Possible titles: PDEs for a Week, PDEs in a Café, Second Course in PDEs; Applied PDEs: Diffusion, Reaction and Fluids

Alex Kiselev[1] Jean-Michel Roquejoffre[2] Lenya Ryzhik[3]

March 9, 2015

[1]Department of Mathematics, University of Wisconsin, 480 Lincoln Drive Madison, WI 53706, USA; kiselev@math.wisc.edu

[2]Institut de Mathématiques Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France; jean-michel.roquejoffre@math.univ-toulouse.fr

[3]Department of Mathematics, Stanford University, Stanford CA, 94305, USA; ryzhik@math.stanford.edu

Chapter 1

Maximum principle and the symmetry of solutions of elliptic equations

1 Act I. The maximum principle enters

We will have two main characters in this chapter: the maximum principle and the sliding method. The latter has a twin, the moving plane method – they are often so indistinguishable that we will count them as one character. They will be introduced separately, and then blended together to study the symmetry properties of the solutions of elliptic equations. In this introductory section, we recall what the maximum principle is. This material is very standard and can be found in almost any undergraduate or graduate PDE text, such as the books by Evans [52], Han and Lin [71], and Pinchover and Rubinstein [104].

We will consider equations of the form

$$\begin{aligned} &\Delta u + F(x,u) = 0 \text{ in } \Omega, \\ &u = g \text{ on } \partial\Omega. \end{aligned} \tag{1.1}$$

Here, Ω is a smooth bounded domain in $\mathbb{R}^n$ and $\partial\Omega$ is its boundary. There are many applications where such problems appear. We will mention just two – one is in the realm of probability theory, where $u(x)$ is an equilibrium particle density for some stochastic process, and the other is in classical physics. In the physics context, one may think of $u(x)$ as the equilibrium temperature distribution inside the domain Ω. The term $F(x,u)$ corresponds to the heat sources or sinks inside Ω, while $g(x)$ is the (prescribed) temperature on the boundary $\partial\Omega$. The maximum principle reflects a basic observation known to any child – first, if $F(x,u) = 0$ (there are neither heat sources nor sinks), or if $F(x,u) \leq 0$ (there are no heat sources but there may be heat sinks), the temeprature inside Ω may not exceed that on the boundary – without a heat source inside a room, you can not heat the interior of a room to a warmer temperature than its maximum on the boundary. Second, if one considers two prescribed boundary conditions and heat sources such that

$$g_1(x) \leq g_2(x) \text{ and } F_1(x,u) \leq F_2(x,u),$$

then the corresponding solutions will satisfy $u_1(x) \leq u_2(x)$ – stronger heating leads to warmer rooms. It is surprising how such mundane considerations may lead to beautiful mathematics.

The maximum principle in complex analysis

Most mathematicians are first introduced to the maximum principle in a complex analysis course. Recall that the real and imaginary parts of an analytic function $f(z)$ have the following property.

Proposition 1.1 *Let $f(z) = u(z) + iv(z)$ be an analytic function in a smooth bounded domain $\Omega \subset \mathbb{C}$, continuous up to the boundary Ω. Then $u(z) = Ref(z)$, $v(z) = Imf(z)$ and $w(z) = |f(z)|$ all attain their respective maxima over Ω on its boundary. In addition, if any of these functions attains its maximum inside Ω, it has to be equal identically to a constant in Ω.*

This proposition is usually proved via the mean-value property of analytic functions (which itself is a consequence of the Cauchy integral formula): for any disk $B(z_0, r)$ contained in Ω we have

$$f(z_0) = \int_0^{2\pi} f(z_0+re^{i\theta})\frac{d\theta}{2\pi}, \quad u(z_0) = \int_0^{2\pi} u(z_0+re^{i\theta})\frac{d\theta}{2\pi}, \quad v(z_0) = \int_0^{2\pi} v(z_0+re^{i\theta})\frac{d\theta}{2\pi}, \tag{1.2}$$

and

$$w(z) \leq \int_0^{2\pi} w(z_0 + re^{i\theta})\frac{d\theta}{2\pi}. \tag{1.3}$$

It is immediate to see that (1.3) implies that if one of the functions u, v and w attains a local maximum at a point z_0 inside Ω, it has to be equal to a constant in a disk around z_0. Thus, the set where it attains its maximum is both open and closed, hence it is all of Ω and this function equals identically to a constant.

The above argument while incredibly beautiful and simple, relies very heavily on the rigidity of analytic functions that is reflected in the mean-value property. The same rigidity is reflected in the fact that the real and imaginary parts of an analytic function satisfy the Laplace equation

$$\Delta u = 0, \quad \Delta v = 0,$$

while $w^2 = u^2 + v^2$ is subharmonic: it satisfies

$$\Delta(w^2) \geq 0.$$

We will see next that the mean-value principle is associated to the Laplace equation and not analyticity in itself, and thus applies to harmonic (and, in a modified way, to subharmonic) functions in higher dimensions as well. This will imply the maximum principle for solutions of the Laplace equation in an arbitrary dimension. One may ask whether a version of the mean-value property also holds for the solutions of general elliptic equations rather than just for the Laplace equation – the answer is "yes if understood properly", and the mean value property survives as the general elliptic regularity theory, an equally beautiful sister of the complex analysis which is occasionally misunderstood as "technical".

Interlude: a probabilistic connection digression

Another good way to understand how the Laplace equation comes about, as well as many of its properties, including the maximum principle, is via its connection to the Brownian motion. It is easy to understand in terms of the discrete equations, which requires only very elementary probability theory. Consider a system of many particles on the n-dimensional integer lattice $\mathbb{Z}^n$. They all perform a symmetric random walk: at each integer time $t = k$ each particle jumps (independently from the others) from its current site $x \in \mathbb{Z}^n$ to one of its $2n$ neighbors, $x \pm e_k$ (e_k is the unit vector in the direction of the x_k-axis), with equal probability $1/(2n)$. At each step we may also insert new particles, the average number of inserted (or eliminated) particles per unit time at each site is $F(x)$. Let now $u_m(x)$ be the average number of particles at the site x at time m. The balance equation for $u_{m+1}(x)$ is

$$u_{m+1}(x) = \frac{1}{2n}\sum_{k=1}^{n}[u_n(x+e_k) + u_n(x-e_k)] + F(x).$$

If the system is in an equilibrium, so that $u_{n+1}(x) = u_n(x)$ for all x, then $u(x)$ (dropping the subscript n) satisfies the discrete equation

$$\frac{1}{2n}\sum_{k=1}^{n}[u(x+e_k) + u(x-e_k) - 2u(x)] + F(x) = 0.$$

If we now take a small mesh size h, rather than one, the above equation becomes

$$\frac{1}{2n}\sum_{k=1}^{n}[u(x+he_k) + u(x-he_k) - 2u(x)] + F(x) = 0.$$

Doing a Taylor expansion in h leads to

$$\frac{h^2}{2n}\sum_{k=1}^{n}\frac{\partial^2 u(x)}{\partial x_k^2} + F(x) = \text{lower order terms}.$$

Taking $F(x) = h^2/(2n)G(x)$ – this prevents us from inserting or removing too many particles, we arrive, in the limit $h \downarrow 0$, at

$$\Delta u + G(x) = 0. \tag{1.4}$$

In this model, we interpret $u(x)$ as the local particle density, and $G(x)$ as the rate at which the particles are inserted (if $G(x) > 0$), or removed (if $G(x) < 0$). When equation (1.4) is posed in a bounded domain Ω, we need to supplement it with a boundary condition, such as

$$u(x) = g(x) \text{ on } \partial\Omega.$$

Here, it means the particle density on the boundary is prescribed – the particles are injected or removed if there are "too many" or "too little" particles at the boundary, to keep $u(x)$ at the given prescribed value $g(x)$.

The mean value property for sub-harmonic and super-harmonic functions

We now return to the world of analysis. A function $u(x)$, $x \in \Omega \subset \mathbb{R}^n$ is harmonic if it satisfies the Laplace equation

$$\Delta u = 0 \text{ in } \Omega. \tag{1.5}$$

This is equation (1.1) with $F \equiv 0$, thus a harmonic function describes a heat distribution in Ω with neither heat sources nor sinks in Ω. We say that u is sub-harmonic if it satisfies

$$-\Delta u \leq 0 \text{ in } \Omega, \tag{1.6}$$

and it is super-harmonic if it satisfies

$$-\Delta u \geq 0 \text{ in } \Omega, \tag{1.7}$$

In other words, a sub-harmonic function satisfies

$$\Delta u + F(x) = 0, \text{ in } \Omega,$$

with $F(x) \leq 0$ – it describes a heat distribution in Ω with only heat sinks present, and no heat sources, while a super-harmonic function satisfies

$$\Delta u + F(x) = 0, \text{ in } \Omega,$$

with $F(x) \geq 0$ – it describes an equilibrium heat distribution in Ω with only heat sources present, and no sinks.

Exercise 1.2 *Give an interpretation of the sub-harmonic and super-harmonic functions in terms of particle probability densities.*

Note that any sub-harmonic function in one dimension is convex:

$$-u'' \leq 0,$$

and then, of course, for any $x \in \mathbb{R}$ and any $l > 0$ we have

$$u(x) \leq \frac{1}{2}\left(u(x+l) + u(x-l)\right), \text{ and } u(x) \leq \frac{1}{2l}\int_{x-l}^{x+l} u(y)dy.$$

The following generalization to sub-harmonic functions in higher dimensions shows that locally $u(x)$ is bounded from above by its spatial average. A super-harmonic function will be locally above its spatial average. A word on notation: for a set S we denote by $|S|$ its volume (or area), and, as before, ∂S denotes its boundary.

Theorem 1.3 *Let $\Omega \subset \mathbb{R}^n$ be an open set and let $B(x,r)$ be a ball centered at $x \in \mathbb{R}^n$ of radius $r > 0$ contained in Ω. Assume that the function $u(x)$ satisfies*

$$-\Delta u \leq 0, \tag{1.8}$$

for all $x \in \Omega$ and that $u \in C^2(\Omega)$. Then we have

$$u(x) \leq \frac{1}{|B(x,r)|}\int_{B(x,r)} u dy, \quad u(x) \leq \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} u dS. \tag{1.9}$$

If the function $u(x)$ is super-harmonic:

$$-\Delta u \geq 0, \tag{1.10}$$

for all $x \in \Omega$ and that $u \in C^2(\Omega)$. Then we have

$$u(x) \geq \frac{1}{|B(x,r)|}\int_{B(x,r)} udy, \quad u(x) \geq \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} udS. \tag{1.11}$$

Moreover, if the function u is harmonic: $\Delta u = 0$, then we have equality in both inequalities in (1.9).

One reason to expect the mean-value property is from physics – if Ω is a ball with no heat sources, it is natural to expect that the temperature in the center of the ball may not exceed the average temperature over any sphere concentric with the ball. The opposite is true if there are no heat sinks (this is true for a super-harmonic function). Another can be seen from the discrete version of inequality (1.8):

$$u(x) \leq \frac{1}{2n}\sum_{j=1}^{n}(u(x+he_j)+u(x-he_j)).$$

Here, h is the mesh size, and e_j is the unit vector in the direction of the coordinate axis for x_j. This discrete equation says exactly that the value $u(x)$ is smaller than the average of the values of u at the neighbors of the point x on the lattice with mesh size h, which is similar to the statement of Theorem 1.3 (though there is no meaning to "nearest" neighbor in the continuous case).

Proof. We will only treat the case of a sub-harmonic function. Let us fix the point $x \in \Omega$ and define

$$\phi(r) = \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} u(z)dS(z). \tag{1.12}$$

It is easy to see that, since $u(x)$ is continuous, we have

$$\lim_{r\downarrow 0}\phi(r) = u(x). \tag{1.13}$$

Therefore, we would be done if we knew that $\phi'(r) \geq 0$ for all $r > 0$ (and such that the ball $B(x,r)$ is contained in Ω). To this end, passing to the polar coordinates $z = x + ry$, with $y \in \partial B(0,1)$, we may rewrite (1.12) as

$$\phi(r) = \frac{1}{|\partial B(0,1)|}\int_{\partial B(0,1)} u(x+ry)dS(y).$$

Then, differentiating in r gives

$$\phi'(r) = \frac{1}{|\partial B(0,1)|}\int_{\partial B(0,1)} y\cdot\nabla u(x+ry)dS(y).$$

Going back to the z-variables gives

$$\phi'(r) = \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} \frac{1}{r}(z-x)\cdot\nabla u(z)dS(z) = \frac{1}{|\partial B(x,r)|}\int_{\partial B(x,r)} \frac{\partial u}{\partial\nu}dS(z).$$

Here, we used the fact that the outward normal to $B(x,r)$ at a point $z \in \partial B(x,r)$ is

$$\nu = (z-x)/r.$$

Using Green's formula

$$\int_U \Delta g dy = \int_U \nabla \cdot (\nabla g) = \int_{\partial U} (\nu \cdot \nabla g) = \int_{\partial U} \frac{\partial g}{\partial \nu} dS,$$

gives now

$$\phi'(r) = \frac{1}{|\partial B(x,r)|} \int_{B(x,r)} \Delta u(y) dy \geq 0.$$

It follows that $\phi(r)$ is a non-decreasing function of r, and then (1.13) implies that

$$u(x) \leq \frac{1}{|\partial B(x,r)|} \int_{\partial B(x,r)} u dS, \tag{1.14}$$

which is the second identity in (1.9).

In order to prove the first equality in (1.9) we use the polar coordinates once again:

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} u dy = \frac{1}{|B(x,r)|} \int_0^r \left(\int_{\partial B(x,s)} u dS \right) ds \geq \frac{1}{|B(x,r)|} \int_0^r u(x) n\alpha(n) s^{n-1} ds$$
$$= u(x)\alpha(n)r^n \frac{1}{\alpha(n)r^n} = u(x).$$

We used above two facts: first, the already proved identity (1.14) about averages on spherical shells, and, second, that the area of an $(n-1)$-dimensional unit sphere is $n\alpha(n)$, where $\alpha(n)$ is the volume of the n-dimensional unit ball. Now, the proof of (1.9) is complete. The proof of the mean-value property for super-harmonic functions works identically. □

The weak maximum principle

The first consequence of the mean value property is the maximum principle that says that a sub-harmonic function attains its maximum over any domain on the boundary and not inside the domain[1]. Once again, in one dimension this is obvious: a smooth convex function does not have any local maxima.

Theorem 1.4 (**The weak maximum principle**) *Let $u(x)$ be a sub-harmonic function in a connected domain Ω and assume that $u \in C^2(\Omega) \cap C(\bar{\Omega})$. Then*

$$\max_{x \in \bar{\Omega}} u(x) = \max_{y \in \partial\Omega} u(y). \tag{1.15}$$

Moreover, if $u(x)$ achieves its maximum at a point x_0 in the interior of Ω, then $u(x)$ is identically equal to a constant in Ω. Similarly, if $u \in C^2(\Omega) \cap C(\bar{\Omega})$ is a super-harmonic function in Ω, then

$$\min_{x \in \bar{\Omega}} u(x) = \min_{y \in \partial\Omega} u(y). \tag{1.16}$$

Moreover, if $u(x)$ achieves its minimum at a point x_0 in the interior of Ω, then $u(x)$ is identically equal to a constant in Ω.

[1]A sub-harmonic function is nothing but the heat distribution in a room without heat sources, hence it is very natural that it attains its maximum on the boundary (the walls of the room)

Proof. Again, we only treat the case of a sub-harmonic function. Suppose that $u(x)$ attains its maximum at an interior point $x_0 \in \Omega$, and set

$$M = u(x_0).$$

Then, for any $r > 0$ sufficiently small (so that the ball $B(x_0, r)$ is contained in Ω), we have

$$M = u(x) \leq \frac{1}{|B(x_0,r)|} \int_{B(x_0,r)} u dy \leq M,$$

with the equality above holding only if $u(y) = M$ for all y in the ball $B(x_0, r)$. Therefore, the set S of points where $u(x) = M$ is open. Since $u(x)$ is continuous, this set is also closed. Since S us both open and closed in Ω, and Ω is connected, it follows that $S = \Omega$, hence $u(x) = M$ at all points $x \in \Omega$. $\Box$

We will often have to deal with slightly more general operators than the Laplacian, of the form

$$Lu = \Delta u(x) + c(x)u. \tag{1.17}$$

We may ask the same question: when is it true that the inequality

$$-\Delta u(x) - c(x)u(x) \leq 0 \text{ in } \Omega \tag{1.18}$$

guarantees that $u(x)$ attains its maximum on the boundary of Ω? It is certainly not always true that any function satisfying (1.18) attains its maximum on he boundary: consider the function $u(x) = \sin x$ on the interval $(0, \pi)$. It satisfies

$$u''(x) + u(x) = 0, \quad u(0) = u(\pi) = 0, \tag{1.19}$$

but achieves its maximum at $x = \pi/2$. In order to understand this issue a little better, consider the following exercise.

Exercise 1.5 *Consider the boundary value problem*

$$-u'' - au = f(x), \quad 0 < x < 1, \quad u(0) = u(1) = 0,$$

with a given non-negative function $f(x)$, and a constant $a \geq 0$. Show that if $a < \pi^2$, then the function $u(x)$ is positive on the interval $(0, 1)$.

One possible answer to our question below (1.18) comes from our childish attempts at physics: if $u(x) \geq 0$, we may interpret $u(x)$ as a heat distribution in Ω. Then, $u(x)$ should not be able to attain its maximum inside Ω if there are no heat sources in Ω. If $u(x)$ satisfies (1.18), the only possible heat source is $c(x)u(x)$. Keeping in mind that $u(x) \geq 0$, we see that absence of heat sources is equivalent to the condition $c(x) \leq 0$ (this, in particular, rules out the counterexample (1.19)). Mathematically, this is reflected in the following.

Corollary 1.6 *Suppose that $c(x) \leq 0$ in Ω, and a function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ satisfies $u \geq 0$ and*

$$\Delta u(x) + c(x)u(x) \geq 0 \text{ in } \Omega.$$

Then u attains its maximum on $\partial\Omega$. Moreover, if $u(x)$ attains its maximum inside Ω then u is identically equal to a constant.

Proof. A non-negative function $u(x)$ that satisfies (1.18) is sub-harmonic, and application of Theorem 1.4 finishes the proof.

Exercise 1.7 *Give an interpretation of this result in terms of particle densities.*

2 Act II. The moving plane method

2.1 The isoperimeteric inequality and sliding

We now bring in our second set of characters, the moving plane and sliding methods. As an introduction, we show how the sliding method can work alone, without the maximum principle. Maybe the simplest situation when the sliding idea proves useful is in an elegant proof of the isoperimetric inequality. We follow here the proof given by X. Cabré in [27][2]. The isoperimetric inequality says that among all domains of a given volume the ball has the smallest surface area.

Theorem 2.1 *Let Ω be a smooth bounded domain in $\mathbb{R}^n$. Then,*

$$\frac{|\partial\Omega|}{|\Omega|^{(n-1)/n}} \geq \frac{|\partial B_1|}{|B_1|^{(n-1)/n}}, \tag{2.1}$$

where B_1 is the open unit ball in $\mathbb{R}^n$, $|\Omega|$ denotes the measure of Ω and $|\partial\Omega|$ is the perimeter of Ω (the $(n-1)$-dimensional measure of the boundary of Ω). In addition, equality in (2.1) holds if and only if Ω is a ball.

A technical aside: the area formula

The proof will use the area formula (see [53] for the proof), a generalization of the usual change of variables formula in the multi-variable calculus. The latter says that if $f : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth one-to-one map (a change of variables), then

$$\int_{\mathbb{R}^n} g(x)Jf(x)dx = \int_{\mathbb{R}^n} g(f^{-1}(y))dy. \tag{2.2}$$

For general maps we have

Theorem 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz map with the Jacobian Jf. Then, for each function $g \in L^1(\mathbb{R}^n)$ we have*

$$\int_{\mathbb{R}^n} g(x)Jf(x)dx = \int_{\mathbb{R}^n} \left[\sum_{x\in f^{-1}\{y\}} g(x) \right] dy. \tag{2.3}$$

We will, in particular, need the following corollary.

Corollary 2.3 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz map with the Jacobian Jf. Then, for each measurable set $A \subset \mathbb{R}^n$ we have*

$$|f(A)| \leq \int_A Jf(x)dx. \tag{2.4}$$

[2]Readers with ordinary linguistic powers may consult [28].

Proof. For a given set S we define its characteristic functions as

$$\chi_S(x) = \begin{cases} 1, & \text{for } x \in S, \\ 0, & \text{for } x \notin S, \end{cases}$$

We use the area formula with $g(x) = \chi_A(x)$:

$$\begin{aligned} \int_A Jf(x)dx &= \int_{\mathbb{R}^n} \chi_A(x) Jf(x)dx = \int_{\mathbb{R}^n} \left[\sum_{x \in f^{-1}\{y\}} \chi_A(x) \right] dy \\ &= \int_{\mathbb{R}^n} [\#x \in A: \; f(x) = y] \, dy \geq \int_{\mathbb{R}^n} \chi_{f(A)}(y) dy = |f(A)|, \end{aligned}$$

and we are done. □

A more general form of this corollary is the following.

Corollary 2.4 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz map with the Jacobian Jf. Then, for each nonnegative function $p \in L^1(\mathbb{R}^n)$ and each measurable set A, we have*

$$\int_{f(A)} p(y)dy \leq \int_A p(f(x)) Jf(x)dx. \tag{2.5}$$

Proof. The proof is as in the previous corollary. This time, we apply the area formula to the function $g(x) = p(f(x))\chi_A(x)$:

$$\begin{aligned} &\int_A p(f(x)) Jf(x)dx = \int_{\mathbb{R}^n} \chi_A(x) p(f(x)) Jf(x)dx = \int_{\mathbb{R}^n} \left[\sum_{x \in f^{-1}\{y\}} \chi_A(x) p(f(x)) \right] dy \\ &= \int_{\mathbb{R}^n} [\#x \in A: \; f(x) = y] \, p(y)dy \geq \int_{f(A)} p(y)dy, \end{aligned}$$

and we are done. □

The proof of the isoperimetric inequality

We now proceed with Cabré's proof of the isoperimetric inequality in Theorem 2.1.

