mathrm{d}x + \frac{\lambda}{2}\int_{\mathbf{R}^d} |v - u_0|^2 \mathrm{d}x, \tag{1.7}$$

where $\lambda > 0$ is the fidelty constant. If v^λ is the unique (H^1) minimizer of (1.7), then v^λ solves the Euler-Lagrange equation of the form

$$-\Delta v + \lambda v = \lambda u_0 \quad \text{in} \quad \mathbf{R}^d. \tag{1.8}$$

We define the *asymptotic sign* of u_0 at x of the form

$$S_a[u_0](x) = \lim_{\lambda\to\infty} \operatorname{sgn} v^\lambda(x). \tag{1.9}$$

The large fidelty formally corresponds to small time in the heat equation. In fact, when one approximates the solution of the heat equation by a fully implicit finite

difference approximation in time, it is interpreted as an Euler-Lagrange equation of the variational problem. So we expect that

$$S_a[u_0](x) = S_D[u_0](x)$$

provided that S_D is well-defined. Indeed, we shall prove it rigorously by writing the Newton potential by a heat semigroup (Theorem 3.1).

In [ROF] the total variation is used in (1.7) instead of the Dirichlet energy for a recovery of blurred image. The idea is to minimize

$$\int_{\mathbf{R}^d} |\nabla v| \, \mathrm{d}x + \frac{\lambda}{2} \int_{\mathbf{R}^d} |v - u_0|^2 \mathrm{d}x \tag{1.10}$$

instead of (1.7). One is able to define

$$S_t[u_0](x) = \lim_{\lambda \downarrow 0} \operatorname{sgn} v^\lambda(x).$$

However, as it turns out this is quite different from S_a or S_D because the speed of diffusion is very slow. The set where diffusive sign is zero is rather thin even if the zero set of u_0 has an interior while the set of zeros of $S_t[u_0]$ may have an interior. We shall see this phenomenon by an example (Theorem 3.3).

We shall apply this diffusive method to separate sets of data. Suppose that each point of $\mathbf{R}^d$ fulfills either property P or Q (with $P \cap Q = \emptyset$) except very thin set. However, we only know that in some subset A of $\mathbf{R}^d$ the property P is fullfilled and in some subset B ($A \cap B = \emptyset$) of $\mathbf{R}^d$ the property Q is fulfilled. We would like to classify other point whether it fulfills the property P or Q in a reasonable way. Usually, people try to find a straight line (or a simple curve) to divide $\mathbf{R}^2$ into two sets so that A belongs to one side of the line and B belongs to another side of the line. The line is taken so that the distance from this line to a closest point of A and B is the same and that this quantity is maximized by taking a suitable normal direction of the line. (In a higher dimensional space the line should be of course a hyperplane.) This is a simple example of support vector machines [CST], [Std] and it is widely used for data separation. This separation line is called a maximal margin classifier [Std, 22.3.1]. We propose here to use the heat equation to find a separation curve which is interpreted as an example of a geometric diffusion approach explained in [CLLMNWZ].

We set

$$u_0(x) = \chi_A(x) - \chi_B(x) \tag{1.11}$$

where χ_A is the characteristic function of A, i.e. $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \neq A$. We implicitly assume that A and B are Lebesgue measurable. We propose to classify a point of $\mathbf{R}^d$ by using the diffusive sign $S_D[u_0]$. We set

$$A^\sharp = \big\{ x \in \mathbf{R}^d \,|\, S_D[u_0](x) = 1 \big\},$$
$$B^\sharp = \big\{ x \in \mathbf{R}^d \,|\, S_D[u_0](x) = -1 \big\}.$$

We expect that x has the property P if $x \in A^\sharp$ and it has the property Q if $x \in B^\sharp$. The complement set of $A^\sharp \cup B^\sharp$ in $\mathbf{R}^d$ is an analytic variety if u_0 fulfills our sufficient condition mentioned before. However, this does not imply that the boundary of $A^\sharp$ has a finite perimeter. It turns out that $\partial A^\sharp = \partial B^\sharp$ is an equi-distance hypersurface from A and B i.e. the set x of point where $d_e(x, A) = d_e(x, B)$. Here d_e denotes the essential distance (Theorem 4.1). Note that even if we replace u_0 by $\chi_A - c\chi_B$ with $c > 0$, the separation hypersurface is the same. The separation hypersurface does not depend on the ratio of size $|A|/|B|$ of A and B, where $|A|$ denotes the Lebesgue measure of A.

We also give a few numerical test to draw a separation curve $\partial A^\sharp$. In [BF] instead of using (1.10) Ginzburg-Landau type energy is proposed

$$\frac{1}{2}\int_{\mathbf{R}^d}\left\{\varepsilon\,|\nabla v|^2 + \frac{1}{\varepsilon}\big(\,|v|^2 - 1\big)^2\right\}\mathrm{d}x + \frac{\lambda}{2}\int_{\mathbf{R}^d}|v - u_0|^2\mathrm{d}x. \tag{1.12}$$

It is essentially known that the Gamma limit as $\varepsilon \to 0$ of (1.12) is (1.10) (if one puts a multiple constant $4/3$ in front of $|\nabla v|$ in (1.10)); see e.g. [MM], [S]. Compared with (1.7) this variational problem emphasizes sign very much. Using (1.12), the authors of [BF] separates several data sets on graphs as well as $\mathbf{R}^d$. It is not clear whether or not our separation by $A^\sharp$ and $B^\sharp$ is the same as theirs. We shall give several speculations in this paper.

2 Sign of a solution of the heat equation

We give a sufficient condition for u_0 so that $S_D[u_0](x)$ is well-defined. We start with one-dimensional problem.

Theorem 2.1. *Assume that u_0 is a (real-valued) bounded measurable function in $\mathbf{R}$ and that u_0 is piecewise continuous (with possibly countably many discoutinuities having at most finitely many accumulation points) and at discontinuities either left or right continuous. Assume that u_0 is continuous at $\hat{x}$ and $u_0(\hat{x}) = 0$. If the number of changes of sign $Z[\bar{u}_0]$ of*

$$\bar{u}_0(x) = \big\{\,u_0(\hat{x} + x) + u_0(\hat{x} - x)\big\}/2$$

is (locally) finite, then $S_D[u_0](\hat{x})$ is well-defined. In other words, there exists $\hat{t} = \hat{t}(\hat{x}) > 0$ such that $u\,(\hat{x}, t)$ has the same sign for $0 < t < \hat{t}\,(\hat{x})$, where u is the solution of (1.1)–(1.2) (i.e. u is given by (1.3)). Moreover, $S_D[u_0](\hat{x}) = 0$ if and only if u_0 is odd with respect to $\hat{x}$. Thus the totality of such $\hat{x}$ is locally finite.

Proof. We may assume that $u_0 \not\equiv 0$. We symmetrize the problem by considering

$$\bar{u}\,(x, t) = \big\{\,u\,(\hat{x} + x, t) + u\,(\hat{x} - x, t)\big\}/2.$$

Evidently, $\bar{u}$ solves the heat equation with initial data $\bar{u}_0$. Assume that $\bar{u}_0 \not\equiv 0$. Since $Z[\bar{u}_0]$ is (locally) finite and $\bar{u}_0$ is even in x, there is an interval $(-\gamma_0, \gamma_0)$ such that u_0 is continuous near $\gamma_0 > 0$ and that

$$\text{(i)} \quad \bar{u}_0 \geq 0 \quad \text{on} \quad (-\gamma_0, \gamma_0) \quad \text{and} \quad \bar{u}_0(\gamma_0) > 0,$$

or

$$\text{(ii)} \quad \bar{u}_0 \leq 0 \quad \text{on} \quad (-\gamma_0, \gamma_0) \quad \text{and} \quad \bar{u}_0(\gamma_0) < 0.$$

Since both cases can be treated similarly, we consider the first case. Since $\bar{u}\,(x,t)$ is continuous at $x = \gamma_0$, $t = 0$ (see e.g. [GGS, Chapter 1]), we may assume that there is $\hat{t} > 0$ such that $\bar{u}\,(x,t) > 0$ on $\{\pm\gamma_0\} \times [0, \hat{t}\,)$. By the strong maximum principle [PW] $\bar{u}\,(x,t) > 0$ in $[-\gamma_0, \gamma_0] \times (0, \hat{t}\,)$. Thus $\bar{u}\,(0,t) > 0$ for $t \in (0, \hat{t}\,)$. This implies that $u\,(\hat{x}, t) > 0$ for $t \in (0, \hat{t}\,)$ so $S_D[u_0](\hat{x})$ is well-defined and equals one. A symmetric argument yields that $S_D[u_0](\hat{x}) = -1$ for case (ii).