Step 1: sliding. Let $v(x)$ be the solution of the Neumann problem

$$\begin{aligned} &\Delta v = k, \;\; \text{in } \Omega, \\ &\frac{\partial v}{\partial \nu} = 1 \text{ on } \partial\Omega. \end{aligned} \tag{2.6}$$

Integrating the first equation above and using the boundary condition, we obtain

$$k|\Omega| = \int_\Omega \Delta v dx = \int_{\partial\Omega} \frac{\partial u}{\partial \nu} = |\partial\Omega|.$$

Hence, solution exists only if

$$k = \frac{|\partial\Omega|}{|\Omega|}. \tag{2.7}$$

It is a classical result that with this particular value of k there exist infinitely many solutions that differ by addition of an arbitrary constant. We let v be any of them. As Ω is a smooth domain, v is also smooth.

Let Γ_v be the lower contact set of v, that is, the set of all $x \in \Omega$ such that the tangent hyperplane to the graph of v at x lies below that graph in all of $\bar{\Omega}$. More formally, we define

$$\Gamma_v = \{x \in \Omega : \; v(y) \geq v(x) + \nabla v(x) \cdot (y - x) \text{ for all } y \in \bar{\Omega}.\} \tag{2.8}$$

The crucial observation is that

$$B_1 \subset \nabla v(\Gamma_v). \tag{2.9}$$

Exercise 2.5 *Explain why this is trivial in one dimension.*

Here, B_1 is the open unit ball centered at the origin. The geometric reason for this is as follows: take any $p \in B_1$ and consider the graphs of the functions

$$r_c(y) = p \cdot y + c.$$

We will now slide this plane upward – we will start with a "very negative" c, and start increasing it, moving the plane up. Note that there exists $M > 0$ so that if $c < -M$, then

$$r_c(y) < v(y) - 100 \text{ for all } y \in \bar{\Omega},$$

that is, the plane is below the graph in all of Ω. On the other hand, if $c > M$, then

$$r_c(y) > v(y) + 100 \text{ for all } y \in \bar{\Omega},$$

in other words, the plane is above the graph in all of Ω. Let

$$\alpha = \sup\{c \in \mathbb{R} : \; r_c(y) < v(y) \text{ for all } y \in \bar{\Omega}\} \tag{2.10}$$

be the largest c so that the plane lies below the graph of v in all of Ω. It is easy to see that the plane $r_\alpha(y) = p \cdot y + \alpha$ has to touch the graph of v: there exists a point $y_0 \in \bar{\Omega}$ such that $r_\alpha(y_0) = v(y_0)$ and

$$r_\alpha(y) \leq v(y) \text{ for all } y \in \bar{\Omega}. \tag{2.11}$$

Furthermore, the point y_0 can not lie on the boundary $\partial\Omega$. Indeed, for all $y \in \partial\Omega$ we have

$$\left|\frac{\partial r_c}{\partial \nu}\right| = |p \cdot \nu| \leq |p| < 1 \text{ and } \frac{\partial v}{\partial \nu} = 1.$$

This means that if $r_c(y) = v(y)$ for some c, and y is on the boundary $\partial\Omega$, then there is a neighborhood $U \in \Omega$ of y such that $r_c(y) > v(y)$ for all $y \in U$. Comparing to (2.11), we see that $c \neq \alpha$, hence it is impossible that $y_0 \in \partial\Omega$. Thus, y_0 is an interior point of Ω, and, moreover, the graph of $r_\alpha(y)$ is the tangent plane to v at y_0. In particular, we have $\nabla v(y_0) = p$, and (2.11) implies that y_0 is in the contact set of v: $y_0 \in \Gamma_v$. We have now shown the inclusion (2.9): $B_1 \subset \nabla v(\Gamma_v)$. Note that the only information about the function $v(x)$ we have used so far is the Neumann boundary condition

$$\frac{\partial v}{\partial \nu} = 1 \text{ on } \partial\Omega,$$

but not the Poisson equation for v in Ω.

Step 2: using the area formula. A trivial consequence of (2.9) is that

$$|B_1| \leq |\nabla v(\Gamma_v)|. \tag{2.12}$$

Now, we will apply Corollary 2.3 to the map $\nabla v: \Gamma_v \to \nabla v(\Gamma_v)$, whose Jacobian is $|\det[D^2 v]|$.

Exercise 2.6 *Show that if Γ_v is the contact set of a smooth function $v(x)$, then det$[D^2 v]$ is non-negative for $x \in \Gamma_v$, and, moreover, all eigenvalues of $D^2 v$ are nonnegative on Γ_v.*

As $\det[D^2 v]$ is non-negative for $x \in \Gamma_v$, we conclude from Corollary 2.3 and (2.12) that

$$|B_1| \leq |\nabla v(\Gamma_v)| \leq \int_{\Gamma_v} \det[D^2 v(x)] dx. \tag{2.13}$$

It remains to notice that by the classical arithmetic mean-geometric mean inequality applied to the (nonnegative) eigenvalues $\lambda_1, \ldots, \lambda_n$ of the matrix $D^2 v(x)$, $x \in \Gamma_v$ we have

$$\det[D^2 v(x)] = \lambda_1 \lambda_2 \ldots \lambda_n \leq \left(\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{n} \right)^n. \tag{2.14}$$

However, by a well-known formula from linear algebra,

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{Tr}[D^2 v],$$

and, moreover, $\mathrm{Tr}[D^2 v]$ is simply the Laplacian Δv. This gives

$$\det[D^2 v(x)] \leq \left(\frac{\mathrm{Tr}[D^2 v]}{n} \right)^n = \left(\frac{\Delta v}{n} \right)^n \text{ for } x \in \Gamma_v. \tag{2.15}$$

Recall that v is the solution of (2.6):

$$\begin{aligned} &\Delta v = k, \quad \text{in } \Omega, \\ &\frac{\partial v}{\partial \nu} = 1 \text{ on } \partial\Omega. \end{aligned} \tag{2.16}$$

with

$$k = \frac{|\partial \Omega|}{|\Omega|}.$$

Going back to (2.13), we deduce that

$$|B_1| \leq \int_{\Gamma_v} \det[D^2 v(x)] dx \leq \int_{\Gamma_v} \left(\frac{\Delta v}{n} \right)^n dx \leq \left(\frac{k}{n} \right)^n |\Gamma_v| = \left(\frac{|\partial \Omega|}{n|\Omega|} \right)^n |\Gamma_v| \leq \left(\frac{|\partial \Omega|}{n|\Omega|} \right)^n |\Omega|.$$

In addition, for the unit ball we have $|\partial B_1| = n|B_1|$, hence the above implies

$$\frac{|\partial B_1|^n}{|B_1|^{n-1}} \leq \frac{|\partial \Omega|^n}{|\Omega|^{n-1}}, \tag{2.17}$$

which is nothing but the isoperimetric inequality (2.1).

In order to see that the inequality in (2.17) is strict unless Ω is a ball, we observe that it follows from the above argument that for the equality to hold in (2.17) we must have equality in (2.14), and, in addition, Γ_v has to coincide with Ω. This means that for each $x \in \Omega$ all eigenvalues of the matrix $D^2 v(x)$ are equal to each other. That is, $D^2 v(x)$ is a multiple of the identity matrix for each $x \in \Omega$.

Exercise 2.7 *Show that if $v(x)$ is a smooth function such that*

$$\frac{\partial^2 v(x)}{\partial x_i^2} = \frac{\partial^2 v(x)}{\partial x_j^2},$$

for all $1 \leq i, j \leq n$ and $x \in \Omega$, and

$$\frac{\partial^2 v(x)}{\partial x_i \partial x_j} = 0,$$

for all $i \neq j$ and $x \in \Omega$, then there exists $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $b \in \mathbb{R}$, so that

$$v(x) = b\left[(x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2\right] + c, \tag{2.18}$$

for all $x \in \Omega$.

Our function $v(x)$ satisfies the assumptions of Exercise 2.7, hence it must have the form (2.18). Finally, the boundary condition $\partial v/\partial \nu = k$ on $\partial\Omega$ implies that Ω is a ball centered at the point $a \in \mathbb{R}^n$. □

3 Act III. Their first meeting

The maximum principle returns, and we study it in a slightly greater depth. At the end of this act the maximum principle and the moving plane method are introduced to each other.

The Hopf lemma and the strong maximum principle

The weak maximum principle leaves open the possibility that u attains its maximum both on the boundary and inside Ω. The strong maximum principle will rule out this possibility unless u is identically equal to a constant. Let us begin with the following exercises.

Exercise 3.1 *Show that if the function $u(x)$ satisfies an ODE of the form*

$$u'' + c(x)u = 0, \quad a < x < b, \tag{3.1}$$

and $u(x_0) = 0$ for some $x_0 \in (a, b)$ then u can not attain its maximum (or minimum) over the interval (a, b) at the point x_0 unless $u \equiv 0$.

This exercise is relatively easy – one has to think about the initial value problem for (3.1) with the data $u(x_0) = u'(x_0) = 0$. Now, look at the next exercise, which is slightly harder.

Exercise 3.2 *Show that, once again, in one dimension, if $u(x)$, $x \in \mathbb{R}$ satisfies a differential inequality of the form*

$$u'' + c(x)u \geq 0, \quad a < x < b,$$

and $u(x_0) = 0$ for some $x_0 \in (a, b)$ then u can not attain its maximum over the interval (a, b) at the point x_0 unless $u \equiv 0$.

The proof of the strong maximum principle relies on the Hopf lemma which guarantees that the point on the boundary where the maximum is attained is not a critical point of u.

Theorem 3.3 *(The Hopf Lemma) Let $B = B(y,r)$ be an open ball in $\mathbb{R}^n$ with $x_0 \in \partial B$, and assume that $c(x) \le 0$ in B. Suppose that a function $u \in C^2(B) \cap C(B \cup x_0)$ satisfies*

$$\Delta u + c(x)u \ge 0 \text{ in } B,$$

and that $u(x) < u(x_0)$ for any $x \in B$ and $u(x_0) \ge 0$. Then, we have $\dfrac{\partial u}{\partial \nu}(x_0) > 0$.

Proof. We may assume without loss of generality that B is centered at the origin: $y = 0$. We may also assume that $u \in C(\bar{B})$ and that $u(x) < u(x_0)$ for all $x \in \bar{B} \backslash \{x_0\}$ – otherwise, we would simply consider a smaller ball $B_1 \subset B$ that is tangent to B at x_0.

The idea is to modify u to turn it into a strict sub-solution of the form

$$w(x) = u(x) + \varepsilon h(x).$$

We also need w to inherit the other properties of u: it should attain its maximum over $\bar{B}$ at x_0, and we need to have $w(x) < w(x_0)$ for all $x \in B$. In addition, we would like to have

$$\frac{\partial h}{\partial \nu} < 0 \text{ on } \partial B,$$

so that the inequality

$$\frac{\partial w}{\partial \nu}(x_0) \ge 0$$

would imply

$$\frac{\partial u}{\partial \nu}(x_0) > 0.$$

An appropriate choice is

$$h(x) = e^{-\alpha|x|^2} - e^{-\alpha r^2},$$

in a smaller domain

$$\Sigma = B \cap B(x_0, r/2).$$

Observe that $h > 0$ in B, $h = 0$ on ∂B (thus, h attains its minimum on ∂B – unlike u which attains its maximum there), and, in addition:

$$\begin{aligned}\Delta h + c(x)h &= e^{-\alpha|x|^2}\left[4\alpha^2|x|^2 - 2\alpha n + c(x)\right] - c(x)e^{-\alpha r^2} \\ &\ge e^{-\alpha|x|^2}\left[4\alpha^2|x|^2 - 2\alpha n + c(x)\right] \ge e^{-\alpha|x|^2}\left[4\alpha^2\frac{|r|^2}{4} - 2\alpha n + c(x)\right] > 0,\end{aligned}$$

for all $x \in \Sigma$ for a sufficiently large $\alpha > 0$. Hence, we have a strict inequality

$$\Delta w + c(x)w > 0, \text{ in } \Sigma,$$

for all $\varepsilon > 0$. Thus, w may not attain its non-negative maximum over Σ inside Σ but only on its boundary. We now show that if $\varepsilon > 0$ is sufficiently small, then w attains this maximum only at x_0. Indeed, as $u(x) < u(x_0)$ in B, we may find δ, so that

$$u(x) < u(x_0) - \delta \text{ for } x \in \partial\Sigma \cap B.$$

Take ε so that
$$\varepsilon h(x) < \delta \text{ on } \partial\Sigma \cap B,$$
then
$$w(x) < u(x_0) = w(x_0) \text{ for all } x \in \partial\Sigma \cap B.$$
On the other hand, for $x \in \partial\Sigma \cap \partial B$ we have $h(x) = 0$ and
$$w(x) = u(x) < u(x_0) = w(x_0).$$
We conclude that $w(x)$ attains its non-negative maximum in $\bar{\Sigma}$ at x_0 if ε is sufficiently small. This implies
$$\frac{\partial w}{\partial \nu}(x_0) \geq 0,$$
and, as a consequence
$$\frac{\partial u}{\partial \nu}(x_0) \geq \varepsilon \frac{\partial h}{\partial \nu}(x_0) = \varepsilon \alpha r e^{-\alpha r^2} > 0.$$
This finishes the proof. □

The strong maximum principle is an immediate consequence of the Hopf lemma.

Theorem 3.4 *(The Strong maximum Principle) Assume that $c(x) \leq 0$ in Ω, and the function $u \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfies*
$$\Delta u + c(x)u \geq 0,$$
and attains its maximum over $\bar{\Omega}$ at a point x_0. Then, if $u(x_0) \geq 0$, then $x_0 \in \partial\Omega$ unless u is a constant. If the domain Ω has the internal sphere property, and $u \not\equiv$ const, then
$$\frac{\partial u}{\partial \nu}(x_0) > 0.$$

Proof. Let $M = \sup_{\bar{\Omega}} u(x)$ and define the set $\Sigma = \{x \in \Omega : \ u(x) = M\}$, where the maximum is attained. We need to show that either Σ is empty or $\Sigma = \Omega$. Assume that Σ is non-empty but $\Sigma \neq \Omega$, and choose a point $p \in \Omega \backslash \Sigma$ such that
$$d_0 = d(p, \Sigma) < d(p, \partial\Omega).$$
Consider the ball $B_0 = B(p, d_0)$ and let $x_0 \in \partial B_0 \cap \partial\Sigma$. Then we have
$$\Delta u + c(x)u \geq 0 \text{ in } B_0,$$
and
$$u(x) < u(x_0) = M,\ M \geq 0 \text{ for all } x \in B_0.$$
The Hopf Lemma implies that
$$\frac{\partial u}{\partial \nu}(x_0) > 0,$$
where ν is the normal to B_0 at x_0. However, x_0 is an internal maximum of u in Ω and hence $\nabla u(x_0) = 0$. This is a contradiction. □

The following corollary of the strong maximum principle is more delicate than our baby physics arguments – we make no assumption on whether $c(x)u(x)$ is a heat source or sink.

Corollary 3.5 *Assume that $c(x)$ is a bounded function, and $u \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfies*

$$\Delta u + c(x)u \geq 0. \tag{3.2}$$

If $u \leq 0$ in Ω then either $u \equiv 0$ in Ω or $u < 0$ in Ω. Similarly, if $u \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfies

$$\Delta u(x) + c(x)u(x) \leq 0 \text{ in } \Omega, \tag{3.3}$$

with $u \geq 0$ in Ω, with a bounded function $c(x)$. Then either $u \equiv 0$ in Ω or $u > 0$ in Ω.

Proof. If $c(x) \leq 0$, this follows directly from the strong maximum principle. In the general case, as $u \leq 0$ in Ω, the inequality (3.2) implies that, for any $M > 0$ we have

$$\Delta u + c(x)u - Mu \geq -Mu \geq 0.$$

However, if $M > \|c\|_{L^\infty(\Omega)}$ then the zero order coefficient satisfies

$$c_1(x) = c(x) - M \leq 0,$$

hence we may conclude, again from the strong maximum principle that either $u < 0$ in Ω or $u \equiv 0$ in Ω. The proof in the case (3.3) holds is identical. □

It is easy to understand the strong maximum principle from the point of view of (3.3) – in this case, a non-negative $u(x)$ can be interpreted as a particle density, and $c(x)u(x)$ is the rate at which the particles are inserted (where $c(x) > 0$) or eliminated (where $c(x) < 0$). The strong maximum principle says that no matter how negative $c(x)$ is, the random particles will always access any point in the domain with a positive probability density.

Separating sub- and super-solutions

A very common use of the strong maximum principle is to re-interpret it as the "untouchability" of a sub-solution and a super-solution of a linear or nonlinear problem – the basic principle underlying what we will see below. Assume that the functions $u(x)$ and $v(x)$ satisfy

$$\Delta u + f(x, u) \geq 0, \quad \Delta v + f(x, v) \leq 0 \text{ in } \Omega. \tag{3.4}$$

We say that $u(x)$ is a sub-solution, and $v(x)$ is a super-solution. Assume that, in addition, we know that

$$u(x) \leq v(x) \text{ for all } x \in \Omega, \tag{3.5}$$

that is, the sub-solution sits below the super-solution. In this case, we are going to rule out the possibility that they touch inside Ω (they can touch on the boundary, however): there can not be an $x_0 \in \Omega$ so that $u(x_0) = v(x_0)$. Indeed, if the function $f(x, s)$ is differentiable (or Lipschitz), the quotient

$$c(x) = \frac{f(x, u(x)) - f(x, v(x))}{u(x) - v(x)}$$

is a bounded function, and the difference $w(x) = u(x) - v(x)$ satisfies

$$\Delta w + c(x)w \geq 0 \text{ in } \Omega. \tag{3.6}$$

As $w(x) \leq 0$ in all of Ω, the strong maximum principle implies that either $w(x) \equiv 0$, so that u and v coincide, or $w(x) < 0$ in Ω, that is, we have a strict inequality: $u(x) < v(x)$ for all $x \in \Omega$. In other words, a sub-solution and a super-solution can not touch at a point – this very simple principle will be extremely important in what follows.

Let us illustrate an application of the strong maximum principle, with a cameo appearance of the sliding method in a disguise as a bonus. Consider the boundary value problem

$$-u'' = e^u, \quad 0 < x < L, \tag{3.7}$$

with the boundary condition

$$u(0) = u(L) = 0. \tag{3.8}$$

If we think of $u(x)$ as a temperature distribution, then the boundary condition means that the boundary is "cold". On the other hand, the positive term e^u is a "heating term", which competes with the cooling by the boundary. A nonnegative solution $u(x)$ corresponds to an equilibrium between these two effects. We would like to show that if the length of the interval L is sufficiently large, then no such equilibrium is possible – the physical reason is that the boundary is too far from the middle of the interval, so the heating term wins. This absence of an equilibrium is interpreted as an explosion, and this model was introduced exactly in that context in late 30's-early 40's. It is convenient to work with the function $w = u + \varepsilon$, which satisfies

$$-w'' = e^{-\varepsilon} e^w, \quad 0 < x < L, \tag{3.9}$$

with the boundary condition

$$w(0) = w(L) = \varepsilon. \tag{3.10}$$

Consider a family of functions

$$v_\lambda(x) = \lambda \sin\left(\frac{\pi x}{L}\right), \quad \lambda \geq 0, \quad 0 < x < L.$$

These functions satisfy (for any $\lambda \geq 0$)

$$v_\lambda'' + \frac{\pi^2}{L^2} v_\lambda = 0, \quad v_\lambda(0) = v_\lambda(L) = 0. \tag{3.11}$$

Therefore, if L is so large that

$$\frac{\pi^2}{L^2} s \leq e^{-\varepsilon} e^s, \quad \text{for all } s \geq 0,$$

we have

$$w'' + \frac{\pi^2}{L^2} w \leq 0, \tag{3.12}$$

that is, w is a super-solution for (3.11). In addition, when $\lambda > 0$ is sufficiently small, we have

$$v_\lambda(x) \leq w(x) \text{ for all } 0 \leq x \leq L. \tag{3.13}$$

Let us now start increasing λ until the graphs of v_λ and w touch at some point:

$$\lambda_0 = \sup\{\lambda : \; v_\lambda(x) \leq w(x) \text{ for all } 0 \leq x \leq L.\} \tag{3.14}$$

The difference

$$p(x) = v_{\lambda_0}(x) - w(x)$$

satisfies

$$p'' + \frac{\pi^2}{L^2} p \geq 0,$$

and $p(x) \leq 0$ for all $0 < x < L$. In addition, there exists x_0 such that $p(x_0) = 0$, and, as $v_\lambda(0) = v_\lambda(L) = 0 < \varepsilon = w(0) = w(L)$, it is impossible that $x_0 = 0$ or $x_0 = L$. We conclude that $p(x) \equiv 0$, which is a contradiction. Hence, no solution of (3.7)-(3.8) may exist when L is sufficiently large.

In order to complete the picture, the reader may look at the following exercise.

Exercise 3.6 *Show that there exists $L_1 > 0$ so that a nonnegative solution of (3.7)-(3.8) exists for all $0 < L < L_1$, and does not exist for all $L > L_1$.*

The maximum principle for narrow domains

Before we allow the moving plane method to return, we describe the maximum principle for narrow domains, which is an indispensable tool in this method. Its proof will utilize the "ballooning method" we have seen in the analysis of the explosion problem. As we have discussed, the usual maximum principle in the form "$\Delta u + c(x)u \geq 0$ in Ω, $u \leq 0$ on $\partial\Omega$ implies either $u \equiv 0$ or $u < 0$ in Ω" can be interpreted physically as follows. If u is the temperature distribution then the boundary condition $u \leq 0$ means that "the boundary is cold" while the term $c(x)u$ can be viewed as a heat source if $c(x) \geq 0$ or as a heat sink if $c(x) \leq 0$. The conditions $u \leq 0$ on $\partial\Omega$ and $c(x) \leq 0$ together mean that both the boundary is cold and there are no heat sources – therefore, the temperature is cold everywhere, and we get $u \leq 0$. On the other hand, if the domain is such that each point inside Ω is "close to the boundary" then the effect of the cold boundary can dominate over a heat source, and then, even if $c(x) \geq 0$ at some (or all) points $x \in \Omega$, the maximum principle still holds.