If $\bar{u} \equiv 0$, then $u\,(x,t)$ is odd with respect to $\hat{x}$. Since u_0 is assumed to be continuous at $\hat{x}$, $u\,(x,t)$ is continuous at $\hat{x}$ and $t = 0$. Since $u\,(\hat{x} + x, t)$ is odd, $u\,(\hat{x} + 0, t) = 0$ for sufficiently small t. This means that $S_D[u_0](\hat{x})$ is well-defined and equals zero. □

Remark 1. (i) The idea using symmetrization is used in many times to prove qualitative properties of solutions of semilinear heat equations. For example, Chen and Matano [CM] proved that the maximum point of $w\,(x,t)$ ($x \in \mathbf{R}$) converges to a unique point as t tends to the blow up time when w solves $w_t = \Delta w + w^p$ ($p > 1$) by considering the symmetrization $\bar{u}$.
(ii) There are several studies about the number of zeros or the number of changes of sign for a solution of a one-dimensional general linear parabolic equation of second-order. It is known that this number is nonincreasing in time. This type of result goes back to Nickel [N] and rediscovered by Matano [M] and Henry [H], where they proved the nonincrease of the number of changes of sign by the strong maximum principle. For nonincrease of the number of zeros the reader is referred to an article by Angenent [A] where it is also analyzed a way of merging zero when the number of zero actually decreases. This paper appeals to an asymtotic analysis near a point of interest by introducing similarity variables (cf. [GGS]).

If the number of changes of sign $Z[\bar{u}_0]$ is infinite, there is a chance that $S_D[\bar{u}_0](\hat{x})$ does not exist. We shall give an explicit example of such u_0. In fact, in our example $u\,(\hat{x}, t)$ is oscillatory in time and it changes signs $+1$ to -1 infinitely many times as $t \downarrow 0$.

Lemma 2.2. *Let U_k be a function of the form*

$$U_k(x) = \begin{cases} \left(-\dfrac{1}{k}\right)^n x, & x \in \left[\dfrac{2^n+1}{2^n}, \dfrac{2^{n-1}+1}{2^{n-1}}\right) \quad \text{with} \quad n \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Let u be the solution of (1.1)–(1.2) with initial data $u_0 = U_k$. Then $u(0,t)$ changes its sign infinitely many times from 1 to -1 as $t \downarrow 0$ provided that $k \geq 8$.

Proof. By a direct calculation we have

$$\begin{aligned} u(0,t) &= \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-y^2/4t} u_0(y) dy \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-z^2} u_0(\sqrt{4t}z) dz. \end{aligned}$$

By the definition of U_k

$$\begin{aligned} u(0,t) &= \frac{1}{\sqrt{\pi}} \sum_{n=1}^{\infty} \int_{(2^n+1)/(2^n\sqrt{4t})}^{(2^{n-1}+1)/(2^{n-1}\sqrt{4t})} \sqrt{4t} z \left(-\frac{1}{k}\right)^n e^{-z^2} dz \\ &= -\frac{\sqrt{4t}}{2\sqrt{\pi}} \sum_{n=1}^{\infty} \int_{(2^n+1)/(2^n\sqrt{4t})}^{(2^{n-1}+1)/(2^{n-1}\sqrt{4t})} \left\{ -2z \left(-\frac{1}{k}\right)^n e^{-z^2} \right\} dz \\ &= -\sqrt{\frac{t}{\pi}} \sum_{n=1}^{\infty} \left(-\frac{1}{k}\right)^n \left(e^{-\left((2^n+1)/(2^{n+1}\sqrt{t})\right)^2} - e^{-\left((2^{n-1}+1)/(2^n\sqrt{t})\right)^2} \right) \\ &= \sqrt{\frac{t}{\pi}} \left[e^{-1/t} + \sum_{n=1}^{\infty} \frac{k+1}{k} \left(-\frac{1}{k}\right)^n e^{-\left((2^n+1)/(2^{n+1})\right)^2/t} \right] \\ &= \sqrt{\frac{t}{\pi}} \left[e^{-1/t} + \sum_{n=1}^{\infty} \frac{k+1}{k} (-1)^n \exp\left\{ -\left(\frac{2^n+1}{2^{n+1}}\right)^2 \frac{1}{t} - n \log k \right\} \right]. \end{aligned}$$

We set

$$a_0 = e^{-1/t}, \quad a_n = \frac{2(k+1)}{k} (-1)^n \exp\left\{ -\left(\frac{2^n+1}{2^{n+1}}\right)^2 \frac{1}{t} - n \log k \right\} (n \in \mathbf{N}).$$

so that

$$u(0,t) = \sqrt{\frac{t}{\pi}} \left[a_0(t) + \sum_{n=1}^{\infty} a_n(t) \right].$$

Note that $\operatorname{sgn} a_n(t) = (-1)^n$.

It is clear that

$$\left|a_0(t)\right| \leq \left|a_1(t)\right| \quad \text{for} \quad t \leq t_0 \equiv \frac{7}{16} \left(\log \frac{k^2}{k+1} \right)^{-1}.$$

If $\left|a_n(t)\right| = \left|a_{n+1}(t)\right|$ for $n \geq 1$, then we obtain

$$\left(\frac{2^n+1}{2^{n+1}}\right)^2 \frac{1}{t} + n \log k = \left(\frac{2^{n+1}+1}{2^{n+2}}\right)^2 \frac{1}{t} + (n+1) \log k.$$

By a direct calculation we see that

$$t = t_n \equiv \frac{2^{n+2}+3}{2^{2n+4}\log k}\ .$$

It is clear that $t_n > t_m$ for $n < m$, and

$$\begin{aligned} \big|a_n(t)\big| &< \big|a_{n+1}(t)\big| \quad \text{for} \quad t < t_n, \\ \big|a_n(t)\big| &> \big|a_{n+1}(t)\big| \quad \text{for} \quad t > t_n. \end{aligned}$$

For $t \in (t_{n+1}, t_n)$ we see that

$$\big|a_{n+1}(t)\big| > \big|a_n(t)\big| > \cdots > \big|a_1(t)\big| > \big|a_0(t)\big| \text{ and } \big|a_{n+1}(t)\big| > \big|a_{n+2}(t)\big| > \cdots . \quad (2.1)$$

By the same argument we obtain

$$\begin{aligned} t = t_n^+ &\equiv \frac{2^{n+2}+3}{2^{2n+4}(\log k - \log 2)} \quad \text{when } \big|a_n(t)\big| = 2\big|a_{n+1}(t)\big|, \\ t = t_n^- &\equiv \frac{2^{n+2}+3}{2^{2n+4}(\log k + \log 2)} \quad \text{when } 2\big|a_n(t)\big| = \big|a_{n+1}(t)\big|. \end{aligned}$$

It is clear that $t_n^- < t_n < t_n^+$, and

$$\begin{aligned} \big|a_n(t)\big| &< 2\big|a_{n+1}(t)\big| \quad \text{for} \quad t < t_n^+, \\ \big|a_n(t)\big| &> 2\big|a_{n+1}(t)\big| \quad \text{for} \quad t > t_n^+, \\ 2\big|a_n(t)\big| &< \big|a_{n+1}(t)\big| \quad \text{for} \quad t < t_n^-, \\ 2\big|a_n(t)\big| &> \big|a_{n+1}(t)\big| \quad \text{for} \quad t > t_n^-. \end{aligned}$$

It is clear that $t_1^- < t_0$ for $k \geq 3$. If $t < t_n^-$ and $t > t_{n+1}^+$, then $|a_{n+1}| > 2|a_n|$ and $|a_{n+1}| > 2|a_{n+2}|$. This implies $\big|a_{n+1}(t)\big| > \big|a_n(t)\big| + \big|a_{n+2}(t)\big|$. To guarantee the existence of such t we need $t_n^- > t_{n+1}^+$ or

$$\frac{2^{n+2}+3}{2^{2n+4}(\log k + \log 2)} > \frac{2^{n+3}+3}{2^{2n+6}(\log k - \log 2)}. \quad (2.2)$$

This condition is fulfilled if k satisfies

$$\log k > \frac{3\cdot 2^{n+3}+15}{2^{n+3}+9}\log 2.$$

If

$$\log k \geq 3\log 2 \quad \text{or} \quad k \geq 8,$$

then this inequality (2.2) holds for all $n = 1, 2, \ldots$. Thus if $k \geq 8$, then $t_{n+1}^+ < t_n^-$. Under the condition $k \geq 8$ we see that

$$\operatorname{sgn}\big(a_n(t) + a_{n+1}(t) + a_{n+2}(t)\big) = (-1)^{n+1} \quad \text{for} \quad t \in (t_{n+1}^+, t_n^-).$$

From the fact $t_{n+1} < t_{n+1}^- < t_n^+ < t_n$ and (2.1) it follows that

$$\operatorname{sgn}\big(a_{n+2j+1}(t) + a_{n+2j+2}(t)\big) = (-1)^{n+1} \quad \text{for} \quad j = 1, 2, \dots, \ t \in (t_{n+1}^+, t_n^-),$$
$$\operatorname{sgn}\big(a_{n-2j+1}(t) + a_{n-2j}(t)\big) = (-1)^{n+1} \quad \text{for} \quad j = 1, 2, \dots, [n/2], \ t \in (t_{n+1}^+, t_n^-),$$

where $[a]$ is a largest integer less than or equal to a. We thus obtain

$$\operatorname{sgn} u\,(0,t) = (-1)^{n+1} \quad \text{for} \quad t \in (t_{n+1}^+, t_n^-).$$

Then the solution of (1.1) with $u_0 = U_k$ for $k \geq 8$ has infinitely many zeros of t at $x = 0$.

□

We are tempted to say that it is enough to assume that u_0 itself has at most finitely many changes of sign to guarantee that $S_D[u_0](x)$ is well-defined. Unfortunately, this is not true in general. In fact, the following example just changes the sign once but $\bar{u}_0$ with respect to zero has infinite $Z[\bar{u}_0]$.

Theorem 2.3. *Assume that*

$$u_0(x) = \big(U_k(x)\big)_+ - \big(-U_k(-x)\big)_+$$

where $a_+ = \max(a, 0)$. *Let* u *be the solution of (1.1)–(1.2). Then* $u\,(0,t)$ *changes its sign infinitely many times from* 1 *to* -1 *as* $t \downarrow 0$ *provided that* $k \geq 8$.

Proof. Since, by symmetry of the Gaussian kernel, $u\,(0,t)$ is the same as Lemma 2.2, our assertion is already proved in Lemma 2.2. (Note that $Z[u_0] = 1$ while $Z[\bar{u}_0] = \infty$, where $\bar{u}_0(x) = \big(u_0(x) + u_0(-x)\big)/2$.) □

We give a sufficient condition for u_0 so that $S_D[u_0](x)$ is well-defined for multi-dimensional case ($d \geq 2$).

Theorem 2.4. *Assume that* u_0 *is a (real-valued) bounded measurable function in* $\boldsymbol{R}^d$. *Assume that* u_0 *is continuous at* $\hat{x}$ *and* $u_0(\hat{x}) = 0$. *Let* $\bar{u}_0$ *be the radial average around* $\hat{x}$ *defined by (1.6). Assume that* $\bar{u}_0$ *is piecewise continuous (with possibly countably many dicontinuities having at most finitely many accumulation points). Moreover, it is left or right continuous at discontinuities. If the number of changes of sign* $Z[\bar{u}_0]$ *of* $\bar{u}_0$ *is (locally) finite then* $S_D[u_0](\hat{x})$ *is well-defined. Moreover,* $S_D[u_0](\hat{x}) = 0$ *if and only if* $\bar{u}_0 \equiv 0$. *The totality of such* $\hat{x}$ *is included in an analytic variety of* $\mathbf{R}^d$ *so it has finite* $d-1$ *dimensional Hausdorff measure.*

Proof. We may assume $u_0 \not\equiv 0$. We study the radial average of the solution

$$\bar{u}\,(r,t) = \int\limits_{|\omega|=1} u\left(\hat{x} + |r|\omega\right) \mathrm{d}\mathcal{H}^{d-1}(\omega), \quad r \in \mathbf{R}.$$

Evidently, $\bar{u}$ is a radial solution of the heat equation (1.1) with initial data $\bar{u}_0$.

Assume that $\bar{u}_0 \not\equiv 0$. Then we proceed exactly as in Theorem 2.1 and observe that $\bar{u}\,(0,t) > 0$ for $t \in (0,\hat{t}\,)$ for the case (i), which implies that $u\,(\hat{x},t) > 0$ for $t \in (0,\hat{t}\,)$. The case (ii) is symmetric. In the case (i) $S_D[u_0](\hat{x}) = 1$ while in the case (ii) $S_D[u_0](\hat{x}) = -1$.

If $\bar{u}_0 \equiv 0$ so that $\bar{u} \equiv 0$. This in particular implies that $u\,(\hat{x},t) = 0$ for all $t > 0$. Thus $S_D[u_0](\hat{x}) = 0$. Moreover, the set

$$\Sigma = \left\{\, \hat{x} \in \mathbf{R}^d \;\middle|\; S_D[u_0](\hat{x}) = 0 \,\right\}$$

is contained in the set of zero of $u\,(x,t)$ for all t since $\bar{u}_0$ at $\hat{x}$ equals zero. Since $u\,(\cdot,t)$ is analytic in space for $t > 0$, Σ is included in an analytic variety, so it is of locally finite $d-1$ Hausdorff measure. □

Remark 2. Even if the heat equation is replaced by a general second-order parabolic equation with non-analytic coefficients it is known that the set of zeros of a solution at a fixed time has at most locally finite $d-1$ Hausdorff dimension [XYChen].

We next study what kind of initial data satisfies the locally finiteness of $Z[\bar{u}_0]$. For further references we say that v is a (locally) *finitely many sign-changing* function if $Z[v]$ is (locally) finite. Evidently, if $u_0(x_1)$ is real analytic like $\sin x_1$, then $\bar{u}_0$ is a locally finitely many sign-changing function. For data separation it is convenient to consider a characteristic function. Assume that A and B are a possibly countably many disjoint union of open intervals whose lengths are bounded from below by a some positive content and $A \cap B = \emptyset$. If one sets

$$u_0(x_1) = \chi_A(x_1) - c\chi_B(x_1)$$

with $c \in \mathbf{R}$, it is easy to set that $\bar{u}_0$ is a locally finitely many sign-changing function. More generally, if $\{A_k\}_{k=1}^{\infty}$ is a set of disjoint intervals with $\inf_k |A_k| > \infty$, then a function

$$f(x_1) = \sum_{k=1}^{\infty} c_k\, \chi_{A_k}(x_1) \tag{2.3}$$

with $c_k \in \mathbf{R}$ is a locally finitely many sign-changing function. If one considers $\bar{f}$, this is again of the form (2.3) with possible modification at locally finitely many points, which does not give any effect to define $Z[\bar{f}]$. Thus one is able to conclude a general statement for a piecewise constant function.