Mathematically, the first step in that direction is the maximum principle for narrow domains. We use the notation $c^+(x) = \max[0, c(x)]$.

Theorem 3.7 *(The maximum principle for narrow domains) Let e be a unit vector. There exists $d_0 > 0$ that depends on the L^∞-norm $\|c^+\|_\infty$ so that if $|(y-x)\cdot e| < d_0$ for all $(x,y) \in \Omega$ then the maximum principle holds for the operator $\Delta + c(x)$. That is, if $u \in C^2(\Omega) \cap C^1(\bar{\Omega})$ satisfies*

$$\Delta u(x) + c(x)u(x) \geq 0 \text{ in } \Omega,$$

and $u \leq 0$ on $\partial\Omega$ then either $u \equiv 0$ or $u < 0$ in Ω.

The main observation here is that in a narrow domain we need not assume $c \leq 0$ – but "the largest possible narrowness", depends, of course, on the size of the positive part $c^+(x)$ that competes against it.

Proof. Note that, according to the strong maximum principle, it is sufficient to show that $u(x) \leq 0$ in Ω. For the sake of contradiction, suppose that

$$\sup_{x\in\Omega} u(x) > 0. \tag{3.15}$$

Without loss of generality we may assume that e is the unit vector in the direction x_1, and that

$$\bar{\Omega} \subset \{0 < x_1 < d\}.$$

Suppose that d is so small that

$$c(x) \le \pi^2/d^2, \quad \text{for all } x \in \Omega, \tag{3.16}$$

and consider the function

$$w(x) = \sin\big(\frac{\pi x_1}{d}\big).$$

It satisfies

$$\Delta w + \frac{\pi^2}{d^2} w = 0, \tag{3.17}$$

and $w(x) > 0$ in $\bar{\Omega}$, in particular

$$\inf_{\bar{\Omega}} w(x) > 0. \tag{3.18}$$

A consequence of the above is

$$\Delta w + c(x) w \le 0, \tag{3.19}$$

Given $\lambda \ge 0$, let us set $w_\lambda(x) = \lambda w(x)$. As a consequence of (3.18), there exists $\Lambda > 0$ so large that $\Lambda w(x) > u(x)$ for all $x \in \Omega$. Now we are going to push w_λ down until it touches $u(x)$: set

$$\lambda_0 = \inf\{\lambda : \ w_\lambda(x) > u(x) \text{ for all } x \in \Omega.\}$$

Note, that, because of (3.15), we know that $\lambda_0 > 0$. The difference

$$v(x) = u(x) - w_{\lambda_0}(x)$$

satisfies

$$\Delta v + c(x) v \ge 0.$$

The difference between $u(x)$, which satisfies the same inequality, and $v(x)$ is that we know already that $v(x) \le 0$ – hence, we may conclude from the strong maximum principle again that either $v(x) \equiv 0$, or $v(x) < 0$ in Ω. The former contradicts the boundary condition on $u(x)$, as $w_\lambda(x) > 0$ on $\partial\Omega$, hence $v(x) < 0$ in Ω. As $v(x) < 0$ also on the boundary $\partial\Omega$, there exists $\varepsilon_0 > 0$ so that

$$v(x) < -\varepsilon_0 \text{ for all } x \in \bar{\Omega},$$

that is,

$$u(x) + \varepsilon_0 < w_{\lambda_0}(x) \text{ for all } x \in \bar{\Omega}.$$

But then we may choose $\lambda' < \lambda_0$ so that we still have

$$w_{\lambda'}(x) > u(x) \text{ for all } x \in \Omega.$$

This contradicts the minimality of λ_0. Thus, it is impossible that $u(x) > 0$ for some $x \in \Omega$, and we are done. $\Box$

The maximum principle for small domains

The maximum principle for narrow domains can be extended, dropping the requirement that the domain is narrow and replacing it by the condition that the domain has a small volume. We begin with the following lemma, which measures how far from the maximum principle a force can push you.

Lemma 3.8 *(The baby ABP Maximum Principle) Assume that $c(x) \le 0$ for all $x \in \Omega$, and let $u \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfy*

$$\Delta u + c(x)u \ge f \ \text{ in } \Omega, \tag{3.20}$$

and $u \le 0$ on $\partial\Omega$. Then

$$\sup_{\Omega} u \le C \mathrm{diam}(\Omega) \left\| f^- \right\|_{L^n(\Omega)}, \tag{3.21}$$

with the constant C that depends only on the dimension n (but not on the function $c(x) \le 0$).

Proof. The idea is very similar to what we did in the proof of the isoperimetric inequality. If $M := \sup_{\Omega} u \le 0$, then there is nothing to prove, hence we assume that $M > 0$. The maximum is then achieved at an interior point $x_0 \in \Omega$, $M = u(x_0)$, as $u(x) \le 0$ on $\partial\Omega$. Consider the function $v = -u^+$, then $v \le 0$ in Ω, $v \equiv 0$ on $\partial\Omega$ and

$$-M = \inf_{\Omega} v = v(x_0) < 0.$$

We proceed as in the proof of the isoperimetric inequality. Let Γ be the lower contact set of the function v. As $v \le 0$ in Ω, we have $v < 0$ on Γ, hence v is smooth on Γ, and

$$\Delta v = -\Delta u \le -f(x) + c(x)u \le -f(x), \ \text{for } x \in \Gamma, \tag{3.22}$$

as $c(x) \le 0$ and $u(x) \ge 0$ on Γ. The analog of the inclusion (2.9) that we will now prove is

$$B(0; M/d) \subset \nabla v(\Gamma), \tag{3.23}$$

with $d = \mathrm{diam}(\Omega)$ and $B(0, M/d)$ the open ball centered at the origin of radius M/d. One way to see that is by sliding: let $p \in B(0; M/d)$ and consider the hyperplane that is the graph of

$$z_k(x) = p \cdot x - k.$$

Clearly, $z_k(x) < v(x)$ for k sufficiently large. As we decrease k, sliding the plane up, let $\bar{k}$ be the first value when the graphs of $v(x)$ and $z_{\bar{k}}(x)$ touch at a point x_1. Then we have $v(x) \ge z_{\bar{k}}(x)$ for all $x \in \Omega$. If x_1 is on the boundary $\partial\Omega$ then $v(x_1) = z_{\bar{k}}(x_1) = 0$, and we have

$$p \cdot (x_0 - x_1) = z_{\bar{k}}(x_0) - z_{\bar{k}}(x_1) \le v(x_0) - 0 = -M,$$

whence $|p| \ge M/d$, which is a contradiction. Therefore, x_1 is an interior point, which means that $x_1 \in \Gamma$ (by the definition of the lower contact set), and $p = \nabla v(x_1)$. This proves the inclusion (3.23).

Mimicking the proof of the isoperimetric inequality we use the area formula (c_n is the volume of the unit bal in $\mathbb{R}^n$):

$$c_n \left(\frac{M}{d}\right)^n = |B(0; M/d)| \le |\nabla v(\Gamma)| \le \int_{\Gamma} |\det(D^2 v(x))| dx. \tag{3.24}$$

Now, as in the aforementioned proof, for every point x in the contact set Γ, the matrix $D^2v(x)$ is non-negative definite, hence (note that (3.22) implies that $f(x) \leq 0$ on Γ)

$$|\det[D^2v(x)]| \leq \left(\frac{\Delta v}{n}\right)^n \leq \frac{(-f(x))^n}{n^n}. \quad (3.25)$$

Integrating (3.25) and using (3.24), we get

$$M^n \leq \frac{(\mathrm{diam}(\Omega))^n}{c_n n^n} \int_\Gamma |f^-(x)|^n dx, \quad (3.26)$$

which is (3.21). □

An important consequence of Lemma 3.8 is a maximum principle for a domain with a small volume [5].

Theorem 3.9 *(The maximum principle for domains of a small volume) Let* $u \in C^2(\Omega) \cap C(\bar{\Omega})$ *satisfy*

$$\Delta u(x) + c(x)u(x) \geq 0 \text{ in } \Omega,$$

and assume that $u \leq 0$ *on* $\partial\Omega$. *Then there exists a positive constant* δ *which depends on the spatial dimension* n, *the diameter of* Ω, *and* $\|c^+\|_{L^\infty}$, *so that if* $|\Omega| \leq \delta$ *then* $u \leq 0$ *in* Ω.

Proof. If $c \leq 0$ then $u \leq 0$ by the standard maximum principle. In general, assume that $u^+ \not\equiv 0$, and write $c = c^+ - c^-$. We have

$$\Delta u - c^- u \geq -c^+ u.$$

Lemma 3.8 implies that (with a constant C that depends only on the dimension n)

$$\sup_\Omega u \leq C\mathrm{diam}(\Omega)\|c^+u^+\|_{L^n(\Omega)} \leq C\mathrm{diam}(\Omega)\|c^+\|_\infty|\Omega|^{1/n}\sup_\Omega u \leq \frac{1}{2}\sup_\Omega u,$$

when the volume of Ω is sufficiently small:

$$|\Omega| \leq \frac{1}{(2C\mathrm{diam}(\Omega)\|c^+\|_\infty)^n}. \quad (3.27)$$

We deduce that $\sup_\Omega u \leq 0$ contradicting the assumption $u^+ \not\equiv 0$, Hence, we have $u \leq 0$ in Ω under the condition (3.27). □

4 Act IV. Dancing together

We will now use a combination of the maximum principle (mostly for small domains) and the moving plane method to prove some results on the symmetry of the solutions to elliptic problems. We show just the tip of the iceberg – a curious reader will find many other results in the literature, the most famous being, perhaps, the De Giorgi conjecture, a beautiful connection between geometry and applied mathematics.

4.1 The Gidas-Ni-Nirenberg theorem

The following result on the radial symmetry of non-negative solutions is due to Gidas, Ni and Nirenberg. It is a basic example of a general phenomenon that positive solutions of elliptic equations tend to be monotonic in one form or other. We present the proof of the Gidas-Ni-Nirenberg theorem from [20]. The proof uses the moving plane method combined with the maximum principles for narrow domains, and domains of small volume.

Theorem 4.1 *Let $B_1 \in \mathbb{R}^n$ be the unit ball, and $u \in C(\bar{B}_1) \cap C^2(B_1)$ be a positive solution of the Dirichlet boundary value problem*

$$\begin{aligned} &\Delta u + f(u) = 0 \quad \textit{in } B_1, \\ &u = 0 \ \textit{on } \partial B_1, \end{aligned} \tag{4.1}$$

with the function f that is locally Lipschitz in $\mathbb{R}$. Then, the function u is radially symmetric in B_1 and

$$\frac{\partial u}{\partial r}(x) < 0 \textit{ for } x \neq 0.$$

To address an immediate question the reader may have, we give the following simple exercise.

Exercise 4.2 *Show that the conclusion that a function u satisfying (4.1) is radially symmetric is false without the assumption that the function u is positive.*

The proof of Theorem 4.1 is based on the following lemma, which applies to general domains with a planar symmetry, not just balls.

Lemma 4.3 *Let Ω be a bounded domain that is convex in the x_1-direction and symmetric with respect to the plane $\{x_1 = 0\}$. Let $u \in C(\bar{\Omega}) \cap C^2(\Omega)$ be a positive solution of*

$$\begin{aligned} &\Delta u + f(u) = 0 \quad \textit{in } \Omega, \\ &u = 0 \ \textit{on } \partial\Omega, \end{aligned} \tag{4.2}$$

with the function f that is locally Lipschitz in $\mathbb{R}$. Then, the function u is symmetric with respect to x_1 and

$$\frac{\partial u}{\partial x_1}(x) < 0 \textit{ for any } x \in \Omega \textit{ with } x_1 > 0.$$

Proof of Theorem 4.1. Theorem 4.1 follows immediately from Lemma 4.3. Indeed, Lemma 4.3 implies that $u(x)$ is decreasing in any given radial direction, since the unit ball is symmetric with respect to any plane passing through the origin. It also follows from the same lemma that $u(x)$ is invariant under a reflection with respect to any hyperplane passing through the origin – this trivially implies that u is radially symmetric. □

Proof of Lemma 4.3

We use the coordinate system $x = (x_1, y) \in \Omega$ with $y \in \mathbb{R}^{n-1}$. We will prove that

$$u(x_1, y) < u(x_1^*, y) \text{ for all } x_1 > 0 \text{ and } -x_1 < x_1^* < x_1. \tag{4.3}$$

This, obviously, implies monotonicity in x_1 for $x_1 > 0$. Next, letting $x_1^* \to -x_1$, we get the inequality

$$u(x_1, y) \le u(-x_1, y) \text{ for any } x_1 > 0.$$

Changing the direction, we get the reflection symmetry: $u(x_1, y) = u(-x_1, y)$.

We now prove (4.3). Given any $\lambda \in (0, a)$, with $a = \sup_\Omega x_1$, we take the "moving plane"

$$T_\lambda = \{x_1 = \lambda\},$$

and consider the part of Ω that is "to the right" of T_λ:

$$\Sigma_\lambda = \{x \in \Omega : \ x_1 > \lambda\}.$$

Finally, given a point x, we let x_λ be the reflection of $x = (x_1, x_2, \ldots, x_n)$ with respect to T_λ:

$$x_\lambda = (2\lambda - x_1, x_2, \ldots, x_n).$$

Consider the difference

$$w_\lambda(x) = u(x) - u(x_\lambda) \text{ for } x \in \Sigma_\lambda.$$

The mean value theorem implies that w_λ satisfies

$$\Delta w_\lambda = f(u(x_\lambda)) - f(u(x)) = \frac{f(u(x_\lambda)) - f(u(x))}{u(x_\lambda) - u(x)} w_\lambda = -c(x, \lambda) w_\lambda$$

in Σ_λ. This is a recurring trick: the difference of two solutions of a semi-linear equation satisfies a "linear" equation with an unknown function c. However, we know a priori that the function c is bounded:

$$|c(x)| \le \mathrm{Lip}(f), \text{ for all } x \in \Omega. \tag{4.4}$$

The boundary $\partial\Sigma_\lambda$ consists of a piece of $\partial\Omega$, where $w_\lambda = -u(x_\lambda) < 0$ and of a part of the plane T_λ, where $x = x_\lambda$, thus $w_\lambda = 0$. Summarizing, we have

$$\begin{aligned} &\Delta w_\lambda + c(x, \lambda) w_\lambda = 0 \text{ in } \Sigma_\lambda \\ &w_\lambda \le 0 \text{ and } w_\lambda \not\equiv 0 \text{ on } \partial\Sigma_\lambda, \end{aligned} \tag{4.5}$$

with a bounded function $c(x, \lambda)$. As the function $c(x, \lambda)$ does not necessarily have a definite sign, we may not apply the maximum principle and immediately conclude from (4.5) that

$$w_\lambda < 0 \text{ inside } \Sigma_\lambda \text{ for all } \lambda \in (0, a). \tag{4.6}$$

Nevertheless, using the moving plane method, we will be able to show that (4.6) holds. This implies in particular that w_λ assumes its maximum (equal to zero) over $\bar{\Sigma}_\lambda$ along T_λ. The Hopf lemma implies then

$$\left.\frac{\partial w_\lambda}{\partial x_1}\right|_{x_1=\lambda} = 2 \left.\frac{\partial u}{\partial x_1}\right|_{x_1=\lambda} < 0.$$

Given that λ is arbitrary, we conclude that

$$\frac{\partial u}{\partial x_1} < 0, \text{ for any } x \in \Omega \text{ such that } x_1 > 0.$$

Therefore, it remains only to show that $w_\lambda < 0$ inside Σ_λ to establish monotonicity of u in x_1 for $x_1 > 0$. Another consequence of (4.6) is that

$$u(x_1, x') < u(2\lambda - x_1, x') \text{ for all } \lambda \text{ such that } x \in \Sigma_\lambda,$$

that is, for all $\lambda \in (0, x_1)$, which is the same as (4.3).

In order to show that $w_\lambda < 0$ one would like to apply the maximum principle to the boundary value problem (4.5). However, as we have mentioned, a priori the function $c(x, \lambda)$ does not have a sign, so the usual maximum principle may not be used. On the other hand, there exists δ_c such that the maximum principle for narrow domains holds for the operator

$$Lu = \Delta u + c(x)u,$$

and domains of the width not larger than δ_c in the x_1-direction. Note that δ_c depends only on $\|c\|_{L^\infty}$ that is controlled in our case by (4.4). Moreover, when λ is sufficiently close to a:

$$a - \delta_c < \lambda < a,$$

the domain Σ_λ does have the width in the x_1-direction which is smaller than δ_c. Thus, for such λ the maximum principle for narrow domains implies that $w_\lambda < 0$ inside Σ_λ. This is because $w_\lambda \leq 0$ on $\partial\Sigma_\lambda$, and $w_\lambda \not\equiv 0$ on $\partial\Sigma_\lambda$.

Let us now decrease λ (move the plane T_λ to the left, hence the name "the moving plane" method), and let (λ_0, a) be the largest interval of values so that $w_\lambda < 0$ inside Σ_λ for all $\lambda \in (\lambda_0, a)$. If $\lambda_0 = 0$, that is, if we may move the plane T_λ all the way to $\lambda = 0$, while keeping (4.6) true, then we are done – (4.6) follows. Assume, for the sake of a contradiction, that $\lambda_0 > 0$. Then, by continuity, we still know that

$$w_{\lambda_0} \leq 0 \text{ in } \Sigma_{\lambda_0}.$$

Moreover, w_{λ_0} is not identically equal to zero on $\partial\Sigma_{\lambda_0}$. The strong maximum principle implies that

$$w_{\lambda_0} < 0 \text{ in } \Sigma_{\lambda_0}. \tag{4.7}$$

We will show that then

$$w_{\lambda_0 - \varepsilon} < 0 \text{ in } \Sigma_{\lambda_0 - \varepsilon} \tag{4.8}$$

for sufficiently small $\varepsilon < \varepsilon_0$. This will contradict our choice of λ_0 (unless $\lambda_0 = 0$).

Here is the key step and the reason why the maximum principle for domains of small volume is useful for us here: choose a simply connected closed set K in Σ_{λ_0}, with a smooth boundary, which is "nearly all" of Σ_{λ_0}, in the sense that

$$|\Sigma_{\lambda_0} \backslash K| < \delta/2$$

with $\delta > 0$ to be determined. Inequality (4.7) implies that there exists $\eta > 0$ so that

$$w_{\lambda_0} \leq -\eta < 0 \text{ for any } x \in K.$$

By continuity, there exits $\varepsilon_0 > 0$ so that

$$w_{\lambda_0 - \varepsilon} < -\frac{\eta}{2} < 0 \text{ for any } x \in K, \tag{4.9}$$

for $\varepsilon \in (0, \varepsilon_0)$ sufficiently small. Let us now see what happens in $\Sigma_{\lambda_0-\varepsilon} \setminus K$. As far as the boundary is concerned, we have

$$w_{\lambda_0-\varepsilon} \leq 0$$

on $\partial\Sigma_{\lambda_0-\varepsilon}$ – this is true for $\partial\Sigma_\lambda$ for all $\lambda \in (0, a)$, and, in addition,

$$w_{\lambda_0-\varepsilon} < 0 \text{ on } \partial K,$$

because of (4.9) We conclude that

$$w_{\lambda_0-\varepsilon} < 0 \text{ on } \partial(\Sigma_{\lambda_0-\varepsilon} \backslash K).$$

However, when ε is sufficiently small we have $|\Sigma_{\lambda_0-\varepsilon} \backslash K| < \delta$. Choose δ (once again, solely determined by $\|c\|_{L^\infty(\Omega)}$), so small that we may apply the maximum principle for domains of small volume to the function $w_{\lambda_0-\varepsilon}$ in the domain $\Sigma_{\lambda_0-\varepsilon} \backslash K$. Then, we obtain

$$w_{\lambda_0-\varepsilon} \leq 0 \text{ in } \Sigma_{\lambda_0-\varepsilon} \backslash K.$$

The strong maximum principle implies that

$$w_{\lambda_0-\varepsilon} < 0 \text{ in } \Sigma_{\lambda_0-\varepsilon} \backslash K.$$

Putting two and two together we see that (4.8) holds. This, however, contradicts the choice of λ_0. The proof of the Gidas-Ni-Nirenberg theorem is complete. □

4.2 The sliding method

The sliding method differs from the moving plane method in that one compares translations of a function rather than its reflections with respect to a plane. We will illustrate it on an example taken from [20].

Theorem 4.4 *Let Ω be an arbitrary bounded domain in $\mathbb{R}^n$ which is convex in the x_1-direction. Let $u \in C^2(\Omega) \cap C(\bar{\Omega})$ be a solution of*

$$\begin{aligned} &\Delta u + f(u) = 0 \text{ in } \Omega \\ &u = \eta(x) \text{ on } \partial\Omega \end{aligned} \tag{4.10}$$

with a Lipschitz function f. Assume that for any three points $x' = (x_1', y)$, $x = (x_1, y)$, and $x'' = (x_1'', y)$ lying on a segment parallel to the x_1-axis, $x_1' < x_1 < x_1''$ with $x', x'' \in \partial\Omega$, the following hold:

$$\eta(x') < u(x) < \eta(x'') \text{ if } x \in \Omega \tag{4.11}$$

and

$$\eta(x') \leq \eta(x) \leq \eta(x'') \text{ if } x \in \partial\Omega. \tag{4.12}$$

Then u is monotone in x_1 in Ω:

$$u(x_1 + \tau) > u(x_1, y) \text{ for } (x_1, y), (x_1 + \tau, y) \in \Omega \text{ and } \tau > 0.$$

Finally, u is the unique solution of (4.10) in $C^2(\Omega) \cap C(\bar{\Omega})$ satisfying (4.11).