Theorem 2.5. *Assume that u_0 is of the form (2.3), i.e.,*

$$u_0(x) = \sum_{k=1}^{\infty} c_k\, \chi_{A_k}(x)$$

with $c_k \in \mathbf{R}$*, where* $\{A_k\}_{k=1}^{\infty}$ *is a disjoint family of intervals with* $\inf_k |A_k| > \infty$*. Then* $\bar{u}_0$ *defined by (1.5) is a locally finitely many sign-changing function for any* $\hat{x} \in \mathbf{R}$*.*

For a higher dimension setting the situation will be more involved. It is expected that if A_k has piecewise real analytic boundary then the radial average $\bar{u}_0$ of u_0 is a locally finitely many sign-changing function. We shall give a proof when A_k is a square in the plane $\mathbf{R}^d$ with $d = 2$.

Lemma 2.6. *Let* A *be a square of the form* $A = [a, a+1] \times [b, b+1]$ *with* $a \geq b > 0$*. Let* v_A *is the radial average of* χ_A*, i.e.,* $v_A = \bar{\chi}_A$ *in* $\mathbf{R}^2$*. Then*

$$2\pi\, v_A(r) = \begin{cases} 0, & r < r_1 \quad \text{or} \quad r \geq r_4, \\ \operatorname{Arc\,cos}\left(h\left(r, a, b\right)/r^2\right), & r_1 \leq r < r_2, \\ \operatorname{Arc\,cos}\left\{ \left(b^2 + b + \sqrt{r^2 - (b+1)^2}\,\sqrt{r^2 - b^2}\right) \,/\, r^2 \right\}, & r_2 \leq r < r_3, \\ \operatorname{Arc\,cos}\left(h\left(r, a+1, b+1\right)/r^2\right), & r_3 \leq r < r_4 \end{cases}$$

with $r_1 = \sqrt{a^2 + b^2}$, $r_2 = \sqrt{a^2 + (b+1)^2}$, $r_3 = \sqrt{(a+1)^2 + b^2}$, $r_4 = \sqrt{(a+1)^2 + (b+1)^2}$ *and* $h\left(r, a, b\right) = a\sqrt{r^2 - b^2} + b\sqrt{r^2 - a^2}$*.*

Moreover, the function (in the RHS of v_A*) in* $[r_2, r_3)$ *can be extended analytically in some neighborhood of* $[r_2, r_3]$ *while the square of the function in* $[r_2, r_3)$ *and* $[r_3, r_4]$ *can be extended analytically in some neighborhood of* $[r_1, r_2]$ *and* $[r_3, r_4]$ *respectively.*

Proof. The formula for v_A follows from a direct calculation. This is a piecewise analytic function. Since

$$\frac{\operatorname{Arc\,cos} x}{\sqrt{1 - x^2}} = \sum_{j=1}^{\infty} \frac{(1 - x^2)^{j-1}}{(2j-1)x^{2j-1}},$$

there exists a real analytic function $b_1(x)$ near 1 such that

$$\operatorname{Arc\,cos} x = b_1(x)\sqrt{1 - x} \quad \text{for} \quad x \in (1 - \delta, 1)$$

with any $\delta \in (0, 1)$. By a direct manipulation we observe that

$$2\pi\, v_A(r) = \begin{cases} 0, & r \leq r_1 \ \text{ or } \ r \geq r_4, \\ b_1(r)\sqrt{r - r_1}, & r_1 < r < r_2, \\ b_2(r), & r_2 \leq r < r_3, \\ b_3(r)\sqrt{r_4 - r}, & r_3 \leq r < r_4 \end{cases}$$

with b_i which is real analytic in a neighborhood of $[r_i, r_{i+1}]$ $(j = 1, 2, 3)$. □

Theorem 2.7. *Let A_k be a square of the form*

$$A_k = [a_k, a_k + 1] \times [b_k, b_k + 1]$$

for $a_k, b_k \in \mathbf{R}$ and $k = 1, \dots, m$. Set $u_0 = \sum_{k=1}^m c_k \chi_{A_k}$ with $c_k \in \mathbf{R}$. Then the radial average $v = \bar{u}_0(r)$ defined by (1.6) is a finitely many sign-changing function for any $\hat{x} \in \mathbf{R}^d$. Thus $S_D[u_0](x)$ is well defined for all $x \in \mathbf{R}^d$.

Proof. We first note that $v_A = v_B$ if we take

$$B = [\bar{a}, \bar{a} + 1] \times [\bar{b}, \bar{b} + 1],\ \bar{a} = \max(\tilde{a}, \tilde{b}),\ \bar{b} = \min(\tilde{a}, \tilde{b})$$

with $\tilde{a} = a$ if $a \geq 0$, $\tilde{a} = |a + 1|$ if $a < 0$ and same for $\tilde{b}$, where

$$A = [a, a + 1] \times [b, b + 1].$$

So we may assume that $a_k > b_k > 0$ to calculate $v = \Sigma_{k=1}^m c_k \chi_{A_k}$. By an explicit form of Lemma 2.6 v is at least continuous.

Let r_i $(i = 1, 2, 3, 4)$ be the singularity of $\bar{\chi}_A$ with $A = A_k$ defined in Lemma 2.6. We denote it by $r_{k,i}$ to clarify the dependence of k. By Lemma 2.6 our v is piecewise real analytic in $\mathbf{R}$ except a singular set

$$S = \left\{ r_{k,i} \;\middle|\; k = 1, \dots m,\ i = 1, 2, 3, 4 \right\}.$$

Let (p, q) be the maximal interval so that v is real analytic. By definition p and q is an element of S. By the identity theorem the number of zeros of v is finite in any compact sunset of (p, q). It remains to exclude the possibility that zeros accumulate at p or q. By Lemma 2.6 near p, v is of the form

$$v(r) = b(r)\sqrt{r - p} + c(r),$$

where b and c is an real analytic function near $r = p$. Thus

$$\big(v(r) - c(r)\big)^2 = b(r)^2(r - p).$$

If there is $\rho_j \downarrow p$ such that $v(\rho_j) = 0$, then by the identity theorem $c(r)^2 = b(r)^2(r-p)$ near $r = p$ since both sides are real analytic in a neighborhood of $r = p$. However, this is impossible since $c(r)$ is analytic near $r = p$. We thus observe that there is no accumulation point of zeros of v. Similarly, there is no accumulation of zeros of v to $r = q$. We have thus proved that v has finitely many zeros and continuous so it is a finitely many sign-changing function. (Note that we need not assume that A_k is mutually disjoint.)

By Theorem 2.4 we now conclude that $S_D[u_0](x)$ is well-defined for such u_0 at all $x \in \mathbf{R}^d$. □

Remark 3. In Theorem 2.7 we may replace finite sum of χ_{A_k}'s by an infinite sum $\sum_{k=1}^\infty c_k \chi_{A_k}$ provided that A_k is mutually disjoint and $\inf_k a_k,\ \inf b_k > 0$. The conclusion should be of course modified by replacing "finitely many" by "locally finitely many". This kind of remark is important if one consider a periodic setting.

3 Variational approach

We shall study the asymptotic sign related to the strictly convex variational problem (1.7), i.e.,

$$v \mapsto \frac{1}{2}\int_{\mathbf{R}^d} |\nabla v|^2 \mathrm{d}x + \frac{\lambda}{2}\int_{\mathbf{R}^d} |v-u_0|^2 \mathrm{d}x \tag{3.1}$$

with $\lambda > 0$. For a given $u_0 \in L^2(\mathbf{R}^d)$ there is a unique H^1-minimizer v^λ, which satisfies the elliptic equation (1.8)

$$(\lambda - \Delta)\, v = \lambda u_0 \quad \text{in} \quad \mathbf{R}^d \tag{3.2}$$

and $S_a[u_0](x) = \lim_{\lambda\to\infty} \big(\operatorname{sgn} v^\lambda(x)\big)$.

Theorem 3.1. *Assure that $u_0 \in L^2(\mathbf{R}^d) \cap L^\infty(\mathbf{R}^d)$ with $d \geq 1$ is continuous at $\hat{x}$ and $S_D[u_0](\hat{x})$ exists. Then $S_a[u_0](\hat{x})$ is well-defined and equals $S_D[u_0](\hat{x})$.*

Proof. The solution of (3.2) is written by use of the heat semigroup $e^{t\Delta} = G_t*$ of the form

$$v^\lambda = (\lambda - \Delta)^{-1}\lambda u_0 = \int_0^\infty e^{(\Delta-\lambda)\,t}(\lambda u_0)\,\mathrm{d}t. \tag{3.3}$$

Here

$$e^{(\Delta-\lambda)\,t} = e^{\Delta t}e^{-\lambda t} \quad \text{and} \quad e^{\Delta t}f = G_t * f.$$

Thus

$$v^\lambda(x) = \int_0^\infty \lambda e^{-\lambda t}(e^{\Delta t}u_0)(x)\,\mathrm{d}t.$$

Assume that $S_D[u_0](\hat{x}) = 1$ so that there is $\hat{t} > 0$ such that $(e^{t\Delta}u_0)(\hat{x}) > 0$ for $t \in (0, \hat{t}\,)$. Since $e^{t\Delta}u_0$ is bounded for all $t > 0$ (actually converges to zero uniformly in t as $t \to \infty$ [GGS]), we now apply the next lemma to conclude that $\operatorname{sgn} v^\lambda(\hat{x}) = 1$ for sufficiently large λ. Thus, $S_a[u_0](\hat{x}) = 1 = S_D[u_0](\hat{x})$. The case $S_D[u_0](\hat{x}) = -1$ can be treated in the same way.

It remains to prove the case $S_D[u_0](\hat{x}) = 0$. In this case $e^{t\Delta}u_0(\hat{x})$ is zero at least for a small t. However, since u is also analytic in time, this implies that $(e^{t\Delta}u_0)(\hat{x}) = 0$ for all $t > 0$. Thus

$$v^\lambda(\hat{x}) = \int_0^\infty \lambda e^{-\lambda t}(e^{t\Delta}u_0)(\hat{x})\,\mathrm{d}t = 0$$

so we conclude that $S_a[u_0](\hat{x}) = 0$. □

Lemma 3.2. *Let f be a bounded real-valued continuous function defined in $[0,\infty)$. Assume that there is $\hat{t} > 0$ such that $f(t) > 0$ for $t \in (0, \hat{t}\,)$. (Note that $f(0)$ may be zero.) Then*

$$a(\lambda) = \int_0^\infty \lambda e^{-\lambda t} f(t)\,\mathrm{d}t$$

is positive for a sufficiently large λ.

Proof. We divide the integral into two parts $(0, \hat{t}\,)$ and $(\hat{t}, \infty)$. We estimate

$$\left| \int_{\hat{t}}^{\infty} \lambda e^{-\lambda t} f(t)\, \mathrm{d}t \right| \leq \|f\|_{\infty} \int_{\hat{t}}^{\infty} \lambda e^{-\lambda t}\, \mathrm{d}t = \|f\|_{\infty} e^{-\lambda \hat{t}},$$

where $\|f\|_{\infty} = \sup \left\{ \left| f(t) \right| \;\middle|\; 0 \leq t < \infty \right\}$. The other part is estimated from below as

$$\int_0^{\hat{t}} \lambda e^{-\lambda t} f(t)\, \mathrm{d}t \geq \lambda e^{-\lambda \hat{t}} \int_0^{\hat{t}} f(t)\, \mathrm{d}t\,.$$

Thus

$$a(\lambda) \geq e^{-\lambda \hat{t}} \left(\lambda \int_0^{\hat{t}} f(t)\, \mathrm{d}t - \|f\|_{\infty} \right).$$

If λ is taken so that

$$\lambda > \|f\|_{\infty} \Big/ \int_0^{\hat{t}} f(t)\, \mathrm{d}t\,,$$

then $a(\lambda) > 0$. □

Remark 4. The equation (3.2) has a unique bounded solution if $u_0 \in L^{\infty}(\mathbf{R}^d)$ without assuming that $u_0 \in L^2(\mathbf{R}^d)$. The unique solution is given by (3.3).

We now study the energy (1.10) involving total variation. This does not diffuse the sign so $S_t[u_0]$ and $S_a[u_0]$ are quite different.

We consider a one-dimensional problem for (1.10) with $u_0 \in L^2(\mathbf{R}^d) \cap BV(\mathbf{R}^d)$. Then the unique minimizer v^{λ} for (1.10) fulfills the Euler-Lagrange equation

$$\begin{aligned} \eta_x - \lambda v = \lambda u_0 \quad &\text{for all} \quad x \in \mathbf{R} \backslash \Sigma_{\lambda}\,, \\ |\eta| \leq 1 \quad &\text{for all} \quad x \in \mathbf{R}\,, \\ \eta\,(x) = \pm 1 \quad &\text{for} \quad |v_x| - \text{a.e.}\,, \end{aligned}$$

where Σ_{λ} is the set that η is not differentiable. See [Ch] and [ACM]. This problem is studied in detail in [BFI].

Theorem 3.3. *Assume that u_0 is Lipschitz and supported in $[-M, M]$. Then the support of the minimizer v^{λ} is also contained in $[-M, M]$ for all $\lambda > 0$.*

Proof. As we know v^{λ} is also Lipschitz continuous since u_0 is Lipschitz continuous [Ch]. If $w \in L^2(\mathbf{R})$ and Lipschitz on $\mathbf{R}$, we have

$$\left(\sup_{x \geq m} \left| w(x) \right| \right)^3 \leq 3\, \|w\|^2_{L^2(x \geq m)} \|w_x\|_{\infty}\,. \tag{3.4}$$

Indeed there is a sequence $\{x_j\}_{j=1}^{\infty}$ such that $x_j \to \infty$ with $w\,(x_j) \to 0$. By a fundamental formula for the calculus we have

$$\left| w^3(m) - w^3(x_j) \right| \leq \int_m^{x_j} \left| (w^3)_x \right| \mathrm{d}x \leq \|w\|^2_{L^2(x \geq m)} \|w_x\|_{\infty}\,.$$

Sending $j \to \infty$ yields the desired estimate. Since v^λ is in L^2, by (3.4) we conclude that $v^\lambda(x) \to \infty$ as $|x| \to \infty$.

Suppose that v^λ is not zero for some $x_0 > M$. Since $v^\lambda(x) \to 0$ as $|x| \to \infty$,

$$\begin{aligned}\int_{\mathbf{R}} |\nabla v^\lambda| + \frac{\lambda}{2}\int_{\mathbf{R}} |v^\lambda - u_0|^2 \mathrm{d}x &\geq \int_{-\infty}^{x_0} |\nabla v^\lambda| + \frac{\lambda}{2}\int_{-\infty}^{x_0} |v^\lambda - u_0|^2 \mathrm{d}x + \left|v^\lambda(x_0)\right| \\ &= \int_{\mathbf{R}} |\nabla w^\lambda| + \frac{1}{2}\int_{\mathbf{R}} |w^\lambda - u_0|^2 \mathrm{d}x\end{aligned}$$

with $w^\lambda(x) = v^\lambda(x)$ for $x \leq x_0$ and $w^\lambda(x) = 0$ for $x > x_0$. Thus the energy of w^λ is smaller or equal to that of v^λ. Since the minimizer is unique, this is a contradiction so we conclude that $v^\lambda = 0$ for $x > M$. A symmetric argument yields that $v^\lambda = 0$ for $x \leq -M$ so that the support of v^λ is contained in $[-M, M]$. □

This is a quite different from the case of (1.7), where the total variation energy is replaced by the Dirichlet energy. Because of diffusion effect of (1.7) v^λ cannot be zero in a larger set for this problem, while the diffusion effect of total variation is limited since it is not strictly parabolic.

Example. If $u_0(x) = \big(1 - |x|\big)_+$, then $v^\lambda = 0$ for $|x| \geq 1$ so for all λ the number $S_t[u_0](x) = 0$ for $|x| \geq 1$. This is strikingly different since the asymptotic sign $S_a[u_0](x) = 1$ no matter how x is. (Note that u_0 is nonnegative.)