Assumption (4.11) is usually checked in applications from the maximum principle and is not as unverifiable and restrictive in practice as it might seem at a first glance. For instance, consider (4.10) in a rectangle $D = [-a, a]_x \times [0, 1]_y$ with the Dirichlet data

$$\eta(-a, y) = 0, \quad \eta(a, y) = 1,$$

prescribed at the vertical boundaries, while the data prescribed along the horizontal lines $y = 0$ and $y = 1$: $\eta_0(x) = u(x, 0)$ and $\eta_1(x) = u(x, 1)$ are monotonic in x. The function f is assumed to vanish at $u = 0$ and $u = 1$:

$$f(0) = f(1) = 0, \; f(s) \leq 0 \text{ for } u \notin [0, 1].$$

The maximum principle implies that then $0 \leq u \leq 1$ so that both (4.11) and (4.12) hold. Then Theorem 4.4 implies that the solution $u(x, y)$ is monotonic in x.

Proof. The philosophy of the proof is very similar to what we did in the proof of the Gidas-Ni-Nirenberg theorem. For $\tau \geq 0$, we let $u^\tau(x_1, y) = u(x_1 + \tau, y)$ be a shift of u to the left. The function u^τ is defined on the set $\Omega^\tau = \Omega - \tau \mathbf{e}_1$ obtained from Ω by sliding it to the left a distance τ parallel to the x_1-axis. The monotonicity of u may be restated as

$$u^\tau > u \text{ in } D^\tau = \Omega^\tau \cap \Omega \text{ for any } \tau > 0, \tag{4.13}$$

and this is what we will prove. As before, we first establish (4.13) for τ close to the largest value τ_0 – that is, those that have been slid almost all the way to the left, and the domain D^τ is both narrow and small. This will be done using the maximum principle for domains of a small volume. Then we will start decreasing τ, sliding the domain Ω^τ to the right, and will show that you may go all the way to $\tau = 0$, keeping (4.13) enforced.

Consider the function

$$w^\tau(x) = u^\tau(x) - u(x) = u(x_1 + \tau, y) - u(x_1, y),$$

defined in D^τ. Since u^τ satisfies the same equation as u, we have from the mean value theorem

$$\begin{aligned} &\Delta w^\tau + c^\tau(x) w^\tau = 0 \text{ in } D^\tau \\ &w^\tau \geq 0 \text{ on } \partial D^\tau \end{aligned} \tag{4.14}$$

where

$$c^\tau(x) = \frac{f(u^\tau(x)) - f(u(x))}{u^\tau(x) - u(x)}$$

is a uniformly bounded function:

$$|c^\tau(x)| \leq \mathrm{Lip}(f). \tag{4.15}$$

The inequality on the boundary ∂D^τ in (4.14) follows from assumptions (4.11) and (4.12). Let

$$\tau_0 = \sup\{\tau > 0 : \; D^\tau \neq \emptyset\}$$

be the largest shift of Ω to the left that we can make so that Ω and Ω^τ still have a non-zero intersection. The volume $|D^\tau|$ is small when τ is close to τ_0. As in the moving plane method, since the function $c^\tau(x)$ is uniformly bounded by (4.15), we may apply the maximum principle for small domains to w^τ in D^τ for τ close to τ_0, and conclude that $w^\tau > 0$ for such τ.

Then we start sliding Ω^τ back to the right, that is, we decrease τ from τ_0 to a critical position τ_1: let (τ_1, τ_0) be a maximal interval with $\tau_1 \geq 0$ so that

$$w^\tau \geq 0 \text{ in } D^\tau \text{ for all } \tau \in (\tau_1, \tau_0].$$

We want to show that $\tau_1 = 0$ and argue by contradiction assuming that $\tau_1 > 0$.

Continuity implies that $w^{\tau_1} \geq 0$ in D^{τ_1}. Furthermore, (4.11) implies that

$$w^{\tau_1}(x) > 0 \text{ for all } x \in \Omega \cap \partial D^{\tau_1}.$$

The strong maximum principle then implies that $w^{\tau_1} > 0$ in D^{τ_1}.

Now we use the same idea as in the proof of Lemma 4.3: choose $\delta > 0$ so that the maximum principle holds for any solution of (4.14) in a domain of volume less than δ. Carve out of D^{τ_1} a closed set $K \subset D^{\tau_1}$ so that

$$|D^{\tau_1} \backslash K| < \delta/2.$$

We know that $w^{\tau_1} > 0$ on K, hence, as the set K is compact, $\inf_K w^{\tau_1}(x) > 0$. Thus, for $\varepsilon > 0$ sufficiently small, the function $w^{\tau_1-\varepsilon}$ is also positive on K. Moreover, for $\varepsilon > 0$ small, we have

$$|D^{\tau_1-\varepsilon} \backslash K| < \delta.$$

Furthermore, since

$$\partial(D^{\tau_1-\varepsilon} \backslash K) \subset \partial D^{\tau_1-\varepsilon} \cup K,$$

we see that

$$w^{\tau_1-\varepsilon} \geq 0 \text{ on } \partial(D^{\tau_1-\varepsilon} \backslash K).$$

Thus, $w^{\tau_1-\varepsilon}$ satisfies

$$\begin{aligned} &\Delta w^{\tau_1-\varepsilon} + c^{\tau_1-\varepsilon}(x) w^{\tau_1-\varepsilon} = 0 \text{ in } D^{\tau_1-\varepsilon} \backslash K \\ &w^{\tau_1-\varepsilon} \geq 0 \text{ on } \partial(D^{\tau_1-\varepsilon} \backslash K). \end{aligned} \tag{4.16}$$

The maximum principle for domains of small volume implies that

$$w^{\tau_1-\varepsilon} \geq 0 \text{ on } D^{\tau_1-\varepsilon} \backslash K.$$

Hence, we have

$$w^{\tau_1-\varepsilon} \geq 0 \text{ in all of } D^{\tau_1-\varepsilon},$$

and, as

$$w^{\tau_1-\varepsilon} \not\equiv 0 \text{ on } \partial D^{\tau_1-\varepsilon},$$

it is positive in $D^{\tau_1-\varepsilon}$. However, this contradicts the choice of τ_1. Therefore, $\tau_1 = 0$ and the function u is monotone in the x_1-variable.

Finally, to show that such solution u is unique, we suppose that v is another solution. We argue exactly as before but with $w^\tau = u^\tau - v$. The same proof shows that $u^\tau \geq v$ for all $\tau \geq 0$. In particular, $u \geq v$. Interchanging the role of u and v we conclude that $u = v$. □

Another beautiful application of the sliding method allows to extend lower bounds obtained in one part of a domain to a different part by moving a sub-solution around the domain and observing that it may never touch a solution. This is a very simple and powerful tool in many problems.

Lemma 4.5 *Let u be a positive function in an open connected set D satisfying*

$$\Delta u + f(u) \leq 0 \text{ in } D$$

with a Lipschitz function f. Let B be a ball with its closure $\bar{B} \subset D$, and suppose z is a function in $\bar{B}$ satisfying

$$\begin{aligned}
&z \leq u \text{ in } B \\
&\Delta z + f(z) \geq 0, \text{ wherever } z > 0 \text{ in } B \\
&z \leq 0 \text{ on } \partial B.
\end{aligned}$$

Then for any continuous one-parameter family of Euclidean motions (rotations and translations) $A(t)$, $0 \leq t \leq T$, so that $A(0) = \mathrm{Id}$ and $A(t)\bar{B} \subset D$ for all t, we have

$$z_t(x) = z(A(t)^{-1}x) < u(x) \text{ in } B_t = A(t)B. \tag{4.17}$$

Proof. The rotational invariance of the Laplace operator implies that the function z_t satisfies

$$\begin{aligned}
&\Delta z_t + f(z_t) \geq 0, \text{ wherever } z_t > 0 \text{ in } B_t \\
&z_t \leq 0 \text{ on } \partial B_t.
\end{aligned}$$

Thus the difference $w_t = z_t - u$ satisfies $\Delta w_t + c_t(x) w_t \geq 0$ wherever $z_t > 0$ in B_t with c_t bounded in B_t, where, as always,

$$c_t(x) = \begin{cases} \dfrac{f(z_t(x)) - f(u(x))}{z_t(x) - u(x)}, & \text{if } z_t(x) \neq u(x) \\ 0, & \text{otherwise.} \end{cases}$$

In addition, $w_t < 0$ on ∂B_t.

We now argue by contradiction. Suppose that there is a first t so that the graph of z_t touches the graph of u at a point x_0. Then, for that t, we still have $w_t \leq 0$ in B_t, but also $w_t(x_0) = 0$. As $u > 0$ in D, and $z_t \leq 0$ on ∂B_t, the point x_0 has to be inside B_t, which means that z_t satisfies

$$z_t + f(z) \geq 0$$

in a neighborhood of x_0. The maximum principle implies then that $w_t \equiv 0$ in the whole component G of the set of points in B_t where $z_t > 0$ that contains x_0. Consequently, $w_t(\widetilde{x}) = 0$ for all $\widetilde{x} \in \partial G$. But then $z_t(\widetilde{x}) = u(\widetilde{x}) > 0$ on ∂G, which contradicts the fact that $z_t = 0$ on ∂G. Hence the graph of z_t may not touch that of u and (4.17) follows. □

Lemma 4.5 is often used to "slide around" a sub-solution that is positive somewhere to show that solution itself is uniformly positive.

5 Monotonicity in unbounded domains

We now consider the monotonicity properties of bounded solutions of

$$\Delta u + f(u) = 0 \text{ in } \Omega \tag{5.1}$$

when the domain Ω is not bounded, so that monotonicity may not be "forced" on the solution as in (4.11)-(4.12). We will consider two examples, the first one is in the whole space, and is part of very deep mathematics but the version we present is relatively simple – the main result is that solution of a semi-linear equation depends only on one variable. The second example is more technically difficult – it addresses domains bounded by a graph of a function and shows monotonicity in any direction that does not touch the graph.

5.1 Monotonicity in $\mathbb{R}^n$

Our first example taken from the paper [15] by Berstycki, Hamel and Monneau deals with the whole space. We consider solutions of

$$\Delta u + f(u) = 0 \text{ in } \mathbb{R}^n \tag{5.2}$$

which satisfy $|u| \leq 1$ together with the asymptotic conditions

$$u(x', x_n) \to \pm 1 \text{ as } x_n \to \pm\infty \text{ uniformly in } x' = (x_1, \dots, x_{n-1}). \tag{5.3}$$

The given function f is Lipschitz-continuous on $[-1, 1]$. We assume that there exists $\delta > 0$ so that

$$f \text{ is non-increasing on } [-1, -1+\delta] \text{ and on } [1-\delta, 1]; \text{ and } f(\pm 1) = 0. \tag{5.4}$$

The prototypical example is $f(u) = u - u^3$. This problem appears in many applications, ranging from biology and combustion to the differential geometry. The main feature of the nonlinearity is that an ODE

$$\frac{du}{dt} = f(u) \tag{5.5}$$

has two stable solutions $u = -1$ and $u = 1$. Solutions of the partial differential differential equation (5.2) describe diffusive transitions between regions in space where u is close to (-1) and those where u is close to $+1$. The prototypical example is the solution of

$$u_0'' + f(u_0) = 0 \text{ in } \mathbb{R}.\ u_0(\pm\infty) = \pm 1.$$

This equation may be solved explicitly: multiplying (5.6) by u_0' and integrating from $-\infty$ to x, using the boundary conditions, leads to

$$\frac{1}{2}(u_0')^2 + F(u_0) = 0, \quad u_0(\pm\infty) = \pm 1. \tag{5.6}$$

Here, we have defined

$$F(s) = \int_{-1}^{s} f(u)du. \tag{5.7}$$

A necessary condition for a solution of (5.6) to exist is that $F(1) = 0$, or

$$\int_{-1}^{1} f(u)du = 0. \tag{5.8}$$

Exercise 5.1 Find a sufficient condition on the nonlinearity $f(u)$ for a monotonically increasing solution of (5.6) to exist.

Rather than study the existence question, we will assume that (5.2) has a solution, and show that the asymptotic conditions (5.3) imply that the solution is actually one-dimensional.

Theorem 5.2 *Let u be any solution of (5.2)-(5.3) such that $|u| \leq 1$. Then $u(x', x_n) = u_0(x_n)$ where u_0 is a solution of*

$$u_0'' + f(u_0) = 0 \text{ in } \mathbb{R},\; u_0(\pm\infty) = \pm 1.$$

Moreover, u is increasing with respect to x_n. Finally, such solution is unique up to a translation.

Without the uniformity assumption in (5.3) this is known as "the weak form" of the De Giorgi conjecture, and was resolved by Savin [108] who showed that all solutions are one-dimensional in $n \leq 8$, and del Pino, Kowalczyk and Wei [44] who showed that non-planar solutions exist $n \geq 9$. The additional assumption of uniform convergence at infinity made in this section makes this question much easier. The full De Giorgi conjecture is that any solution of (5.4) in dimension $n \leq 8$ with $f(u) = u - u^3$ such that $-1 \leq u \leq 1$ is one-dimensional. It is still open in this generality, to the best of our knowledge. The motivation for the conjecture comes from the study of the minimal surfaces in differential geometry but we will not discuss this connection here.

First, we state a version of the maximum principle for unbounded domains.

Lemma 5.3 *Let D be an open connected set in $\mathbb{R}^n$, possibly unbounded. Assume that $\bar{D}$ is disjoint from the closure of an infinite open (solid) cone Σ. Suppose that a function $z \in C(\bar{D})$ is bounded from above and satisfies*

$$\begin{aligned} &\Delta z + c(x) z \geq 0 \text{ in } D \\ &z \leq 0 \text{ on } \partial D. \end{aligned} \tag{5.9}$$

with some continuous function $c(x) \leq 0$, then $z \leq 0$.

Proof. If the function $z(x)$ would, in addition, vanish at infinity:

$$\limsup_{|x|\to+\infty} z(x) = 0, \tag{5.10}$$

then the proof would be easy. Indeed, if (5.10) holds then we can find a sequence $R_n \to +\infty$ so that

$$\sup_{\bar{D}\cap\{|x|=R_n\}} z(x) \leq \frac{1}{n}. \tag{5.11}$$

The usual maximum principle in the domain $D_n = D \cap B(0; R_n)$ imples that $z(x) \leq 1/n$ in D_n. Letting $n \to \infty$ gives

$$z(x) \leq 0 \text{ in } D.$$

Our next task is to reduce the case of a bounded function z to (5.11). To do this we will construct a harmonic function $g(x) > 0$ in D such that

$$|g(x)| \to +\infty \text{ as } |x| \to +\infty. \tag{5.12}$$

Since g is harmonic, the ratio $\sigma = z/g$ will satisfy a differential inequality in D:

$$\Delta\sigma + \frac{2}{g}\nabla g \cdot \nabla\sigma + c\sigma \geq 0.$$

This is similar to (5.9) but now σ does satisfy the asymptotic condition

$$\limsup_{x \in D, |x| \to \infty} \sigma(x) \leq 0,$$

uniformly in $x \in D$. Moreover, $\sigma \leq 0$ on ∂D. Hence one may apply the above argument to the function $\sigma(x)$, and conclude that $\sigma(x) \leq 0$, which, in turn, implies that $z(x) \leq 0$ in D.

In order to construct such harmonic function $g(x)$ in D, the idea is to decrease the cone Σ to a cone $\widetilde{\Sigma}$ and to consider the principal eigenfunction ψ of the spherical Laplace-Beltrami operator in the region $G = \mathbb{S}^{n-1} \backslash \widetilde{\Sigma}$ with $\psi = 0$ on ∂G:

$$\begin{aligned} &\Delta_S \psi + \mu\psi = 0, \quad \psi > 0 \text{ in } G, \\ &\psi = 0 \text{ on } \partial G. \end{aligned}$$

Note that the eigenvalue $\mu > 0$. Then, going to the polar coordinates $x = r\xi$, $r > 0$, $\xi \in \mathbb{S}^{n-1}$, we set $g(x) = r^\alpha \psi(\xi)$, $\xi \in G$, defined on D, with

$$\alpha(n + \alpha - 2) = \mu.$$

With this choice of α, the function g is harmonic:

$$\Delta g = \frac{\partial^2 g}{\partial r^2} + \frac{n-1}{r}\frac{\partial g}{\partial r} + \frac{1}{r^2}\Delta_S g = [\alpha(\alpha - 1) + \alpha(n-1) - \mu] r^{\alpha - 2}\Psi = 0.$$

Moreover, as $\mu > 0$ (the operator $(-\Delta_S)$ is positive), we have $\alpha > 0$, thus (5.12) also holds, and the proof is complete. □

This lemma will be most important in the proof of the Berestycki-Caffarelli-Nirenberg result later on. For now we will need the following corollary that we will use for half-spaces.

Corollary 5.4 *Let f be a Lipschitz continuous function, non-increasing on $[-1, -1+\delta]$ and on $[1-\delta, 1]$ for some $\delta > 0$. Assume that u_1 and u_2 satisfy*

$$\Delta u_i + f(u_i) = 0 \text{ in } \Omega$$

and are such that $|u_i| \leq 1$. Assume furthermore that $u_2 \geq u_1$ on $\partial\Omega$ and that either $u_2 \geq 1-\delta$ or $u_1 \leq -1+\delta$ in Ω. If $\Omega \subset \mathbb{R}^n$ is an open connected set so that $\mathbb{R}^n \backslash \bar{\Omega}$ contains an open infinite cone then $u_2 \geq u_1$ in Ω.

Proof. Assume, for instance, that $u_2 \geq 1-\delta$, and set $w = u_1 - u_2$. Then

$$\Delta w + c(x, z)w = 0 \text{ in } \Omega$$

with

$$c(x) = \frac{f(u_1) - f(u_2)}{u_1 - u_2} \leq 0 \text{ where } w \geq 0.$$

Hence if the set $G = \{w > 0\}$ is not empty, we may apply the maximum principle of Lemma 5.3 to the function w in G (note that $w = 0$ on ∂G), and conclude that $w \leq 0$ in G giving a contradiction. □

Proof of Theorem 5.2

We are going to prove that

$$u \text{ is increasing in any direction } \nu = (\nu_1, \ldots, \nu_n) \text{ with } \nu_n > 0. \tag{5.13}$$

This will mean that

$$\frac{1}{\nu_n}\frac{\partial u}{\partial \nu} = \frac{\partial u}{\partial x_n} + \sum_{j=1}^{n-1} \alpha_j \frac{\partial u}{\partial x_j} > 0$$

for any choice of $\alpha_j = \nu_j/\nu_n$. It follows that all $\partial u/\partial x_j = 0$, $j = 1, \ldots, n-1$, so that u depends only on x_n, and, moreover, $\partial u/\partial x_n > 0$. Hence, (5.13) implies the conclusion of Theorem 5.2 on the monotonicity of the solution.

We now prove (5.13). Monotonicity in the direction ν can be restated as

$$u^t(x) \geq u(x), \text{ for all } t \geq 0 \text{ and all } x \in D, \tag{5.14}$$

where $u^t(x) = u(x + t\nu)$ are the shifts of the function u in the direction ν. We start the sliding method with a very large t. Observe that there exists a real $a > 0$ so that

$$u(x', x_n) \geq 1 - \delta \text{ for all } x_n \geq a,$$

and

$$u(x', x_n) \leq -1 + \delta \text{ for all } x_n \leq -a.$$

Take $t \geq 2a/\nu_n$, then the functions u and u^t are such that

$$\begin{aligned} u^t(x', x_n) &\geq 1 - \delta && \text{for all } x' \in \mathbb{R}^{n-1} \text{ and for all } x_n \geq -a \\ u(x', x_n) &\leq -1 + \delta && \text{for all } x' \in \mathbb{R}^{n-1} \text{ and for all } x_n \leq -a \\ u^t(x', -a) &\geq u(x', -a) && \text{for all } x' \in \mathbb{R}^{n-1}. \end{aligned} \tag{5.15}$$

Hence, we may apply Corollary 5.4 separately in $\Omega_1 = \mathbb{R}^{n-1} \times (-\infty, -a)$ and $\Omega_2 = \mathbb{R}^{n-1} \times (-a, +\infty)$. In both cases, we conclude that $u^t \geq u$ and thus

$$u^t \geq u \text{ in all of } \mathbb{R}^n \text{ for } t \geq 2a/\nu_n.$$

Following the sliding method, we start to decrease t, and let

$$\tau = \inf\{t > 0,\ u^t \geq u \text{ in } \mathbb{R}^n\}.$$

By continuity, we still have $u^\tau \geq u$ in $\mathbb{R}^n$. We need to show that $\tau = 0$, and argue by contradiction. Assume that $\tau > 0$ and consider two cases.

Case 1. Suppose that

$$\inf_{D_a}(u^\tau - u) > 0, \quad D_a = \mathbb{R}^{n-1} \times [-a, a]. \tag{5.16}$$

The function u is globally Lipschitz continuous – this follows from the standard elliptic estimates [69]. This implies that there exists $\eta_0 > 0$ so that for all $\tau - \eta_0 < t < \tau$ we still have

$$u^t(x', x_n) > u(x', x_n) \text{ for all } x' \in \mathbb{R}^{n-1} \text{ and for all } -a \leq x_n \leq a. \tag{5.17}$$

As $u(x', x_n) \geq 1 - \delta$ for all $x_n \geq a$, it follows that

$$u^t(x', x_n) \geq 1 - \delta \text{ for all } x_n \geq a \text{ and } t > 0. \tag{5.18}$$

We may then apply Corollary 5.4 in the half-spaces $\{x_n > a\}$ and $\{x_n < -a\}$ to conclude that

$$u^{\tau-\eta}(x) > u(x)$$

everywhere in $\mathbb{R}^n$ for all $\eta \in [0, \eta_0]$. This contradicts the choice of τ. Thus, the case (5.16) is impossible.