If we calculate the Euler-Lagrange equation, one can prove that

$$v^\lambda(x) = \min\left(1 - \sqrt{\frac{2}{\lambda}},\ \big(1 - |x|\big)_+\right) .$$

along the line of [BFI]. Note that from the Euler-Lagrange equation one observes that if v^λ is continuous near x_0 and $v^\lambda(x_0) \neq u_0(x_0)$, then v^λ is a constant near x_0 [BFI, Proposition 2.2]. This observation is a key to have the above solution.

4 Application to separation of data

We are interested in characterizing

$$S_\pm = \big\{x \in \mathbf{R}^d \;\big|\; S_D[u_0](x) = \pm 1\big\}$$

when u_0 has a special structure like $\chi_A - \chi_B$ where A and B are two disjoint measurable sets in $\mathbf{R}^d$. Roughly speaking the interior of S_+ (resp. S_-) is the set where distance from A is longer (resp. shorter) than that from B. Since our A and B is just measurable, we have to use essential distance instead of usual distance. We define the essential distance $d_e(x, A)$ by

$$d_e(x, A) = \sup\Big\{r \in \mathbf{R} \;\Big|\; \big|B_r(x) \cap A\big| = 0\Big\},$$

where $B_r(x)$ denotes the closed ball of radius r centered at x and $|\cdot|$ denotes the Lebesgue measure. By definition it is clear that $d_e(x,A)$ is Lipschitz continuous on $\mathbf{R}^d$. We consider more general data u_0 than $\chi_A - \chi_B$.

Theorem 4.1. *Assume that* $u_0 \in L^\infty(\mathbf{R}^d)$ *satisfies* ess. $\inf_A u_0 > 0$ *and* ess. $\sup_B u_0 < 0$ *and* $u_0 = 0$ *outside* $A \cup B$*, where* A *and* B *are two disjoint measurable sets or more weakly* $|A \cap B| = 0$*. Then*

$$\operatorname{int} S_+ = \left\{ x \in \mathbf{R}^d \;\middle|\; d_e(x,B) > d_e(x,A) \right\}, \;\; \operatorname{int} S_- = \left\{ x \in \mathbf{R}^d \;\middle|\; d_e(x,B) < d_e(x,A) \right\}. \tag{4.1}$$

Moreover, the complement of $\operatorname{int} S_+ \cup \operatorname{int} S_-$ *has no interior.*

Proof. Since d_e is continuous, it suffices to prove that $d_e(x,A) < d_e(x,B)$ (resp. $d_e(x,B) < d_e(x,A)$) implies $x \in S_+$ (resp. $x \in S_-$). The argument is symmetric so we just give a proof for the case $d_e(x,A) < d_e(x,B)$. By the assumption there exists $\rho > 0$ such that

$$|B_\rho(x) \cap A| = M > 0 \quad \text{and} \quad |B_\rho(x) \cap B| = 0.$$

We divide the integral of $u = G_t * u_0$ as

$$u(x,t) = \int_{B_\rho(x)} G_t(x-y) u_0(y) \mathrm{d}y + \int_{\mathbf{R}^d \setminus B_\rho(x)} G_t(x-y) u_0(y) \mathrm{d}y = I_1 + I_2.$$

Since

$$(4\pi t)^{d/2} I_1 \geq c_A \int_{B_\rho(x) \cap A} \exp\left(-\frac{|x-y|^2}{4t} \right) \mathrm{d}y \geq c_A M \exp(-\rho^2/4t)$$

with $c_A = $ ess. $\inf_A u_0$ and

$$\begin{aligned} (4\pi t)^{d/2} I_2 &\geq -C_B \int_{\mathbf{R}^d \setminus B_\rho(x)} \exp\left(-\frac{|x-y|^2}{4t} \right) \mathrm{d}y \\ &= -C_B \int_\rho^\infty \exp(-r^2/4t) r^{d-1} \mathrm{d}r \mathcal{H}^{d-1}(S^{d-1}) \end{aligned}$$

with $C_B = $ ess. $\sup_B |u_0|$, we observe that

$$\begin{aligned} u(\bar{x},t) \geq & (4\pi t)^{-d/2} \exp\left(-\frac{\rho^2}{4t} \right) \\ & \times \left[c_A M - C_B \mathcal{H}^{d-1}(S^{d-1}) \int_\rho^\infty \exp(-r^2/4t) r^{d-1} \mathrm{d}r \exp\left(\frac{\rho^2}{4t} \right) \right]. \end{aligned}$$

Since

$$0 \leq \exp\left((\rho^2 - r^2)/4t \right) r^{d-1} \leq \exp(\rho^2 - r^2) r^{d-1} \text{ for } \rho \leq r \text{ and } t \leq 1/4$$

and the righthand side is integrable in (ρ, ∞), Lebesgue's dominated convergence theorem implies that

$$\int_{\rho}^{\infty} \exp\left((\rho^2 - r^2)/4t\right) r^{d-1} \mathrm{d}r \to 0$$

as $t \downarrow 0$. We thus conclude that $u(x,t) > 0$ for sufficiently small t. □

It is easy to see that the set where $\mathrm{d}_e(x, A) = \mathrm{d}_e(x, B)$ has no interior so the last assertion follows.

Corollary 1. *Assume that $u_0 = \chi_A - c\chi_B$ with $c > 0$ where A and B are two disjoint measurable set. Then (4.1) holds.*

Note that c does not play any role to determine $\operatorname{int} S_+$ and $\operatorname{int} S_-$.

We consider a data separation problem. Suppose that each point of $\mathbf{R}^d$ fulfills either propery P or Q (with $P \cap Q = \emptyset$) except very thin set. We have to classify a point of $\mathbf{R}^d$ either it fulfills P or Q. We know that in $A \subset \mathbf{R}^d$ the property P is fulfilled while in $B \subset \mathbf{R}^d$. Our S_+ and S_- in Corollary 1 give a way to classify $\mathbf{R}^d$ by the properties P and Q. Corollary 1 gives a characterization of our $\operatorname{int} S_+$ and $\operatorname{int} S_-$ by a completely geometric way. The set

$$C = \mathbf{R}^d \backslash \big((\operatorname{int} S_+) \cup (\operatorname{int} S_-) \big) = \big\{ x \in \mathbf{R}^d \;\big|\; d_e(x, B) = d_e(x, A) \big\}$$

is a separation curve or hypersurface consisting of points having the same distance from A and B. We call this set as an *equi-distance hypersurface*. Note that in general C may not be of finite perimeter.

From Theorem 4.1 and Corollary 1 we have an algorithm of data separation. Set u_0 as

$$u_0(x) = \chi_A(x) - \chi_B(y)$$

for given data A and B, and obtain the solution $u(t,x)$ of (1) with the above u_0. For a reasonable data separation it is better to choose t very small. We observe that our method provides the hypersurface version of a maximal margin classifier [CST] without any technique of data transfer to a higher dimensional space.

We give here a few examples of data separations. In the following numerical examinations we consider the heat equation in a square $\Omega = [-1,1] \times [-1,1]$ with the boundary condition — periodic in x_1 and the homogeneous Neumann condition at $x_2 = -1, 1$ for $x = (x_1, x_2)$. Note that for the heat equation homogeneous Neumann problem in one direction (x_2-direction) is reduced to a periodic boundary value problem by extending a function in x_2 in an even way (a symmetric way) with respect to $x_2 = +1, -1$. Thus under this interpretation our solution with the Neumann in x_2 and periodic in x_1 boundary condition is regarded as the solution of the Cauchy problem (1.1) with an initial data whose restriction in Ω equals our u_0.