Case 2. Suppose that

$$\inf_{D_a}(u^\tau - u) = 0, \quad D_a = \mathbb{R}^{n-1} \times [-a, a]. \tag{5.19}$$

This would be a contradiction to the maximum principle if we could conclude from (5.19) that the graphs of u^τ and u touch at one internal point. This, however, is not clear, as there may exist a sequence of points ξ_k with $|\xi_k| \to +\infty$, such that $u^\tau(\xi_k) - u(\xi_k) \to 0$, without the graphs ever touching. In order to deal with this issue, we will use the usual trick of moving "the interesting part" of the domain to the origin and passing to the limit. We know from (5.19) that there exists a sequence $\xi_k \in D_a$ so that

$$u^\tau(\xi_k) - u(\xi_k) \to 0 \text{ as } k \to \infty.$$

Let us re-center: set

$$u_k(x) = u(x + \xi_k).$$

Then the standard elliptic regularity estimates imply that $u_k(x)$ converge along a subsequence to a function $u_\infty(x)$, uniformly on compact sets. We have

$$u^\tau_\infty(0) = u_\infty(0),$$

and

$$u^\tau_\infty(x) \geq u_\infty(x), \quad \text{for all } x \in \mathbb{R}^n,$$

because $u^\tau_k \geq u_k$ for all k. The strong maximum principle implies that $u^\tau_\infty = u_\infty$, that is,

$$u_\infty(x + \tau\nu) = u_\infty(x),$$

that is, the function u_∞ is periodic in the ν-direction. However, as all $\xi_k \in D_a$, their n-th components are uniformly bounded $|(\xi_k)_n| \leq a$. Therefore, when we pass to the limit we do not lose the boundary conditions in x_n: the function u_∞ must satisfy the boundary conditions (5.3). This is a contradiction. Hence, this case is also impossible, and thus $\tau = 0$. This proves monotonicity of $u(x)$ in x_n and the fact that u depends only on x_n: $u(x) = u(x_n)$.

In order to prove the uniqueness of such solution, assuming there are two such solutions u and v, one repeats the sliding argument above but applied to the difference

$$w^\tau(x_n) = u(x_n + \tau) - v(x).$$

The same sliding argument will now imply that $u(x_n + \tau) \geq v(x_n)$ for all $x_n \in \mathbb{R}$ and all $\tau \geq 0$, meaning that, in particular, $u(x_n) \geq v(x_n)$. Reversing the role of u and v we will conclude that $u(x_n) = v(x_n)$, showing uniqueness of such solution. □

5.2 Monotonicity in general unbounded domains

We now consider the monotonicity properties of the bounded solutions of

$$\begin{aligned} &\Delta u + f(u) = 0 \text{ in } \Omega \\ &u = 0 \text{ on } \partial\Omega, \end{aligned} \tag{5.20}$$

when the domain Ω is not all of $\mathbb{R}^n$ but it is not bounded so that monotonicity may not be "forced" on the solution as in (4.11)-(4.12). We assume that the solution u is uniformly bounded: $0 < u \leq M < \infty$ in Ω and the domain Ω is defined by

$$\Omega = \{x \in \mathbb{R}^n : \ x_n > \phi(x_1, \ldots, x_{n-1})\}. \tag{5.21}$$

For simplicity, we assume that $\phi : \mathbb{R}^{n-1} \to \mathbb{R}$ is a smooth, globally Lipschitz function. An interested reader should consult [13] for the additional slightly technical arguments required if we only assume that ϕ is a globally Lipschitz continuous function. A typical example would be a half-space Ω – the main result we are going to prove says that u has to be a monotonic function of the single variable x_n in this case.

We will assume that f is Lipschitz continuous on $\mathbb{R}^+$, $f(s) > 0$ on $(0,1)$ and $f(s) \leq 0$ for $s \geq 1$. Furthermore, we assume that f satisfies

$$f(s) \geq \delta_0 s \text{ on } [0, s_0] \text{ for some } s_0 > 0, \tag{5.22}$$

and

$$\text{there exists } s_1 \text{ so that } f \text{ is non-increasing on } (s_1, 1). \tag{5.23}$$

The prototypical example is[3] $f(s) = s(1-s)$. The main result of Berestycki, Caffarelli and Nirenberg in [13] says that such u is unique, monotonic in x_n and tends to one as distance to the boundary tends to infinity. Note that unless the boundary is flat, that is, it has the form $\partial\Omega = \{x_n = c_0\}$ with some $c_0 \in \mathbb{R}$, there is no reason to expect that the solution will depend only on x_n.

Theorem 5.5 *The function u has the following properties:*
(i) it is monotonic with respect to x_n:

$$\frac{\partial u}{\partial x_n} > 0 \ \text{in } \Omega,$$

(ii) $0 < u < 1$ *in* Ω
(iii) $u(x) \to 1$ *as* $\mathrm{dist}(x, \partial\Omega) \to \infty$, *uniformly in* Ω.
(iv) u is the unique bounded solution of (5.20) that is positive inside Ω.
(v) Let κ be the Lipschitz constant of the graph function ϕ, then given any collection of constants a_j, $j = 1, \ldots, n-1$ so that $\sum_{j=1}^{n-1} a_j^2 < \frac{1}{\kappa^2}$, we have

$$\frac{\partial u}{\partial x_n} + \sum_{j=1}^{n-1} a_j \frac{\partial u}{\partial x_j} > 0 \ \text{in } \Omega. \tag{5.24}$$

[3]Such nonlinearities arise naturally in reaction-diffusion modeling and are known as nonlinearities of the Fisher-KPP (for Kolmogorov, Petrovskii, Piskunov) type. Apart from the original papers [63, 82] which are both masterpieces, good recent introductions to reaction-diffusion problems are the books [14, 118], and the review [117], where many more references can be found.

Part (v) implies that u is increasing in any direction ξ such that there exists an orthonormal change of variables $x \to z$ with the z_n-axis in the direction of ξ and $\partial\Omega = \{z_n = \widetilde{\phi}(z')\}$ with a smooth function $\widetilde{\phi}$.

When $\phi = 0$, that is, when Ω is a half-space, the constants α_j in (5.24) may be arbitrary which immediately implies that

$$\frac{\partial u}{\partial x_j} = 0 \text{ for all } j = 1, \dots, n-1,$$

so that u has to be a function of x_n only in this case.

Let us explain heuristically why the limit in (iii) holds. Under our assumptions on the function f, the ODE

$$\dot{u} = f(u)$$

has two steady states: $u = 0$ is unstable,while $u = 1$ is stable. Solutions of the elliptic problem (5.20) can be thought of as steady solutions of the parabolic problem

$$v_t = \Delta v + f(v) \text{ in } \Omega \tag{5.25}$$
$$v = 0 \text{ on } \partial\Omega,$$
$$v(0, x) = v_0(x). \tag{5.26}$$

The parabolic problem inherits from the ODE the stability of the steady state $v = 1$. The boundary condition $v = 0$ on $\partial\Omega$ prevents v from being close to one near the boundary but far away from the boundary its effect is weak, hence solutions tend to one as both distance from the boundary and time tend to infinity. This, in turn, is reflected in the behavior of the solutions of the elliptic problem as $|x| \to +\infty$.

Outline of the proof

The proof of Theorem 5.5 is fairly long and we prove each part separately. The general flow is as follows. First, one uses the maximum principle of Lemma 5.3 to show that $0 < u < 1$, so that (ii) holds. Second, we show that $f(u) \to 0$ as $\mathrm{dist}(x, \partial\Omega)) \to \infty$ – roughly speaking, because otherwise u would satisfy

$$\Delta u < -\varepsilon_0, \tag{5.27}$$

at infinity, with some $\varepsilon_0 > 0$ which is impossible as $0 < u < 1$.

It is easy to conclude from $f(u) \to 0$ that $u \to 1$. In the third step, uniqueness is proved by the sliding method. Finally, monotonicity is established by constructing a solution that is positive and monotonic. Uniqueness implies that the original solution coincides with that one and hence is itself monotonic. Such solution is constructed first on bounded domains and then we pass to the limit of the full domain. The tricky part is to make sure that the limit is positive – this is done by ensuring that solution we construct stays above u.

Proof of (ii) in Theorem 5.5

Let us assume that $u > 1$ somewhere and let D be a connected component of the set $\{u > 1\}$. The set D lies outside an open cone since the function ϕ that defines the boundary $\partial\Omega$ is

Lipschitz. Consider the function $z = u - 1$ in D. It satisfes

$$\Delta z = -f(u) \geq 0 \text{ in } D,$$

as $f(u) \leq 0$ in D. Furthermore, z vanishes on ∂D and is bounded in D. Thus, Lemma 5.3 implies that $z \leq 0$ in D which is a contradiction. Therefore, we have $u \leq 1$ in Ω. If $u(x_0) = 1$ for some $x_0 \in \Omega$, the function $z = u - 1$ satisfies $z \leq 0$ in Ω, $z(x_0) = 0$ and

$$\Delta z + c(x) z = 0 \text{ in } \Omega,$$

with

$$c(x) = \begin{cases} \dfrac{f(u(x))}{u(x) - 1}, & \text{if } u(x) < 1 \\ 0, & \text{if } u(x) = 1. \end{cases}$$

The function $c(x)$ is bounded, hence the strong maximum principle implies that $z \equiv 0$ in Ω which contradicts the fact that $z = -1$ on $\partial\Omega$. $\square$

Proof of (iii) in Theorem 5.5

The proof that $u(x) \to 1$ as $\operatorname{dist}(x, \partial\Omega) \to \infty$ is in two steps. First, we show that u is bounded away from zero at a fixed distance away from the boundary: $u(x) \geq \varepsilon_1$ if $\operatorname{dist}(x, \partial\Omega) > R_0$. Second, we show that $f(u(x)) \to 0$ as $\operatorname{dist}(x, \partial\Omega) \to \infty$. This implies that $u \to 1$, as u is bounded away from zero in this region, and $u = 0$ and $u = 1$ are the only zeros of $f(u)$ in the interval $0 \leq u \leq 1$.

Step 1: u is strictly positive away from the boundary.

Lemma 5.6 *There exist $\varepsilon_1 > 0$ and $R_0 > 0$ so that*

$$u(x) > \varepsilon_1 \text{ if } \operatorname{dist}(x, \partial\Omega) > R_0. \tag{5.28}$$

Proof. Let R_0 be so large that the principle eigenvalue λ_1 of the Dirichlet Laplacian in a ball $B(0; R_0)$ of radius R_0 is smaller than the constant δ_0 in (5.22), so that $f(u) \geq \lambda_1 u$ for $u \in [0, s_0]$. Let ϕ_1 be the corresponding positive eigenfunction with $\max \phi_1 = 1$:

$$\begin{aligned} &-\Delta\phi_1 = \lambda_1 \phi_1 \text{ in } B(0; R_0), \\ &\phi_1 = 0 \text{ on } \partial B(0; R_0), \\ &\phi_1 > 0 \text{ in } B(0; R_0), \\ &\max_{x \in B(0;R_0)} \phi_1(x) = 1. \end{aligned} \tag{5.29}$$

Then the function $z = \varepsilon\phi_1$ is a sub-solution of our equation for $0 < \varepsilon \leq s_0$, that is,

$$\begin{aligned} &\Delta z + f(z) \geq 0 \text{ in } B(0; R_0) \\ &z = 0 \text{ on } \partial B(0; R_0). \end{aligned} \tag{5.30}$$

Let us choose $a \in \Omega$ "large enough" so that $\overline{B(a; R_0)} \subset \Omega$ and set

$$\varepsilon_0 = \min_{\bar{B}(a;R_0)} u.$$

The maximum principle implies that $\varepsilon_0 > 0$. We set $\varepsilon_1 = \min(\varepsilon_0, s_0)$, then, as $\phi_1 \le 1$ we have

$$\varepsilon_1\phi_1(x-a) \le u(x) \text{ in } B(a; R_0).$$

We may now slide the ball $B(a; R_0)$ around the domain Ω and use Lemma 4.5 to deduce that

$$\varepsilon_1\phi_1(x-y) \le u(x) \text{ in } B(y; R_0)$$

for all $y \in \Omega$ with $\text{dist}(y, \partial\Omega) > R_0$. In particular, $u(y) > \varepsilon_1$ for such y. □

We see from the proof that the distance R_0 depends only on the function $f(s)$, though the constant ε_1 in the above proof depends on the function u. However, we will next show that $u(x) \to 1$ as $\text{dist}(x, \partial\Omega) \to +\infty$. Therefore, we may choose the center a in the proof of Lemma 5.6 sufficiently far from the boundary so that

$$\min_{x \in B(a,R_0)} u(x) > s_0.$$

This will allow us to set $\varepsilon_1 = s_0$, hence both R_0 and ε_1 in the statement of Lemma 5.6 depend only on the function $f(s)$.

Step 2: the nonlinearity vanishes at infinity. As we have explained above, in order to complete the proof of part (iii) of Theorem 5.5 we show that $f(u(x)) \to 0$ as $\text{dist}(x, \partial\Omega) \to \infty$. Once again, the heuristic reason is that the function u can not have a uniformly negative Laplacian in too big a region without violating the condition $0 < u < 1$. Here is how that is formalized. Let $v(x)$ be the solution of

$$\begin{aligned} &-\Delta v = 1 \text{ in } B(0;1) \\ &v = 0 \text{ on } \partial B(0;1). \end{aligned}$$

It is given explicitly by

$$v(x) = \frac{1-|x|^2}{2n},$$

with

$$\max_{B(0;1)} v = v(0) = 1/(2n).$$

Let also $|y| \ge R_0$ with R_0 as in the previous lemma, and set

$$\gamma(y) = \min\{f(s) : \ s \in [\varepsilon_1, u(y)]\}.$$

Here ε_1 is also as in (5.28). We claim that

$$\gamma(y) \le \frac{2n}{[\text{dist}(y, \partial\Omega) - R_0]^2}. \tag{5.31}$$

This estimate immediately implies that $f(u(x)) \to 0$ as $\text{dist}(x, \partial\Omega) \to +\infty$.

In order to prove(5.31) we argue by contradiction: suppose that (5.31) fails, that is,

$$\gamma(y_0) > \frac{2n}{[\text{dist}(y_0, \partial\Omega) - R_0]^2}$$

for some y_0 with $\mathrm{dist}(y_0, \partial\Omega) > R_0$. Fix $R < \mathrm{dist}(y_0, \partial\Omega) - R_0$ so that

$$\frac{\gamma(y_0)}{2n} > \frac{1}{R^2}. \tag{5.32}$$

The function u cannot have a local minimum at y_0 since

$$\Delta u(y_0) = -f(y_0) < 0.$$

Thus, we may find y_1 close to y_0 so that $u(y_1) < u(y_0)$ and $\mathrm{dist}(y_1, \partial\Omega) > R_0 + R$. Lemma 5.6 implies that $u \geq \varepsilon_1$ in $B := B(y_1; R)$. Let

$$z(y) = \gamma(y_0) R^2 v\left(\frac{y - y_1}{R}\right),$$

then $\max_B z = \gamma(y_0)R^2/(2n)$ and z satisfies

$$\begin{aligned} -\Delta z &= \gamma(y_0) \text{ in } B \\ z &= 0 \text{ on } \partial B. \end{aligned} \tag{5.33}$$

Now we do "ballooning" (as opposed to "sliding") of z: let $z^\tau(x) = \tau z(x)$. Then for $\tau > 0$ small we have

$$z^\tau(x) < \varepsilon_1 \leq u(x) \text{ in } B.$$

As we increase τ, there is the first value τ_0 so that the graph of z^{τ_0} touches the graph of $u(y)$ at some point x_0. Since $z = 0$ on ∂B, x_0 has to be inside B, hence $u(x_0) > \varepsilon_1$. Also, as $z^{\tau_0}(y_1) \leq u(y_1)$, we have

$$u(x_0) = \tau_0 z(x_0) \leq \tau_0 z(y_1) = \frac{\tau_0 \gamma(y_0) R^2}{2n} \leq u(y_1) < u(y_0) < 1 \tag{5.34}$$

Hence, by the choice of R (see (5.32)) we have

$$\tau_0 < \frac{2n}{\gamma(y_0)R^2} < 1.$$

It follows that

$$w := \tau_0 z - u \leq 0 \text{ in } B,\ w(x_0) = \tau_0 z(x_0) - u(x_0) = 0. \tag{5.35}$$

Note that, according to (5.34), $u(x_0) < u(y_0)$ and so in an neighborhood N of x_0 we still have $u(x) < u(y_0)$, thus $\varepsilon_1 < u(x) < u(y_0)$. The definition of $\gamma(y_0)$ implies that

$$\Delta u(x) \leq -\gamma(y_0) \text{ for } x \in N,$$

and thus

$$\Delta w(x) \geq -\tau_0 \gamma(y_0) + \gamma(y_0) > 0 \text{ for } x \in N,$$

as $\tau_0 < 1$. This contradicts the fact that w has a local maximum at x_0.

Therefore, (5.31) holds, and $\gamma(y)$ satisfies $\gamma(y) \to 0$ as $\mathrm{dist}(y, \partial\Omega) \to \infty$. As a consequence, we conclude that $f(u(y)) \to 0$ in this limit, which implies $u(y) \to 1$, since $u(y) \geq \varepsilon_1$ in this region. Moreover, the above proof shows that the rate at which $u(x) \to 1$ as $\mathrm{dist}(x, \partial\Omega) \to \infty$ depends only on function $f(s)$, that is, for any $\varepsilon > 0$ there exists L_ε so that for any positive solution $u(x)$ of (5.20) we have

$$u(x) > 1 - \varepsilon, \text{ if } \mathrm{dist}(x, \partial\Omega) > L_\varepsilon.$$

Proof of (iv) in Theorem 5.5

We now show uniqueness of a positive bounded solution of (5.20). Naturally, we will do this by sliding. In order to start sliding, we will need the following estimate in strips.

Lemma 5.7 *For any $h > 0$ the solution is bounded away from 1 in the strip*

$$\Omega_h = \{x \in \Omega : \phi(x') < x_n < \phi(x') + h\}.$$

Proof. This follows immediately from the regularity of the solution $u(x)$ up to the boundary that follows from the standard elliptic estimates [69]. It is instructive also to see the argument by contradiction and shifting. Assume there exists a sequence $\xi_j \in \Omega_h$ so that $u(\xi_j) \to 1$. We shift all ξ_j to the origin: set $u_j(x) = u(x+\xi_j)$, with Ω shifted to a domain $\Omega_j = \{x_n > \phi_j(x')\}$. The shifted functions $\phi_j(x')$ are all translations of $\phi(x')$ (up to an additive constant) and thus all have the same Lipschitz constant. Thus, along a subsequence they converge to a function $\hat{\phi}(x)$. The shifted domains converge to a domain $\hat{\Omega} = \{x_n > \hat{\phi}(x')\}$, with 0 an interior point of $\hat{\Omega}$, while the shifted solutions converge along a subsequence (this also follows from the standard elliptic estimates) to a solution of

$$\begin{aligned} &\Delta\hat{u} + f(\hat{u}) = 0, \quad \text{in } \hat{\Omega} \\ &0 \le \hat{u} \le 1, \quad \text{in } \hat{\Omega}, \\ &\hat{u} = 0 \text{ on } \partial\Omega, \end{aligned}$$

such that $\hat{u}(0) = 1$. This is impossible according to part (ii) of the present theorem that we have already proved. □

Let now u and w be a pair of positive bounded solutions of (5.20). Note that the condition

$$d(x) := x_n - \phi(x') \to \infty$$

implies $\mathrm{dist}(x, \partial\Omega) \to \infty$, as the function ϕ is Lipschitz. Hence, part (iii) of Theorem 5.5 that we have already proved implies that both $u(x)$ and $w(x)$ tend to one uniformly as $d(x) \to \infty$. Hence there exists $A > 0$ so that

$$u(x), w(x) \ge s_1 \text{ if } d(x) \ge A, \tag{5.36}$$

with s_1 as in (5.23): $f(s)$ is non-increasing on $(s_1, 1)$. We set $\Omega^\varepsilon = \{x \in \Omega : \ d(x) > A\}$ and $\Omega_A = \{x \in \Omega : \ d(x) < A\}$. A key point is that once we show $u \ge w$ in Ω_A then this inequality propagates to the whole Ω. More generally, we have the following lemma.

Lemma 5.8 *Suppose that for some $\tau \ge 0$ the inequality*

$$u^\tau(x) = u(x + \tau\mathbf{e}_n) \ge w(x) \tag{5.37}$$

holds in $\bar{\Omega}_A$. Then (5.37) holds in all of Ω.

Proof. The proof is very similar to that of Corollary 5.4. Assume that (5.37) holds. The function $z = w - u^\tau$ satisfies an equation of the form

$$\Delta z + c(x)z = 0 \text{ in } \Omega^\varepsilon.$$

Both $w, u^\tau \geq s_1$ in Ω^ε, thus

$$c(x) = \frac{f(w(x)) - f(u^\tau(x))}{w(x) - u^\tau(x)} \leq 0,$$

wherever $z(x) \geq 0$. Moreover, by assumption $z \leq 0$ on $\partial\Omega^\varepsilon$ – this is all we need from (5.37). We may now apply Lemma 5.3, the maximum principle for unbounded domains, to Ω^ε and conclude that $z \leq 0$ in Ω^ε, that is (5.37) holds in all of Ω. □

Let us now show that $u(x) \geq w(x)$ in Ω_A. We do that by the sliding method. Note that

$$u^\tau(x) = u(x + \tau \mathbf{e}_n) > w(x) \text{ in } \Omega_A \text{ for large } \tau > 0,$$

because $u^\tau(x) \to 1$ as $\tau \to +\infty$ (according to the already proved part (iii) of the present theorem) while Lemma 5.7 implies that $w(x)$ is bounded away from 1 in Ω_A. As has been our common practice, we let

$$T = \inf\{\tau > 0 : \; u^\tau(x) \geq w(x) \text{ in } \Omega_A\}.$$

By continuity,

$$u^T(x) \geq w(x) \text{ in } \Omega_A. \tag{5.38}$$

We have to prove that $T = 0$. Suppose that $T > 0$, then there is a sequence of points $x_j \in \Omega_A$ and a sequence $\tau_j < T$, $\tau_j \to T$, so that

$$u(x_j + \tau_j \mathbf{e}_n) < w(x_j). \tag{5.39}$$

Once again, we shift the points x_j to the origin. The domain Ω is moved to Ω_j, and, as before, along a subsequence, Ω_j converge to a domain $\hat{\Omega} = \{x : \; x_n > \hat{\phi}(x')\}$ with a Lipschitz function $\hat{\phi}$. The shifts of u and w converge to positive solutions $\hat{u}$ and $\hat{w}$ of (5.20) in $\hat{\Omega}$ – this also follows from the standard elliptic regularity estimates. In addition, as $T > 0$, we know that $\hat{u}(x) \geq q_0 > 0$ in $\hat{\Omega}$, with some $q_0 > 0$. As follows from (5.38), we have

$$\hat{u}(x + T\mathbf{e}_n) \geq \hat{w}(x) \text{ in } \hat{\Omega}_A. \tag{5.40}$$

Lemma 5.8 implies that this inequality holds in the whole domain $\hat{\Omega}$. But passing to the limit in (5.39) we obtain that at the origin

$$0 < \hat{u}(T\mathbf{e}_n) \leq \hat{w}(0).$$

This implies that

$$\hat{u}^T(0) = \hat{w}(0), \tag{5.41}$$

and, in particular, 0 is an interior point of $\hat{\Omega}$, as on the boundary of $\hat{\Omega}$ we have $\hat{w} = 0$ while $\hat{u}^T > 0$ on $\partial\hat{\Omega}$ since $T > 0$. We have reached a contradiction: the function $\hat{z} = \hat{w} - \hat{u}^T$ satisfies an elliptic equation

$$\Delta \hat{z} + \hat{c}(x)\hat{z} = 0,$$

and inequality (5.40) means that $\hat{z} \leq 0$, while (5.41) implies that $\hat{z}(0) = 0$ and $0 \in \hat{\Omega}$. It follows that $\hat{z} \equiv 0$. However, at the boundary $\partial\hat{\Omega}$, $\hat{w} = 0$ while $\hat{u}^T > 0$, hence $z < 0$ on $\partial\hat{\Omega}$, which is a contradiction. Thus, $T = 0$ and $u(x) \geq w(x)$ for all $x \in \Omega_A$, hence in all of Ω. Similarly, we can show that $w(x) \geq u(x)$ for all $x \in \Omega$, and uniqueness follows.

Proof of (i) in Theorem 5.5

One would like to use Theorem 4.4 to show monotonicity. The problem is that the domain Ω is unbounded. This should be remedied by the fact that $u \to 1$ as $\text{dist}(x, \partial\Omega) \to \infty$ – hence, one may think of infinity as another part of the boundary where the value $u = 1$ is prescribed that guarantees that condition (4.11) still "holds" with the boundary condition "$\eta(x') = 0$ and $\eta(x'') = 1$". In order to make this precise we will consider a family of approximating domains

$$\Omega_h = \{\phi(x') < x_n < \phi(x') + h\} \tag{5.42}$$

and consider a sequence $h_n \to \infty$. We will construct a monotonic solution w_h on Ω_h and let $h \to \infty$. The sequence w_{h_n} will converge to a limit function w along a subsequence $h_n \to \infty$. The function w will be a monotonic solution of (5.20) and uniqueness (the already proved part (iv) of the Theorem 5.5) will finish the proof. Moreover, in order to make sure that $w \neq 0$ identically, we will construct w_h so that $w_h \geq u$ in Ω_h.

Let us construct the monotonic solution w. This is done in two steps. First, we consider the cylinder

$$\Omega_{h,R} = \{x \in \Omega_h : \ |x'| < R\}$$

with Ω_h as in (5.42). The standard Hölder regularity estimates up to the boundary (recall that the boundary of Ω is smooth) imply that there exist $M > 0$ and $\alpha > 0$ so that

$$|u(x) - u(y)| \leq M|x - y|^\alpha \text{ for } x, y \in \Omega. \tag{5.43}$$

Using the constants α and M as above, we define

$$\sigma(t) = \begin{cases} Mt^\alpha, \text{ for } 0 \leq t \leq M^{-1/\alpha}, \\ 1, \text{ for } t \geq M^{-1/\alpha}. \end{cases}$$

We consider $h > h_0 = 1 + M^{-1/\alpha}$ and define a continuous function σ_R on $\partial\Omega_{h,R}$:

$$\sigma_R(t) = \begin{cases} 0, & \text{for } x \in \partial\Omega, \\ 1, & \text{for } x \text{ s.t. } x_n = \phi(x') + h, \\ \sigma(x_n - \phi(x')), & \text{otherwise on } \partial\Omega_{h,R}. \end{cases}$$

Note that $\sigma_R \geq u$ on $\partial\Omega_h$ by (5.43). Let w the solution of

$$\begin{aligned} &\Delta w_{h,R} + f(w_{h,R}) = 0 \text{ in } \Omega_{h,R} \\ &w = \sigma_R \text{ on } \partial\Omega_{h,R}. \end{aligned} \tag{5.44}$$

Existence of a solution to (5.44) follows from the fact that it has a sub-solution $\underline{w} = u$ and a super-solution $\overline{w} = 1$. Indeed, start with $w_0 = \underline{w}$ and solve

$$\begin{aligned} &\Delta w_{j+1} - kw_{j+1} = -f(w_j) - kw_j \text{ in } \Omega_{h,R} \\ &w_{j+1} = \sigma_R \text{ on } \partial\Omega_{h,R}. \end{aligned}$$

Here k is the Lipschitz constant of f. First, we have

$$\Delta w_1 - kw_1 = -f(w_0) - kw_0 = -f(u) - ku = \Delta u - ku$$

and hence

$$\Delta(w_1 - u) - k(w_1 - u) = 0 \text{ in } \Omega_{h,R},$$

while $w_1 \geq u$ on $\partial\Omega_{h,R}$. Hence $w_1 \geq u \geq w_0$. The induction argument shows that

$$w_0 \leq w_1 \leq \cdots \leq \overline{w}, \tag{5.45}$$

because

$$\Delta(w_{j+1} - w_j) - k(w_{j+1} - w_j) = -k(w_j - w_{j-1}) - [f(w_j) - f(w_{j-1})] \leq 0$$

by the induction hypothesis. The last inequality in (5.45) also follows from induction applied to

$$\Delta(w_{j+1} - \bar{w}) - k(w_{j+1} - \bar{w}) = -k(w_j - \bar{w}) - (f(w_j) - f(\bar{w})) \geq 0.$$

Hence, w_j converge to a limit $w_{h,R}$ as $j \to +\infty$ – elliptic regularity implies that $w_{h,R}$ is a solution of (5.44).

Theorem 4.4 implies that $w_{h,R}$ is monotonic in x_n, and the maximum principle implies that $w_{h,R} \geq u$.

We now pass to the limit $R \to \infty$. The standard elliptic estimates as before imply that $w_{h,R}$ converges along a subsequence $R_n \to \infty$ to a function $w_h \geq u$ that satisfies

$$\begin{aligned}
&\Delta w_h + f(w_h) = 0 \text{ in } \Omega_h \\
&w_h = 0 \text{ on } \partial\Omega \\
&w_h = 1 \text{ on } \{x_n = \phi(x') + h\}.
\end{aligned}$$

Finally we let $h \to \infty$, and by the same argument conclude that, along a subsequence, w_{h_n} converges to a monotonic solution of

$$\begin{aligned}
&\Delta w + f(w) = 0 \text{ in } \Omega_h \\
&w = 0 \text{ on } \partial\Omega
\end{aligned}$$

with $w \geq u$. Hence uniqueness of a positive bounded solution implies that u has to coincide with w and we are done. □

Proof of (v) of Theorem 5.5. This one is a trivial consequence of part (i): all it says is that u is monotonic in any direction ξ such that there exists an orthonormal basis with $\mathbf{e}_n$ along ξ so that the boundary $\partial\Omega$ may be represented as $z = \phi(z')$ in the new variables. □

Chapter 2

The parabolic maximum principle and the principal eigenvalue

1 The parabolic maximum principle

The parabolic maximum principle in a bounded domain

The next step in the background review is to recall some basics on the maximum principle for parabolic equations, which are very similar in spirit to what we have just described for the elliptic equations. Now, we consider a more general elliptic operator of the form

$$Lu(x) = a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + b_j(x)\frac{\partial u}{\partial x_j}, \tag{1.1}$$

in a bounded domain $\Omega \subset \mathbb{R}^n$. Note that the zero-order coefficient $c(x)$ is equal to zero. We assume that the matrix $a_{ij}(x)$ is uniformly elliptic and bounded: there exist two positive constants $\lambda > 0$ and $\Lambda > 0$ so that, for any $\xi \in \mathbb{R}^n$ and any $x \in \Omega$, we have

$$\lambda|\xi|^2 \leq a_{ij}(x)\xi_i\xi_j \leq \Lambda|\xi|^2. \tag{1.2}$$

We also assume that all coefficients $a_{ij}(x)$ and $b_j(x)$ are continuous and uniformly bounded in Ω. Given a time $T > 0$, define the parabolic cylinder $\Omega_T = [0, T) \times \Omega$ and the parabolic boundary

$$\Gamma_T = \{x \in \Omega,\ 0 \leq t \leq T :\ \text{either } x \in \partial\Omega \text{ or } t = 0\},$$

that is, Γ_T is the part of the boundary of Ω_T without "the top" $\{(t, x) :\ t = T, x \in \Omega\}$.

Theorem 1.1 *(The weak maximum principle) Let a function $u(t, x)$ satisfy*

$$\frac{\partial u}{\partial t} = Lu, \quad x \in \Omega, \quad 0 \leq t \leq T, \tag{1.3}$$

and assume that Ω is a smooth bounded domain. Then $u(t, x)$ attains its maximum over Ω_T on the parabolic boundary Γ_T, that is,

$$\sup_{\Omega_T} u(t, x) = \sup_{\Gamma_T} u(t, x). \tag{1.4}$$

Proof. Take $\varepsilon > 0$ and consider the (once again, "corrected" to produce a strict sub-solution) function

$$v(t,x) = u(t,x) - \varepsilon t,$$

which satisfies

$$\frac{\partial v}{\partial t} - Lv = -\varepsilon. \tag{1.5}$$

In other words, v is, indeed, a strict sub-solution of the parabolic problem. The function $v(t,x)$ must attain its maximum over the set (the parabolic cylinder with the "top" included)

$$\bar{\Omega}_T = \Omega_T \bigcup \{(T,x) : \ x \in \Omega\},$$

at some point $(t_0, x_0) \in \bar{\Omega}_T$. We claim that this point has to lie on the parabolic boundary Γ_T. Indeed, if $0 < t_0 < T$ and x_0 is not on the boundary $\partial\Omega$, then the point (t_0, x_0) is an interior maximum of $v(t,x)$, so $v(t,x)$ should satisfy

$$\frac{\partial v(t_0,x_0)}{\partial t} = 0, \quad \nabla v(t_0,x_0) = 0,$$

and the matrix $D^2 v(t_0,x_0)$ should be non-positive definite. This implies

$$a_{ij}(x_0)\frac{\partial^2 v(t_0,x_0)}{\partial x_i \partial x_j} \leq 0.$$

The last two conditions imply that

$$\frac{\partial v(x_0,t_0)}{\partial t} - Lv(t_0,x_0) \geq 0,$$

which is impossible because of (1.5). On the other hand, if $t_0 = T$, x_0 is an interior point of Ω, and v attains its maximum over $\bar{\Omega}$ at this point, then we should have

$$\frac{\partial v(t_0,x_0)}{\partial t} \geq 0, \quad \nabla v(t_0,x_0) = 0, \quad a_{ij}(x_0)\frac{\partial^2 v(t_0,x_0)}{\partial x_i \partial x_j} \leq 0,$$

which, once again, contradicts (1.5). Hence, the function v attains its maximum over $\bar{\Omega}_T$ at a point (t_0, x_0) that belongs to the parabolic boundary Γ_T. It means that

$$\max_{(t,x)\in\bar{\Omega}_T} v(t,x) = \max_{(t,x)\in\Gamma_T} v(t,x) \leq \max_{(t,x)\in\Gamma_T} u(t,x).$$

However, we also have

$$\max_{(t,x)\in\bar{\Omega}_T} u(t,x) \leq \varepsilon T + \max_{(t,x)\in\bar{\Omega}_T} v(t,x).$$

Putting the last two inequalities together gives

$$\max_{(t,x)\in\bar{\Omega}_T} u(t,x) \leq \varepsilon T + \max_{(t,x)\in\Gamma_T} u(t,x).$$

As $\varepsilon > 0$ is arbitrary, it follows that

$$\max_{(t,x)\in\bar{\Omega}_T} u(t,x) \leq \max_{(t,x)\in\Gamma_T} u(t,x),$$

and the proof is complete. □

As in the elliptic case, we also have the strong maximum principle.

Theorem 1.2 *(The strong maximum principle) Let a smooth function* $u(t,x)$ *satisfy*

$$\frac{\partial u}{\partial t} = Lu, \quad x \in \Omega, \quad 0 \le t \le T, \tag{1.6}$$

in a smooth bounded domain Ω. *Then if* $u(t,x)$ *attains its maximum over* $\bar{\Omega}_T$ *at an interior point* $(t_0, x_0) \notin \Gamma_T$ *then* $u(t,x)$ *is constant in* Ω_T.

We will not prove it here, the reader may either use it as an exercise, or consult [52] for a proof.

The comparison principle

A consequence of the maximum principle is the comparison principle, a result that holds also for operators with zero order coefficients and in unbounded domains (under a proper restriction on the growth at infinity).

Theorem 1.3 *Let the smooth uniformly bounded functions* $u(t,x)$ *and* $v(t,x)$ *satisfy*

$$\frac{\partial u}{\partial t} = Lu + c(x)u, \quad 0 \le t \le T, \quad x \in \Omega \tag{1.7}$$

and

$$\frac{\partial v}{\partial t} = Lv + c(x)v, \quad 0 \le t \le T, \quad x \in \Omega, \tag{1.8}$$

in a smooth (and possibly unbounded) domain Ω. *Assume that* $u(0,x) \ge v(0,x)$ *and*

$$u(t,x) \ge v(t,x) \text{ for all } 0 \le t \le T \text{ and } x \in \partial\Omega.$$

Then, we have

$$u(t,x) \ge v(t,x) \text{ for all } 0 \le t \le T \text{ and all } x \in \Omega.$$

Moreover, if in addition, $u(0,x) > v(0,x)$ *on an open subset of* Ω *then* $u(t,x) > v(t,x)$ *for all* $0 < t < T$ *and all* $x \in \Omega$.

The assumption that both $u(t,x)$ and $v(t,x)$ are uniformly bounded is important – without this condition even the Cauchy problem for the standard heat equation in $\mathbb{R}^n$ may have more than one solution, and the comparison principle implies uniqueness trivially. Note that the special case $\Omega = \mathbb{R}^n$ is included in Theorem 1.3, and in that case only the comparison at the initial time $t = 0$ is needed for the conclusion to hold. Once again, a reader who is not interested in treating the proof as an exercise should consult [52].

A standard corollary of the parabolic maximum principle is the following estimate.

Exercise 1.4 Let Ω be a bounded domain, and $u(t,x)$ be the solution of the initial boundary value problem

$$\begin{aligned} &u_t = Lu + c(x)u, \text{ in } \Omega, \\ &u(t,x) = 0 \text{ for } x \in \partial\Omega, \\ &u(0,x) = u_0(x). \end{aligned} \tag{1.9}$$

Assume that the function $c(x)$ is bounded, with $c(x) \le M$ for all $x \in \Omega$, then $u(t,x)$ satisfies

$$|u(t,x)| \le \|u_0\|_{L^\infty} e^{Mt}, \quad \text{for all } t > 0 \text{ and } x \in \Omega. \tag{1.10}$$

The estimate (1.10) on the possible growth (or decay) of the solution of (1.9) is by no means optimal, and we will soon see how it can be improved.

2 The principal eigenvalue and the maximum principle

The principal eigenvalue

The maximum principle for elliptic and parabolic problems has a beautiful connection to the eigenvalue problems, which also allows to extend it to operators with a zero-order term. Here, we will consider operators of the form

$$Lu(x) = a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + b_j(x)\frac{\partial u}{\partial x_j} + c(x)u, \tag{2.1}$$

with a uniformly elliptic matrix $a_{ij}(x)$ and bounded and continuous coefficients $a_{ij}(x)$, $b_j(x)$ and $c(x)$.

We assume that Ω is a bounded smooth domain and consider the corresponding eigenvalue problem with the Dirichlet boundary conditions:

$$\begin{aligned} -Lu &= \lambda u, \quad x \in \Omega \\ u &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{2.2}$$

Let us recall that the eigenvalues λ_k, $k \in \mathbb{N}$ are discrete and have finite multiplicity[1]. As the operator L is not necessarily self-adjoint, the eigenvalues need not be real.

The key spectral property of the operator L comes from the comparison principle. To this end, we need to recall the Krein-Rutman theorem[2] which says if M is a compact operator which preserves a solid cone K of functions in the space $C(\Omega)$, and maps the boundary of K into its interior, then it has an eigenfunction ϕ that lies in this cone:

$$M\phi = \lambda\phi.$$

Moreover, the corresponding eigenvalue λ has the largest real part of all eigenvalues of the operator M. How can we apply this theorem to the elliptic operators? The operator L given by (2.1) is not compact, nor does it preserve any interesting cone. However, let us assume momentarily that $c(x) \leq 0$ for all $x \in \Omega$. Then the boundary value problem

$$\begin{aligned} -Lu &= f, \quad \text{in } \Omega \\ u &= 0 \ \text{ on } \partial\Omega, \end{aligned} \tag{2.3}$$

has a unique solution, and, in addition, if $f(x) \geq 0$ and $f \not\equiv 0$, then $u(x) > 0$ for all $x \in \Omega$. This means that we may define the inverse operator $M = (-L)^{-1}$. This operator preserves the cone of the positive functions, and maps its boundary (non-negative functions that vanish somewhere in Ω) into its interior. Hence, the inverse operator satisfies the assumptions of the Krein-Rutman theorem[3]. Thus, there exists a positive function f and $\mu \in \mathbb{R}$ so that

[1] This comes from the elliptic regularity estimates but we will not dwell on this issue here, the reader should really consult [52] or another reference for the proof.

[2] The classical reference [43] has an excellent discussion and proof of this theorem, as well as many of its beautiful its consequences.

[3] Once again, compactness of the inverse would remain under the rug for the moment, as it follows from the elliptic regularity estimates – please, see [52]

the function $u = \mu f$ satisfies (2.3). Positivity of f implies that the solution of (2.3) is also positive, hence $\mu > 0$. As μ is the eigenvalue of $(-L)^{-1}$ with the largest real part, $\lambda = \mu^{-1}$ is the eigenvalue of $(-L)$ with the smallest real part (in particular, all λ_k have a positive real part).

If the assumption $c(x) \le 0$ does not hold, we may take

$$M > \max_{x \in \Omega} c(x),$$

and consider the operator

$$L'u = Lu - Mu.$$

The zero-order coefficient of L' is $c'(x) = c(x) - M \le 0$. Hence, we may apply the previous argument to the operator L' and conclude that $(-L')$ has an eigenvalue μ_1 that corresponds to a positive eigenfunction, and has the smallest real part among all eigenvalues of $(-L')$. The same is true for the operator $(-L)$, with the eigenvalue $\lambda_1 = \mu_1 - M$. We say that λ_1 is the principal (Dirichlet) eigenvalue of the operator $(-L)$. The same conclusion holds if the domain Ω is a torus, and the Dirichlet boundary condition in (2.2) is replaced by the requirement that the eigenfunction is periodic.

The comparison principle revisited

Let us now connect the principal eigenvalue and the comparison principle. The principal eigenfunction $\phi_1 > 0$, solution of

$$-L\phi_1 = \lambda_1 \phi_1, \text{ in } \Omega, \tag{2.4}$$

$$\phi_1 = 0 \text{ on } \partial\Omega, \tag{2.5}$$

in particular, provides a special solution

$$\psi(t,x) = e^{-\lambda_1 t}\phi_1(x) \tag{2.6}$$

for the linear parabolic problem

$$\begin{aligned} &\frac{\partial \psi}{\partial t} = L\psi, \quad t > 0, x \in \Omega \\ &\psi = 0 \text{ on } \partial\Omega. \end{aligned} \tag{2.7}$$

Consider then the Cauchy problem

$$\begin{aligned} &\frac{\partial v}{\partial t} = Lv, \quad t > 0, x \in \Omega \\ &v = 0 \text{ on } \partial\Omega, \\ &v(0,x) = g(x), \quad x \in \Omega, \end{aligned} \tag{2.8}$$

with a smooth bounded function $g(x)$ that vanishes at the boundary $\partial\Omega$. We can find a constant $M > 0$ so that

$$-M\phi_1(x) \le g(x) \le M\phi_1(x), \quad \text{for all } x \in \Omega.$$

The comparison principle then implies that for all $t > 0$ we have a bound

$$-M\phi_1(x)e^{-\lambda_1 t} \le v(t,x) \le M\phi_1(x)e^{-\lambda_1 t}, \quad \text{for all } x \in \Omega, \tag{2.9}$$

which is very useful, especially if $\lambda_1 > 0$. The assumption that the initial data g vanishes at the boundary $\partial\Omega$ is not necessary but removes the technical step of having to show that even if $g(x)$ does not vanish on the boundary, for any positive time $t_0 > 0$ we can find a constant C_0 so that $|v(t_0,x)| \le C_0\phi_1(x)$. This leads to the bound (2.9) for all $t > t_0$.

Let us now apply the above considerations to solutions of the elliptic problem

$$\begin{aligned} -a_{ij}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} - b_j(x)\frac{\partial u}{\partial x_j} - c(x)u = g(x), \\ u = 0 \text{ on } \partial\Omega, \end{aligned} \tag{2.10}$$

with a non-negative function $g(x)$. When can we conclude that the solution $u(x)$ is also non-negative? Solution of (2.10) can be formally written as

$$u(x) = \int_0^\infty v(t,x)dt. \tag{2.11}$$

Here, the function $v(t,x)$ satisfies the Cauchy problem (2.8). If the principal eigenvalue λ_1 of the operator $\mathcal{L}$ is positive, then the integral (2.11) converges for all $x \in \Omega$ because of the estimates (2.9), and solution of (2.10) is, indeed, given by (2.11). On the other hand, if $g(x) \ge 0$ and $g(x) \not\equiv 0$, then the parabolic comparison principle implies that $v(t,x) > 0$ for all $t > 0$ and all $x \in \Omega$.