We calculate the solution by an explicit difference method with the space lattice grid size $h = 0.01$ and the time grid size $\tau = 0.1 \times h^2$. The difference equation is

$$u_{i,j}^{k+1} = u_{i,j}^k + \frac{\tau}{h^2}(u_{i-1,j}^k + u_{i+1,j}^k + u_{i,j-1}^k + u_{i,j+1}^k - 4u_{i,j}^k),$$

where $u_{i,j}^k = u(hi, hj, \tau k)$ for $-100 \le i, j \le 100$ and $k \ge 0$.

The first one clarifies the difference between our method and support vector machine. Let

$$\begin{cases} A = ([-a, -a+h) \cup (a-h, a]) \times [c, c+h), \\ B = ((-b-h, -b] \cup [b, b+h)) \times (-c-h, -c], \end{cases} \tag{4.2}$$

where $b > a > 0$. It is easy to find that the separation line by the maximal margin classifier (a separation line by a support vector machine) is $\mathbf{R} \times \{0\}$. However, our method provides the equi-distance curve

$$\left\{ (x_1, x_2) \in \mathbf{R}^2 \,\middle|\, \begin{array}{ll} (b-a, 2c) \cdot (x_1 + (a+b)/2, x_2) = 0 & \text{on } (-\infty, 0] \times \mathbf{R}, \\ (a-b, 2c) \cdot (x_1 - (a+b)/2, x_2) = 0 & \text{on } [0, \infty) \times \mathbf{R} \end{array} \right\}$$

for the curve of data separation. If $(\pm a, c)$, $(\pm b, c)$ are on lattice points for numerics and h is smaller than the span of the lattice points, then the initial data for numerics should be given as

$$u_0(x) = \chi_{\{(\pm a, c)\}}(x) - \chi_{\{(\pm b, -c)\}}(x).$$

The calculation in very short time, like as figure 1, provides the separation curve which is very close to the equi-distance curve. If one calculates longer, the curve is smoothened as in the right one of figure 1. If one calculates for a long time i.e. for large t, the curve may fail to classify data as shown in the next example.

The second example is the two-moon type data, which a simple maximal margin classifier cannot draw a separation curve. We give each 100 points of random data for A and B around $\{0.5(\cos\theta, \sin\theta) - (0.25, 0.15) |\ \theta \in [0, \pi]\}$ and $\{0.5(\cos\theta, \sin\theta) + (0.25, 0.15) |\ \theta \in [-\pi, 0]\}$, and set

$$u_0(x) = \chi_A(x) - \chi_B(x).$$

In the right one of figure 3 some points are failed to separate by our method. We thus observe that it is necessary to take t sufficiently small for exact data separation.

If we calculate by an implicit scheme, the results of separation curves are almost the same; however, evidently it takes more time to calculate.

In [CLLMNWZ] a geometric diffusion approach is given for a data separation procedure. Their approach is very general including the data separation by the heat equation given here. Although they discussed the problem on a graph or a manifold, we just explain the idea of their method when A and B are disjoint subsets of $\mathbf{R}^d$. Let S be a compact self-adjoint operator in periodic L^2 space in $\mathbf{R}^d$, i.e. $L^2(\mathbf{T}^d)$ (or

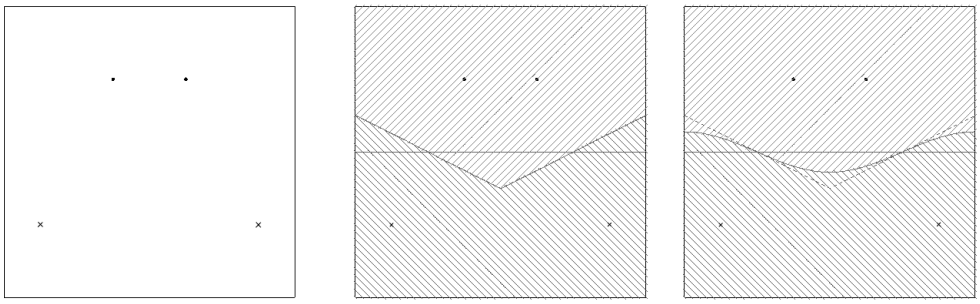

Figure 1: Example of data separation for (4.2). The left figure denotes the profile of the initial data such that $u_0 = 1$ at black dots, $u_0 = -1$ at cross, and $u_0 = 0$ at the others. The center and right figures express the profiles of $\{x|\ u(x,t) = 0\}$ at $t = 0.005$ and $t = 0.2$ (500 and 20000 steps respectively by the explicit difference scheme). The area with slash line is the place where $u(\cdot, t) > 0$, and the other area is the place where $u(\cdot, t) < 0$. The maximal margin classifier [CST], [Std, 22.3.1] provides the solid straight line, which is $[-1, 1] \times \{0\}$, in center and right figures. The dashed line in the right figure is the equi-distance curve, which almost agree with the curve $\{x|\ u(x, 0.005) = 0\}$ in the center figure.

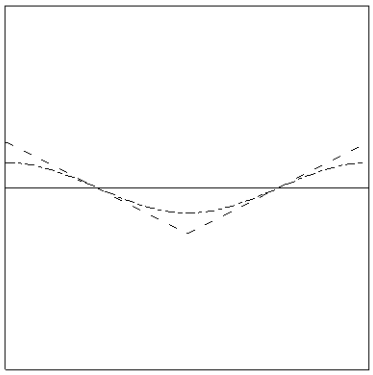

Figure 2: Difference between $\{x|\ u(x,t) = 0\}$ with $t = 0.005$(dashed line), 0.2(chain line) by our diffusive method, and the curve by maximal margin classifier (solid line).

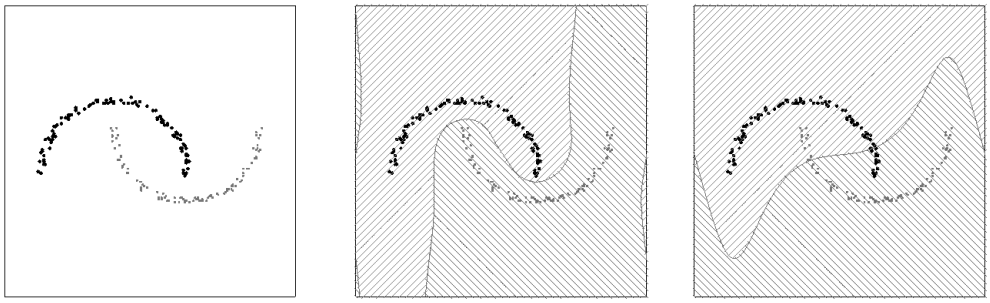

Figure 3: Example of a data separation for two-moon type data. The left figure is the distribution of given data, 100 points of black dots by $u_0 = 1$ and 100 points of gray dots by $u_0 = -1$, and $u_0 = 0$ on the other area. The center and right figures are the profiles of $\{x|\ u(x,t) = 0\}$ with $t = 0.003$ and $t = 0.05$ (300 and 5000 steps respectively by the explicit difference scheme), respectively. The area with slash line is the place where $u(\cdot, t) > 0$, and the other area is the place where $u(\cdot, t) < 0$.

weighted L^2 type Hilbert space). A typical example is an iteration of a resolvent of the Laplacian, i.e.

$$S = S_m := (I - \Delta/m)^{-m} \qquad (m = 1, 2, \ldots).$$

Its limit as $m \to \infty$ is of course $S_\infty = e^\Delta$ which is also a typical example. (The kernel of S is regarded as a similarity function which is constracted by using feature vectors derived from a neighborhood of each pixel for practical purpose. In other words, S is chosen depending upon feature of data sets as explained in [BF].) Let A be a subset where property Q is fulfilled. We set $u_0 = \chi_A - \chi_B$ and introduce a parameter $t > 0$. Then we give a separation

$$A_t^\sharp = \left\{x \in \mathbf{R}^d \;\middle|\; S^t[u_0](x) > 0\right\},$$
$$B_t^\sharp = \left\{x \in \mathbf{R}^d \;\middle|\; S^t[u_0](x) < 0\right\}$$

of $\mathbf{R}^d$. (In practice, t is chosen so that $A_t^\sharp \supset A$, $B_t^\sharp \supset B$.) If $S = S_\infty$, the limit as $t \downarrow 0$ yields our separation.