Therefore, we have proved the following theorem that succinctly relates the notions of the principal eigenvalue and the comparison principle.

Theorem 2.1 *If the principal eigenvalue of the operator $(-L)$ is positive then solutions of the elliptic equation (2.10) satisfy the comparison principle: $u(x) > 0$ in Ω if $g(x) \ge 0$ in Ω and $g(x) \not\equiv 0$.*

This theorem allows to look at the maximum principle in narrow domains from a slightly different point of view: the narrowness of the domain implies that the principal eigenvalue of $(-L)$ is positive no matter what the sign of the free coefficient $c(x)$ is. This is because the “size” of the second order term in L increases as the domain narrows, while the “size” of the zero-order term does not change. Therefore, in a sufficiently narrow domain the principal eigenvalue of $(-L)$ will be positive (recall that the required narrowness does depend on the size of $c(x)$).

We conclude this section with another characterization of the principal eigenvalue of an elliptic operator in a bounded domain, which we leave as an (important) exercise for the reader. Let us define

$$\mu_1(\Omega) = \sup\{\lambda : \ \exists \phi \in C^2(\Omega) \cap C^1(\bar{\Omega}), \ \phi > 0 \text{ and } (L+\lambda)\phi \le 0 \text{ in } \Omega\}, \tag{2.12}$$

and

$$\mu_1'(\Omega) = \inf\{\lambda : \ \exists \phi \in C^2(\Omega) \cap C^1(\bar{\Omega}), \ \phi = 0 \text{ on } \partial\Omega, \ \phi > 0 \text{ in } \Omega, \text{ and } -(L+\lambda)\phi \le 0 \text{ in } \Omega\}. \tag{2.13}$$

Exercise 2.2 Let L be an elliptic operator in a smooth bounded domain Ω, and let λ_1 be the principal eigenvalue of the operator $(-L)$, and $\mu_1(\Omega)$ and $\mu_1'(\Omega)$ be as above. Show that

$$\lambda_1 = \mu_1(\Omega) = \mu_1'(\Omega). \tag{2.14}$$

As a hint, say, for the equality $\lambda_1 = \mu_1(\Omega)$, we suggest, assuming existence of some $\lambda > \lambda_1$ and $\phi > 0$ such that

$$(L + \lambda)\phi \le 0,$$

to consider the Cauchy problem

$$u_t = (L + \lambda)u, \quad \text{in } \Omega$$

with the initial data $u(0, x) = \phi(x)$, and with the Dirichlet boundary condition $u(t, x) = 0$ for $t > 0$ and $x \in \partial\Omega$. One should prove two things: first, that $u_t(t, x) \le 0$ for all $t > 0$, and, second, that there exists some constant $C > 0$ so that

$$u(t, x) \ge C\bar{\phi}(x)e^{(\lambda-\lambda_1)t},$$

where $\bar{\phi}$ is the principal Dirichlet eigenfunction of $(-L)$. This will lead to a contradiction. The second equality in (2.14) is proved in a similar way.

3 The periodic principal eigenvalue in unbounded domains

We will now give a superficial but, hopefully, tempting discussion of the principle eigenvalue of an elliptic operator in an unbounded domain (we will consider, for simplicity, only $\Omega = \mathbb{R}^n$). The first issue is simply to understand what one could mean by the principal eigenvalue in an unbounded domain. A "simpleton" way is to look for a true positive eigenfunction and the corresponding eigenvalue, solution of

$$-L\phi = \lambda\phi, \quad x \in \mathbb{R}^n$$

with $\phi > 0$. Alas, such positive eigenfunction may exist for infinitely many λ – just consider the problem

$$-u'' = \lambda u \ \text{ in } \mathbb{R},$$

which has two positive eigenfunctions (exponentials) for all $\lambda > 0$.

A natural "brute force" idea to overcome this issue would be to look at the Dirichlet eigenvalue problems in a sequence of bounded domains Ω_k that would fill $\mathbb{R}^n$ as $k \to +\infty$, and pass to the limit. Let us illustrate what can happen with this approach – there are good and bad news. We start with the good news – they come from the self-adjoint operators with the periodic coefficients, of the form

$$Lu = -\Delta u - \mu(x)u, \tag{3.1}$$

with a 1-periodic (in all x_j) function $\mu(x)$. A natural candidate for the principal eigenvalue of the operator L in the whole space is the principal eigenvalue of the periodic problem

$$\begin{aligned} &-\Delta\phi - \mu(x)\phi = \lambda_1\phi, \\ &\phi(x) \text{ is 1-periodic in all its variables.} \end{aligned} \tag{3.2}$$

Indeed, he Krein-Rutman theorem implies that this problem has a positive eigenfunction that corresponds to a simple eigenvalue. The corresponding function ϕ is also a positive bounded eigenfunction of the operator L in the whole space in the most literal sense. Let us see whether it can be obtained as a limit with the "exhaustion by bounded domains" procedure we have mentioned above – consider the Dirichlet eigenvalue problem in the ball $B(0;R)$

$$\begin{aligned} &-\Delta\psi_R(x) - \mu(x)\psi_R = \lambda_R\psi_R(x), \quad |x| < R, \\ &\psi_R(x) > 0 \text{ for } |x| < R, \\ &\psi_R(x) = 0 \text{ on } \{|x| = R\}, \end{aligned} \tag{3.3}$$

and investigate the limit $R \to +\infty$. It turns out that the two approaches are equivalent for an operator of the form (3.1).

Theorem 3.1 *Let λ_1 be the principal periodic eigenvalue of the problem (3.2), and λ_R be the principal Dirichlet eigenvalue of the problem (3.3), then*

$$\lim_{R\to+\infty} \lambda_R = \lambda_1. \tag{3.4}$$

Proof. Let us first recall the variational principles for λ_1 and λ_R in terms of the Rayleigh quotient (a reader not yet familiar with these formulations should consult Section 6.5 of [52]):

$$\lambda_1 = \inf_{v\in H^1(\mathbb{T}^n)} \frac{\displaystyle\int_{\mathbb{T}^n} (|\nabla v|^2 - \mu(x)v^2)dx}{\displaystyle\int_{\mathbb{T}^n} |v|^2 dx} \tag{3.5}$$

and

$$\lambda_R = \inf_{v\in H_0^1(B(0,R))} \frac{\displaystyle\int_{\mathbb{T}^n} (|\nabla v|^2 - \mu(x)v^2)dx}{\displaystyle\int_{\mathbb{T}^n} |v|^2 dx}. \tag{3.6}$$

The difference between the two expressions is in the collection of test functions: 1-periodic H^1 functions in the case of λ_1 and $H_0^1(B(0;R))$ functions in the case of λ_R. Uniqueness of the positive eigenfunction shows that, for any positive integer m, λ_1 is also the principal periodic eigenvalue on the larger torus $T_m = [0,m]^n$. Hence, λ_1 can be written as

$$\lambda_1 = \inf_{v\in H^1(T_m)} \frac{\displaystyle\int_{T_m} (|\nabla v|^2 - \mu(x)v^2)dx}{\displaystyle\int_{T_m} |v|^2 dx}. \tag{3.7}$$

That is, the infimum can be also taken over all m-periodic functions. Let us take $m > 27R$, set the vector $e = (1, 1, \ldots, 1)$, and consider an m-periodic function $v_{R,m}$ (defined in the period cell $[0, m]^n$) that equals $\psi_R(x - (m/2)e)$ in the ball $B(me/2, R)$, and to zero everywhere else in $T_m = [0, m]^n$. Note that $B(me/2, R) \subset T_m$. The Rayleigh quotient of $v_{R,m}$ is exactly λ_R, hence

$$\lambda_1 \leq \lambda_R. \tag{3.8}$$

In order to establish the opposite bound, let ϕ_1 be the 1-periodic eigenfunction and set

$$w_R(x) = \chi_R(x)\phi_1(x),$$

where $\chi_R(x)$ is a smooth cut-off function such that $0 \leq \chi_R(x) \leq 1$, $\chi_R(x) = 1$ for $|x| \leq R - 1$, and $\chi_R(x) = 0$ for $|x| \geq R$. We may assume that $\|\chi_R\|_{C^2} \leq K$ with a constant K that does not depend on R. The L^2-norm of the gradient of w_R is

$$\begin{aligned}\int\limits_{B(0,R)} |\nabla w_R(x)|^2 dx &= \int_{B(0,R)} |\nabla \chi_R(x)\phi_1(x) + \chi_R(x)\nabla\phi_1(x)|^2 dx \\ &= \int\limits_{B(0,R)} (|\nabla\chi_R|^2|\phi_1(x)|^2 + 2(\phi_1(x)\chi_R(x)\nabla\chi_R(x)\cdot\nabla\phi_1(x))dx + \int\limits_{B(0,R)} |\chi_R(x)|^2|\nabla\phi_1(x)|^2 dx.\end{aligned}$$

As $\nabla\chi_R(x) = 0$ and $\chi_R(x) = 1$ for x outside the annulus $R - 1 \leq |x| \leq R$, it is easy to see from the above that

$$\int_{B(0,R)} |\nabla w_R(x)|^2 dx = \int_{B(0,R)} |\nabla\phi_1(x)|^2 dx + O(R^{n-1}).$$

Furthermore, we can estimate, using the same idea:

$$\int_{B(0,R)} \mu(x)|w_R(x)|^2 dx = \int_{B(0,R)} \mu(x)|\phi_1(x)|^2 dx + O(R^{n-1}).$$

The notation above means that the integrals in the left and right side differ by expressions that can be bounded by CR^{n-1}. And, finally, we have, in the same way:

$$\int_{B(0,R)} |w_R(x)|^2 dx = \int_{B(0,R)} |\phi_1(x)|^2 dx + O(R^{n-1}).$$

The last observation is that, for instance,

$$\int_{B(0,R)} |\phi_1(x)|^2 dx = N_R \int_{[0,1]^n} |\phi_1(x)|^2 dx + O(R^{n-1}),$$

and similarly for the other integrals appearing in the Rayleigh quotient for w_R. Here N_R is the number of disjoint $[0, 1]^n$ cubes that fit into the ball $B(0, R)$. We deduce that

$$\begin{aligned}\lambda_R &\leq \frac{\displaystyle\int_{B_R} |(|\nabla w_R(x)|^2 - \mu(x)|w_R(x)|^2)dx}{\displaystyle\int_{B_R} |w_R(x)|^2 dx} = \frac{\displaystyle\int_{[0,1]^n} |(|\nabla\phi_1(x)|^2 - \mu(x)|\phi_1(x)|^2)dx}{\displaystyle\int_{[0,1]^n} |\phi_1(x)|^2 dx} + O(R^{-1}) \\ &= \lambda_1 + O(R^{-1}).\end{aligned} \tag{3.9}$$

This estimate, together with (3.8) shows that

$$\lim_{R\to+\infty}\lambda_R=\lambda_1, \tag{3.10}$$

and the proof of Theorem 2.4 is complete. □

Theorem 2.4 can be generalized to other self-adjoint operators with periodic coefficients, as the same argument using the Rayleigh quotient would still apply. To ensure the reader that the matter is less trivial in general, consider the following example from [23]. The operator

$$Lu=-u''+u'$$

is periodic on the real line, with an arbitrary period. The principal periodic eigenvalue is $\lambda_1=0$ and the principal eigenfunction is $\phi_1(x)\equiv 1$. On the other hand, solution of the Dirichlet problem on the interval $[-R,R]$

$$-\phi_R''+\phi_R'=\lambda_R\phi_R,\quad -R<x<R,\ \ \phi_R(-R)=\phi_R(R)=0,$$

is explicit:

$$\phi_R(x)=e^{x/2}\cos(\frac{\pi x}{2R}),$$

and the corresponding principal eigenvalue is

$$\lambda_R=\frac{1}{4}+\frac{\pi^2}{4R^2}.$$

Thus, we have

$$\lim_{R\to+\infty}\lambda_R=\frac{1}{4}\neq\lambda_1=0, \tag{3.11}$$

and the result of Theorem 2.4 fails spectacularly in the non-self-adjoint case.

It is easy to understand the above phenomenon from the probabilistic interpretation: continue here

Let us now revert to the philosophy of Exercise 2.2 and see what this would give us, and whether it can reconcile the discrepancy in (3.11). For simplicity, we will consider periodic operators in $\mathbb{R}^n$ – an interested reader should consult [23] for a much fuller picture. Given a periodic elliptic operator

$$Lu=a_{ij}(x)\frac{\partial^2 u}{\partial x_i\partial x_j}+b_i(x)\frac{\partial u}{\partial x_i}+c(x)u,$$

we let λ_1 be the principal periodic eigenvalue of $(-L)$, and also define

$$\mu_1=\sup\{\lambda:\ \exists\phi\in C^2(\mathbb{R}^n),\ \phi>0 \text{ and } (L+\lambda)\phi\le 0 \text{ in } \mathbb{R}^n\}, \tag{3.12}$$

and

$$\mu_1'=\inf\{\lambda:\ \exists\phi\in C_b^2(\mathbb{R}^n),\ \phi>0 \text{ and } -(L+\lambda)\phi\le 0 \text{ in } \mathbb{R}^n\}. \tag{3.13}$$

Note the subtle difference in the definition of μ_1 and μ_1' – in the first case, the test functions are just smooth, in the second, they are also bounded, a distinction unnecessary in a bounded domain. Recall that μ_1 and μ_1' coincide in a bounded domain – both are equal to the principal Dirichlet eigenvalue in Ω. What happens in the whole space? It turns out that μ_1 and μ_1' may be different.

Theorem 3.2 *(i) We have $\mu_1' = \lambda_1$.*
(ii) Let λ_R be the principal Dirichlet eigenvalue of the operator $(-L)$ in the ball $B(0;R)$, then

$$\lim_{R\to+\infty} \lambda_R = \mu_1.$$

Proof. We first prove (i). Taking $\phi = \phi_1$, the periodic eigenfunction of $(-L)$, and $\lambda = \lambda_1$ in the definition of μ_1' we get $\mu_1' \le \lambda_1$. Next, take any $\lambda < \lambda_1$. As $\phi_1 > 0$ is periodic, we can find $\varepsilon > 0$ so that

$$-(L+\lambda)\phi_1 = (\lambda_1 - \lambda)\phi_1 \ge \varepsilon > 0.$$

Assume now that there exists some bounded function $\phi \in C_b^2(\mathbb{R}^n)$ so that

$$-(L+\lambda)\phi \le 0.$$

We will show that ϕ has to be non-positive. Assume that $\sup \phi(x) > 0$, set

$$m = \sup_{x\in\mathbb{R}^n} \frac{\phi(x)}{\phi_1(x)},$$

and define $w(x) = m\phi_1(x) - \phi(x) \ge 0$. We will show that

$$\inf_{x\in\mathbb{R}^n} w(x) > 0,$$

which will give a contradiction to the definition of m. The function $w(x)$ satisfies

$$-(L+\lambda)w = m(\lambda_1 - \lambda)\phi_1 + (L+\lambda)\phi \ge m\varepsilon > 0. \tag{3.14}$$

If $w(x)$ attains its infimum (which is equal to zero by the definition of m), we would get an immediate contradiction to the maximum principle from (3.14). If there is a sequence x_k such that $w(x_k) \to 0$ as $k \to +\infty$, we need to do a little more work. Take a smooth function $\theta(x) \ge 0$ such that $\theta(0) = 0$, and

$$\lim_{|x|\to+\infty} \theta(x) = 1.$$

We can find a constant $K > 0$ so that any translation $\theta_y(x) = \theta(x-y)$ (with any $y \in \mathbb{R}^n$) satisfies

$$-(L+\lambda)\theta_y(x) \ge -\frac{m\varepsilon K}{2}, \quad \text{for all } x \in \mathbb{R}^n.$$

Let us then choose a point x_0 so that $w(x_0) \le \varepsilon_1 < m/K$, where ε_1 is a sufficiently small constant (which will depend on K, but also on ε and the coefficients of the operator L). We may then choose $R > 0$ so large that

$$\frac{\theta(x-x_0)}{K} > \varepsilon_1 > w(x_0), \quad \text{for all } x \in \partial B(x_0, R).$$

The function

$$\widetilde{w}(x) = w(x) + \frac{\theta(x-x_0)}{K},$$

satisfies then $\widetilde{w}(x) > w(x_0) = \widetilde{w}(x_0)$ on the boundary $\partial B(x_0, R)$, hence $\widetilde{w}$ has to attain its infimum over $B(x_0; R)$ at some interior point $y_0 \in B(x_0, R)$. Finally, consider the function

$$\chi(x) = \widetilde{w}(x) - \widetilde{w}(y_0).$$

The function χ is non-negative in $B_R(x_0, R)$ and vanishes at y_0. But we also have

$$\begin{aligned}
-(L+\lambda)\chi &= -(L+\lambda)\widetilde{w} + (c(x)+\lambda)\widetilde{w}(y_0) \\
&= -(L+\lambda)w - \frac{1}{K}(L+\lambda)\theta_{x_0} + (c(x)+\lambda)\widetilde{w}(y_0) \\
&\geq m\varepsilon - \frac{m\varepsilon}{2} - (\|c\|_\infty + |\lambda|)\varepsilon_1 > \frac{\varepsilon}{4},
\end{aligned}$$

if we choose ε_1 sufficiently small. This, however, contradicts the fact that χ attains an interior minimum at the point y_0 where $\chi(y_0) = 0$. This contradiction shows that $\lambda_1 \leq \mu_1'$, and thus $\lambda_1 = \mu_1'$.

We now prove part (ii) of the theorem:

$$\lim_{R\to+\infty} \lambda_R = \mu_1.$$

Notice that, by definition, we have

$$\mu_1(B_{R_1}) \leq \mu_1(B_{R_2}), \quad \text{for all } R_1 > R_2,$$

and

$$\mu_1 \leq \mu_1(B_R), \quad \text{for all } R > 0.$$

As $\mu_1(B_R) = \lambda_R$, it follows that

$$\mu_1 \leq \bar{\lambda} := \lim_{R\to+\infty} \lambda_R. \tag{3.15}$$

Consider now the eigenfunction ϕ_R:

$$\begin{aligned}
&-L\phi_R = \lambda_R\phi_R, \quad \text{in } B(0; R), \\
&\phi_R = 0 \text{ on } \partial B(0; R), \\
&\phi_R > 0 \text{ in } B(0; R),
\end{aligned}$$

normalized so that $\phi_R(0) = 1$. We will now, unfortunately, have to appeal to the elliptic regularity estimates – a reader unfamiliar with them may either wait until Chapter 3, or find them in [52, 69] and other classical books on elliptic equations. These estimates say that the L^∞_{loc}-bounds on ϕ_R and the uniform bounds on λ_R imply that the family $\phi_R(x)$ is uniformly bounded in $C^{2,\alpha}_{loc}(\mathbb{R}^n)$ for some $\alpha > 0$. It follows that we may extract a sequence $R_n \to +\infty$ so that the sequence ϕ_{R_n} converges, locally uniformly, to a limit function $\bar{\phi} \in C^{2,\alpha}_{loc}(\mathbb{R}^n)$, which, in addition, satisfies

$$\begin{aligned}
&-L\bar{\phi} = \bar{\lambda}\bar{\phi}, \quad \text{in } \mathbb{R}^n, \\
&\bar{\phi} > 0 \text{ in } \mathbb{R}^n.
\end{aligned}$$

This means that $\mu_1 \geq \bar{\lambda}$, whence $\mu_1 = \bar{\lambda}$, and the proof is complete. □

The above discussion gives just a glimpse at what happens to the principal eigenvalue in unbounded domains – an interested reader should investigate further, starting with the variational formulations of [1] and [99], and continuing with the more recent papers [18, 23] that we have followed in this section.

Chapter 3

Heat kernel bounds

In this chapter we will mostly consider the Cauchy problem for the parabolic equations of the form

$$\begin{aligned}&\phi_t = \nabla\cdot(a(x)\nabla\phi),\\&\phi(0,x) = \phi_0(x),\end{aligned}\tag{0.1}$$

in the whole space $x \in \mathbb{R}^n$, $t > 0$. The diffusion matrix $a(x)$ is assumed to be bounded and uniformly elliptic. We are interested in estimates for the solutions of (0.1) that would exhibit both temporal and spatial decay, as in the heat equation. Let us recall that solutions of the heat equation

$$\psi_t = \Delta\psi,\tag{0.2}$$

with the initial data $\psi(0,x) = \psi_0(x)$ are given by

$$\psi(t,x) = \int_{\mathbb{R}^n} G_0(t,x,y)\psi_0(y)dy.\tag{0.3}$$

The Green's function for the heat equation is given explicitly by

$$G_0(t,x,y) = \frac{1}{(4\pi t)^{n/2}}e^{-(x-y)^2/(4t)}.\tag{0.4}$$

Similarly, solutions of the inhomogeneous diffusion equation (0.1) can be expressed in terms of the Green's function for this problem as

$$\phi(t,x,y) = \int_{\mathbb{R}^n} G(t,x,y)\phi_0(y)dy.\tag{0.5}$$

However, the Green's function is no longer given explicitly and the best we can hope for are interesting bounds for $G(t,x,y)$. We will show that, in some sense, solution of (0.1) behaves "almost exactly" as a solution of the heat equation. More precisely, there exists a constant $C > 0$ so that

$$\frac{1}{Ct^{n/2}}e^{-Cx^2/t} \le G(t,x,y) \le \frac{C}{t^{n/2}}e^{-x^2/(Ct)}.\tag{0.6}$$

This result is originally due to Nash [97]. We will follow here a more recent version of the proof due to Fabes and Stroock [56].