5 Remark on the method based on the Ginzburg-Landau energy

We now compare minimizers of the Ginzburg-Landau energy (1.12) with the minimizer of (1.7). There exists at least one H^1 minimizer $v^{\varepsilon,\lambda}$ satisfying the Euler-Lagrange equation

$$-\varepsilon\Delta v - 2\left(1 - v^2\right)v/\varepsilon + \lambda v = \lambda u_0. \tag{5.1}$$

For fix $\varepsilon > 0$ if $1/\lambda$ is regarded as the time grid τ, then (5.1) gives an implicit scheme of time discretization on the evolution equation (called the Allen-Cahn equation)

$$\partial u/\partial t = \varepsilon \Delta u + 2\,(1-u^2)\,u/\varepsilon. \tag{5.2}$$

We often rescale t by microscopic time t', i.e. $t = t'/\varepsilon$ when we discuss a phase separation. The equation (5.2) becomes

$$\partial u/\partial t' = \Delta u + 2\,(1-u^2)\,u/\varepsilon^2. \tag{5.3}$$

Starting from initial data u_0, we know that u quickly tends to either 1 or -1 as time develops [XChen] for small ε. As for the heat equation it is natural to define the *diffusive sign by the Allen-Cahn equation* by

$$S_D^{AC\varepsilon}[u_0](x) = \lim_{t\downarrow 0} \operatorname{sgn} u^\varepsilon(x,t)$$

when u^ε is the solution of (5.3) with initial data u_0. We also define the asymptotic sign by the Allen-Cahn equation by

$$S_a^{AC\varepsilon}[u_0](x) = \lim_{\lambda'\to\infty} \operatorname{sgn} v^{\varepsilon,\lambda'\varepsilon}(x). \tag{5.4}$$

We conclude this paper by stating several problems related to relation of diffusive signs.

Problem. *(i) Does $S_D^{AC\varepsilon}$ agree with S_D? In particular, is $S_D^{AC\varepsilon}$ independent of $\varepsilon > 0$?*
(ii) Does $S_D^{AC\varepsilon}$ agree with $S_a^{AC\varepsilon}$? In particular, is $S_a^{AC\varepsilon}$ independent of the choice of a minimizer?
(iii) Is the convergence (5.4) uniform with respect to ε? In other words, does there exist $\hat\lambda = \hat\lambda(x)$ independent of $\varepsilon \in (0,1)$ so that $\operatorname{sgn} v^{\varepsilon,\lambda'\varepsilon}(x)$ is constant for $\lambda' > \hat\lambda(x)$? Or more weakly, $\operatorname{sgn} v^{\varepsilon,\lambda'}(x)$ is constant for $\lambda' > \hat\lambda(x)$?

We know

$$E_{\varepsilon,\lambda}[v] = \frac{1}{2}\int_{\mathbf{R}^d} \left\{ \varepsilon\,|\nabla v|^2 + \frac{1}{\varepsilon}\left(|v|^2-1\right)^2 \right\} \mathrm{d}x + \frac{\lambda}{2}\int_{\mathbf{R}^d} |v-u_0|^2 \mathrm{d}x$$

converges to

$$E_\lambda[v] = \begin{cases} \displaystyle\int_{\mathbf{R}^d} \left|\nabla W(v)\right| + \frac{\lambda}{2}\int_{\mathbf{R}^d} |v-u_0|^2 \mathrm{d}x, \ v(x) \in \{-1,1\}, \ v \in BV(\mathbf{R}^d) \cap L^2(\mathbf{R}^d), \\ \infty, \quad \text{otherwise} \end{cases}$$

with $W(u) = \int_{-1}^{u} (1-r^2)\,\mathrm{d}r$ in the sense of L^1_{loc} Gamma convergence as $\varepsilon \to 0$; see e.g. [MM], [S] for a given $u_0 \in L^2(\mathbf{R}^d)$. The first item $\int \left|\nabla W(v)\right|$ is regarded as

$\frac{4}{3}\int |\nabla v|$ for the valued function $v \in \{-1, 1\}$ since $W(1) = 4/3$. Thus E_λ is essentially the same as (1.10) (with restricting the value of v in $\{-1, 1\}$). Since

$$\varepsilon\,|\nabla v|^2 + \frac{1}{\varepsilon}\left(|v|^2 - 1\right)^2 \geq 2\big|\nabla W(v)\big|$$

for a function v whose value is in $[-1, 1]$, if one knows a bound for $\min E_{\varepsilon\,\lambda}[v]$ for a fixed λ, then it gives a bound for

$$\int_{\mathbf{R}^d} \big|\nabla W(v^{\varepsilon\,\lambda})\big|$$

independent of $\varepsilon \in (0, 1)$. If λ is fixed sufficiently large and $\varepsilon \to 0$, then the limit is a two-valued function whose total variation is finite. This separation seems to give a regularized way of equi-distance separation.

References

[ACM] F. Andreu-Vaillo, V. Caselles and J. M. Mazón, "Parabolic quasilinear equations minimizing linear growth functionals," Progress in Mathematics, **223**, Birkhäuser Verlag, 2004.

[A] S. Angenent, *The zero set of a solution of a parabolic equation*, J. Reine Angew. Math. **390** (1988), 79–96.

[BFI] M. Bonforte and A. Figalli, *Total variation flow and sign fast diffusion in one dimension*, preprint, 2011.

[BF] A. L. Bertozzi and A. Flenner, *Diffuse interface models on graphs for classification of high dimensional data*, preprint, 2011.

[Ch] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces and free boundaries, **6** (2004), 195–218.

[XChen] X. Chen, *Generation and propagation of interfaces for reaction-diffusion equations*, J. Differential Equations, **96** (1992), 116–141.

[CM] X.-Y. Chen and H. Matano, *Convergence, asymptotic* p*eriodicity, and finite-point blow-up in one-dimensional semilinear heat equations*, J. Differential Equations, **78** (1989), 160–190.

[XYChen] X.-Y. Chen, *A strong unique continuation theorem for parabolic equations*, Math. Ann. **311** (1998), no. 4, 603–630.

[CLLMNWZ] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner and S. W. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proc. Natl. Acad. Sci. **102** no. 21 (2005), 7426–7431.

[CST] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," Cambridge University Press, 2000.

[GG] M.-H. Giga and Y. Giga, *Very singular diffusion equations: second and fourth order problems*, Japan J. Indust. Appl. Math. (2010), 323–345.

[GGS] M.-H. Giga, Y. Giga and J. Saal, "Nonlinear partial differential Equations: asymptotic behavior of solutions and self-similar solutions," Progress in Nonlinear Differential Equations and Their Applications, **79**, Birkhäuser, Boston, 2010.

[H] D. Henry, *Some infinite-dimensional Morse-Smale systems defined by parabolic partial differential equations*, J. Differential Equations, **598** (1985), 165–205.

[M] H. Matano, *Nonincrease of the lap-number of a solution for a one-dimensional semilinear parabolic equation*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **29** (1982), no. 2, 401–441.

[MM] L. Modica and S. Mortola, *Un esempio di* Γ*-convergenza: il Teorema di Modica-Mortola*, Boll. Un. Mat. Ital., **B(5)** (1977), 285–299.

[N] K. Nickel, *Gestaltaussagen über Lösungen parabolischer Differentialgleighungen*, J. Reine Angew. Math., **211** (1962), 78–94.

[PW] M. H. Protter and H. F. Weinberger, "Maximum principles in differential equations," Englewood Cliffs, 1967.

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

[Std] G. Steidl, *Supervised learning by support vector machines*, "Handbook of Mathematical Methods in Imaging", (ed. O. Scherzer), vol 3, Springer, 2011, 959–1013.

[S] P. Sternberg *The effect of a singular perturbation on nonconvex variational problems*, Arch. Rational Mech. Anal., **101** (1988), 209–260.

[W] D. G. Widder, "The heat equations," Academic Press, New York, 1975.