Following Fabes and Stroock we will also explain that the heat kernel estimates (0.6) imply "everything you ever wanted to know about parabolic equations": such as the Hölder regularity of solutions, and Harnack inequality. There is a physical reason for that: thermodynamics tells us that solutions of heat equations tend to equilibrate. The bounds in (0.6) are simply a quantification of that. We will see that the Gaussian bounds imply that the oscillation $\mathrm{Osc}_{D_R}\phi$ of any solution ϕ over a set $D_R = \{|x - x_0| < R, |t - t_0| < R^2\}$ goes down to zero exponentially as $R \to 0$:

$$\mathrm{Osc}_{D_R} < \gamma \mathrm{Osc}_{D_{2R}},$$

with a constant $\gamma < 1$. This implies the Hölder bound on ϕ.

Divergence and non-divergence forms: intuition or integration?

We will consider in this section only equations in the divergence form, occasionally with an incompressible drift:

$$\phi_t + u \cdot \nabla\phi = \nabla \cdot (a(x)\nabla\phi), \tag{0.7}$$

with a prescribed flow $u(x)$ such that $\nabla \cdot u = 0$. The "practical reason" to consider equations in the divergence form is that they are much more amenable to integration by parts than their counterpart in the non-divergence form

$$\phi_t + u \cdot \nabla\phi = a_{ij}(x)\frac{\partial^2\phi}{\partial x_i \partial x_j}. \tag{0.8}$$

Of course, there is also a physical reason: advection-diffusion equations in the divergence form appear in many physical problems where the total mass of u is conserved, there is an external incompressible drift and diffusion is present due to the heterogeneous Fourier law. On the other hand, solutions of the non-divergence form equations (0.8) have a nice probabilistic interpretation. Consider the stochastic differential equation

$$dX_t = -u(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x. \tag{0.9}$$

Here $W_t = (W_t^1, \dots, W_t^n)$ is the n-dimensional Brownian motion, that is, every component W_t^j, $j = 1, \dots, n$ is a standard Brownian motion, and W_t^j and W_t^k are independent for $k \neq j$. The matrix $\sigma(x)$ is symmetric and satisfies $\sigma^2(x) = 2a(x)$. This SDE is related to the PDE (0.8) in a way very similar to the connection between first order hyperbolic equations and ODE's. Let $\phi(t,x)$ be the solution of (0.8) with the initial data $\phi(0,x) = \phi_0(x)$. Then it is given "explicitly" by

$$\phi(t,x) = \mathbb{E}_x(\phi_0(X(t)). \tag{0.10}$$

Here X_t is the solution of the stochastic differential equation (0.9), and the subscript x in $\mathbb{E}_x$ refers to the fact that $X(t)$ starts at the point x at time $t = 0$. It would be far too ambitious for us to review the general theory of diffusions here, the reader may either think of solutions of stochastic differential equations as "randomized" solutions of classical ODEs, or consult [8, 9, 100] that all consider the connections between PDEs and diffusions. A useful exercise for the reader would be to try to convince oneself that the non-divergence form equations can be recovered from the discrete approximation by inhomogeneous random walks on an integer lattice, generalizing what we have done before for the Laplace equation.

This probabilistic interpretation provides a very good intuition for how solutions of the equations in the non-divergence form should behave. Much of this intuition applies also to solutions of equations in the divergence form (though the probabilistic interpretation has to be modified to take into account the additional drift coming from $\nabla a(x)$), and we will often appeal to it even when we consider equations in the divergence form. Of course, the physical intuition about propagation of heat in a heterogeneous environment is also very useful.

1 The Nash inequality

Spreading in the heat equation

Before analyzing the inhomogeneous diffusions, let us first review "why" solutions of the constant coefficients heat equation spread and decay on the mathematical level (on the physical level this is very intuitive – heat likes to equilibrate). One can, of course, deduce everything starting with an explicit expression for the heat kernel:

$$G(t,x) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/(4t)},$$

but such expressions are not available for heterogenous diffusions so we need to learn to live without them. The Nash inequality will be an indispensable tool here.

Ballpark arguments

We start with some integral balances that tell us that solutions spread as $x \sim \sqrt{t}$. There are two basic balances: first, the integral of the solutions of the heat equation

$$\begin{aligned} &\psi_t = \Delta\psi, \\ &\psi(0,x) = \psi_0(x) \end{aligned} \tag{1.1}$$

posed in $\mathbb{R}^n$, with rapidly decaying initial data $\psi_0(x)$, is conserved:

$$\int_{\mathbb{R}^n} \psi(t,x)dx = \int_{\mathbb{R}^n} \psi_0(x)dx := M_0. \tag{1.2}$$

The evolution also preserves positivity: if $\psi_0 \geq 0$ then $\psi(t,x) > 0$ for all $t > 0$. The second balance is for the second moment: multiplying (1.1) by $|x|^2$ and integrating gives

$$\frac{d}{dt}\int_{\mathbb{R}^n} |x|^2\psi(t,x)dx = \int_{\mathbb{R}^n} |x|^2\Delta\psi(t,x)dx = 2n\int_{\mathbb{R}^n} \psi(t,x)dx = 2nM_0, \tag{1.3}$$

so that

$$M_2(t) := \int_{\mathbb{R}^n} |x|^2\psi(t,x)dx = 2nM_0t + \int_{\mathbb{R}^n} |x|^2\psi_0(x)dx. \tag{1.4}$$

Hence, the second moment $M_2(t)$ grows linearly in time, while the total mass stays constant. It follows that solutions of the heat equation have to spread keeping their mass fixed –

otherwise, $M_2(t)$ would not grow in time. We learn from (1.4) two things: first, as an upper bound

$$\int_{\mathbb{R}^n} |x|^2\psi(t,x)dx \le 2nM_0 t + \int_{\mathbb{R}^n} |x|^2\psi_0(x)dx, \tag{1.5}$$

it tells you that "the mass outside of the ball $B(0,R)$ is small for any $R \gg t^{-1/2}$":

$$\int_{|x|\ge Nt^{1/2}} \psi(t,x)dx \le \frac{2nM_0}{N^2} + \frac{1}{N^2 t}\int_{\mathbb{R}^n} |x|^2\psi_0(x)dx. \tag{1.6}$$

On the other hand, as a lower bound

$$\int_{\mathbb{R}^n} |x|^2\psi(t,x)dx \ge 2nM_0 t, \tag{1.7}$$

it tells you that there has to be some mass at distance $O(t^{1/2})$ from the origin – the mass can not be concentrated in a ball $B(0,R)$ of radius $R \ll t^{1/2}$. Hence, solutions of the constant coefficients heat equation have to spread over the ball of radius $O(t^{1/2})$. On the other hand, as they spread in a mass preserving way, the mass balance tells us that its maximum should roughly satisfy

$$I_0 \sim \psi_{max}(t)(t^{1/2})^n, \tag{1.8}$$

hence the maximum should decay as $\psi_{max} \sim t^{-n/2}$. The very last step (1.8), is, of course, just a rough ballpark estimate but a combination of the above bounds lies at the heart of the rigorous proof.

Careful accounting

In order to estimate the decay for the heat equation in a more careful way, let us multiply (1.1) by ψ and integrate:

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n} |\psi(t,x)|^2dx = -\int_{\mathbb{R}^n} |\nabla\psi(t,x)|^2dx. \tag{1.9}$$

If the problem were posed on a torus $\mathbb{T}^n$ and

$$\int_{\mathbb{T}^n} \psi_0(x)dx = 0,$$

then we would have

$$\int_{\mathbb{T}^n} \psi(t,x)dx = 0,$$

for all $t > 0$. Therefore, the Poincaré inequality would hold:

$$\int_{\mathbb{T}^n} |\psi(t,x)|^2dx \le C\int_{\mathbb{T}^n} |\nabla\psi(t,x)|^2dx.$$

Using this in (1.9) we would get

$$frac12\frac{d}{dt}\int_{\mathbb{T}^n} |\psi(t,x)|^2dx \le -C\int_{\mathbb{T}^n} |\psi(t,x)|^2dx, \tag{1.10}$$

implying the exponential intake decay of the L^2-norm of $\psi(t,x)$:

$$\|\psi(t,\cdot)\|_{L^2(\mathbb{T}^n)} \le \|\psi_0\|_{L^2(\mathbb{T}^n)} e^{-Ct}. \tag{1.11}$$

However, in the whole space the Poincaré inequality does not hold, hence we need a different way to relation the dissipation

$$D = \int_{\mathbb{R}^n} |\nabla\psi(t,x)|^2 dx, \tag{1.12}$$

the conserved mass

$$M_0 = \int_{\mathbb{R}^n} \psi(t,x) dx, \tag{1.13}$$

and the L^2-norm of ψ itself. It is given by the Nash inequality.

Theorem 1.1 *(The Nash inequality) There exists a constant $C > 0$, which depends only on the dimension n so that for any $\phi \in H^1(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ we have*

$$\|\nabla\phi\|_{L^2}^2 \ge \frac{C\|\phi\|_{L^2}^{2+4/n}}{\|\phi\|_{L^1}^{4/n}}. \tag{1.14}$$

Proof. Using the Fourier transform

$$\hat\phi(\xi) = \int e^{-2\pi i\xi\cdot x}\phi(x)dx,$$

we have, for any $R > 0$:

$$\begin{aligned}\int |\phi(x)|^2 dx = \int |\hat\phi(\xi)|^2 d\xi &= \int_{|\xi|\le R} |\hat\phi(\xi)|^2 d\xi + \int_{|\xi|\ge R} |\hat\phi(\xi)|^2 d\xi \\ &\le C_n R^n \|\hat\phi\|_{L^\infty}^2 + \frac{1}{R^2}\int_{|\xi|\ge R} |\xi|^2|\hat\phi(\xi)|^2 d\xi.\end{aligned} \tag{1.15}$$

We may now estimate the two terms in the right side as

$$\|\hat\phi\|_{L^\infty} \le \|\phi\|_{L^1},$$

and

$$\int_{|\xi|\ge R} |\xi|^2|\hat\phi(\xi)|^2 d\xi \le \int_{\mathbb{R}^n} |\xi|^2|\hat\phi(\xi)|^2 d\xi = \frac{1}{4\pi^2}\int |\nabla\phi(x)|^2 dx.$$

Going back to (1.15) we conclude that, for any $R > 0$ we have

$$\int |\phi(x)|^2 dx \le C\left[R^n\|\phi\|_{L^1}^2 + R^{-2}\|\nabla\phi\|_{L^2}^2\right]. \tag{1.16}$$

Choosing

$$R = \left(\frac{\|\nabla\phi\|_{L^2}^2}{\|\phi\|_{L^1}^2}\right)^{1/(n+2)}$$

leads to

$$\int |\phi(x)|^2 dx \le C\|\nabla\phi\|_{L^2}^{2n/(n+2)}\|\phi\|_{L^1}^{4/(n+2)},$$

which is the same as (1.14). $\Box$

Exercise 1.2 *The L^1-norm of the solutions of the heat equation does not increase in time: if $\psi(t,x)$ satisfies (1.1) then*

$$\|\psi(t)\|_{L^1} \leq I_0 := \|\psi_0\|_{L^1} \textit{for all } t \geq 0. \tag{1.17}$$

The Nash inequality together with (1.17) implies that the L^2-norm of the solution of (1.1) satisfies

$$\|\nabla\phi\|_{L^2}^2 \geq \frac{C}{I_0^{4/n}}\|\phi\|_{L^2}^{2+4/n}. \tag{1.18}$$

Using this in (1.9) gives

$$\frac{dM}{dt} \leq -\frac{C}{I_0^{4/n}}\left(M(t)\right)^{1+2/n}, \tag{1.19}$$

with

$$M(t) = \int |\phi(t,x)|^2 dx.$$

Integrating this ODE in time gives

$$\frac{1}{M(0)^{2/n}} - \frac{1}{M(t)^{2/n}} \leq -\frac{Ct}{M_0^{4/n}}. \tag{1.20}$$

It follows that the decay of the L^2-norm can be estimated as

$$\|\psi(t)\|_{L^2}^2 = M(t) \leq \frac{CM_0^2}{t^{n/2}} = \frac{C}{t^{n/2}}\|\psi_0\|_{L^1}^2. \tag{1.21}$$

This estimate could, of course, be easily obtained directly from the explicit form of the heat kernel but our goal here is exactly the opposite: to devise a method that would work without the explicit formulas. This is just a baby example of how the strategy works: one starts with physical balances and then tries to estimate the dissipation in terms of the conserved quantities.

2 The temporal decay

Divergence form equations

We now show how the above strategy using the Nash inequality can be used to obtain the temporal decay of solutions of the parabolic problem

$$\begin{aligned} \phi_t &= \nabla\cdot(a(x)\nabla\phi), \\ \phi(0,x) &= \phi_0(x), \end{aligned} \tag{2.1}$$

in the whole space $x \in \mathbb{R}^n$, $t > 0$. The proof is essentially identical to what we have done above for the heat equation. We assume that the matrix $a(x)$ is bounded and uniformly elliptic: for any $\xi \in \mathbb{R}^n$ and all $x \in \mathbb{R}^n$ we have

$$\lambda|\xi|^2 \leq a_{ij}(x)\xi_i\xi_j \leq \Lambda|\xi|^2, \tag{2.2}$$

with some $\lambda, \Lambda > 0$. The next theorem shows that solutions of (2.1) obey the same decay bounds as solutions of the heat equation with constant coefficients.

Theorem 2.1 *Let the diffusion matrix $a(x)$ satisfy (2.2) and assume that the initial data $\psi_0(x)$ is sufficiently rapidly decaying. There exists a constant C that depends only on the dimension n so that the function $\psi(t,x)$ satisfies*

$$|\phi(t,x)| \leq \frac{C}{(\lambda t)^{n/2}} \|\phi_0\|_{L^1}, \tag{2.3}$$

for all $x \in \mathbb{R}^n$ and all $t > 0$.

The estimate (2.3) is exactly the same as for the solutions of the heat equation with a constant diffusivity. The $L^1 - L^\infty$ decay (small initial L^1-norm implies small L^∞-norm at times $t > 0$) comes from physics: if the initial data has small mass, then solution will "spread around", and, as it has to preserve the total mass, it has no choice but to have a small L^∞-norm. It also gives a guess of a scale on which the solution will spread by time t: as the total mass is preserved, and the maximum decays at least as $t^{-n/2}$, then the "spreading scale" $L(t)$ should satisfy, roughly,

$$\int_{\mathbb{R}^n} \psi_0(x)dx = \int_{\mathbb{R}^n} \psi(t,x)dx \approx |L(t)|^n \frac{1}{t^{n/2}},$$

thus

$$L(t) \sim \sqrt{t},$$

the familiar estimate for diffusive spreading. Note, however, that here this estimate holds not just for the solution of the heat equation with a constant diffusivity but for a much more general class of equations.

The proof of this theorem proceeds as in the homogeneous case. First, we will show, using the Nash inequality that $\phi(t,x)$ satisfies an $L^1 \to L^2$ decay estimate:

$$\|\phi(t)\|_{L^2} \leq \frac{C}{(\lambda t)^{n/4}} \|\phi_0\|_{L^1}. \tag{2.4}$$

Next, we use a bit of functional analysis. Define the solution operator S_t as the mapping of the initial data ϕ_0 to the solution of (2.1) at time t:

$$S_t[\phi_0] = \phi(t).$$

The estimate (2.4) means that S_t is a bounded operator from L^1 to L^2 for each $t > 0$, with its norm bounded as

$$\|S_t\|_{L^1 \to L^2} \leq \frac{C}{(\lambda t)^{n/4}}. \tag{2.5}$$

Therefore, the adjoint operator S_t^* maps L^2 to L^∞ with the bound

$$\|S_t^*\|_{L^2 \to L^\infty} \leq \frac{C}{(\lambda t)^{n/4}}. \tag{2.6}$$

We claim that the operator S_t is self-adjoint. Indeed, let $\phi(t,x)$ and $\psi(t,x)$ be the solutions of (2.1) with the initial data $\phi(0,x) = f(x)$ and $\psi(0,x) = g(x)$. Symmetry of S_t means that

$$\int f(x)\psi(t,x)dx = \int g(x)\phi(t,x)dx. \tag{2.7}$$

In order to see that this identity holds for all $t \geq 0$, set

$$B(s) = \int \phi(s,x)\psi(t-s,x)dx,$$

then

$$\frac{dB}{ds} = \int [\nabla \cdot (a(x)\nabla\phi(s,x))\psi(t-s,x) - \phi(s,x)\nabla \cdot (a(x)\nabla\psi(t-s,x))dx$$
$$= \int [(a(x)\nabla\phi(s,x)) \cdot \nabla\psi(t-s,x) - \nabla\phi(s,x) \cdot (a(x)\nabla\psi(t-s,x))dx = 0.$$

It follows that

$$B(s) = B(0) \text{ for all } 0 \leq s \leq t.$$

Setting $s = t$ gives (2.7). Hence, the solution operator S_t is, indeed, self-adjoint and (2.6) means nothing but

$$\|S_t\|_{L^2 \to L^\infty} \leq \frac{C}{(\lambda t)^{n/4}}. \tag{2.8}$$

The next observation is that the operators S_t form a semi-group so that

$$S_t = S_{t/2} \circ S_{t/2}, \tag{2.9}$$

which simply says that solving the Cauchy problem with the data given at $t = 0$ until a time $T > 0$ is equivalent to solving it until the time $T/2$, and using the result as the initial data to run the evolution again for the time $T/2$. As $S_{t/2}$ maps L^1 to L^2 and, as we have just shown, it also maps L^2 to L^∞, we know from (2.9) that S_t maps L^1 to L^∞ with the norm bounded as

$$\|S_t\|_{L^1 \to L^\infty} \leq \|S_{t/2}\|_{L^1 \to L^2}\|S_{t/2}\|_{L^2 \to L^\infty} \leq \frac{C}{(\lambda t)^{n/4}}\frac{C}{(\lambda t)^{n/4}} = \frac{C'}{(\lambda t)^{n/2}}. \tag{2.10}$$

This exactly means that estimate (2.3) holds. Therefore, it only remains to prove the $L^1 \to L^2$ estimate (2.4).

In order to show that (2.4) holds we proceed essentially exactly as in the heat equation case: multiply (2.1) by ϕ and integrate:

$$\frac{1}{2}\frac{d}{dt}\int |\phi(t,x)|^2 dx = -\int (a(x)\nabla\phi(t,x) \cdot \nabla\phi(t,x))dx \leq -\lambda \int |\nabla\phi(t,x)|^2 dx. \tag{2.11}$$

We also integrate (2.1) in space to get

$$\int \phi(t,x)dx = \int \phi_0(x)dx := M_0. \tag{2.12}$$

We may assume that $\phi_0(x) \geq 0$, otherwise we decompose $\phi = \phi_1 - \phi_2$. Here ϕ_1 and ϕ_2 are solutions ϕ_1 and ϕ_2 of (2.1) with the initial data $\phi^+(x)$ and $\phi^-(x)$, respectively. The bound we prove for solutions with non-negative initial data will apply both to ϕ_1 and ϕ_2, hence to their difference ϕ.

If $\phi_0 \geq 0$, then (2.12) means that $\|\phi(t)\|_{L^1} = M_0$ for all $t > 0$. The Nash inequality implies then that

$$\int |\nabla\phi(t,x)|^2 dx \geq \frac{C}{M_0^{4/n}} \left(M(t)\right)^{1+2/n}. \tag{2.13}$$

Here we have set

$$M(t) = \int |\phi(t,x)|^2 dx.$$

We may now rewrite the inequality (2.11) as

$$\frac{dM}{dt} \leq -\frac{C\lambda}{I_0^{4/n}} \left(M(t)\right)^{1+2/n}. \tag{2.14}$$

Integrating this ODE in time gives

$$\frac{1}{M(0)^{2/n}} - \frac{1}{M(t)^{2/n}} \leq -\frac{C\lambda t}{M_0^{4/n}}. \tag{2.15}$$

It follows that

$$M(t) \leq \frac{CM_0^2}{(\lambda t)^{n/2}}. \tag{2.16}$$

This is exactly (2.4), hence the proof of Theorem 2.1 is complete. □

Equations with an incompressible drift

It turns out that the previous argument can be easily generalized to the Cauchy problem for parabolic equations with an incompressible drift, yielding decay estimates that are uniform in the drift. Consider the initial value problem

$$\begin{aligned} &\phi_t + u \cdot \nabla\phi = \nabla \cdot (a(x)\nabla\phi), \\ &\phi(0,x) = \phi_0(x), \end{aligned} \tag{2.17}$$

with a uniformly elliptic matrix $a(x)$ satisfying (2.2), and a divergence-free flow $u(x)$:

$$\nabla \cdot u(x) = 0 \text{ in } \mathbb{R}^n. \tag{2.18}$$

The divergence-free condition (2.18) means that the fluid is incompressible, that is, the solution map of an ODE

$$\dot{X}(t;x) = u(X), \quad X(0;x) = x \tag{2.19}$$

is measure-preserving. In other words, given any measurable set A and any $t > 0$ we have the following property: the Lebesgue measure of A equals to the Lebesgue measure of the set

$$A(t) = \{y \in \mathbb{R}^n : \ y = X(t;x) \text{ for some } x \in A\}. \tag{2.20}$$

That is, the set A is not compressed, hence the term ”incompressible”. This property plays an enormously important role in the theory of fluids.

Theorem 2.2 *Solutions of (2.17) satisfy the estimate*

$$|\phi(t,x)| \leq \frac{C}{(\lambda t)^{n/2}}\|\phi_0\|_{L^1}, \tag{2.21}$$

with a constant $C > 0$ that depends only dimension n and does not depend on the flow u, provided that the incompressibility constraint (2.18) holds.

The assumption that the flow $u(x)$ is divergence-free is very important and the conclusion of this theorem is false without this condition. The reason why estimate (2.21) holds with a constant that is uniform in all incompressible flows can be seen from the probabilistic interpretation of the solutions of (2.17) in the special case $a(x)$ is the identity matrix. Let X_t be the solution of a stochastic differential equation

$$dX_t = -u(X_t)dt + \sqrt{2}dW_t, \quad X_0 = x. \tag{2.22}$$

Here $W_t = (W_t^1, \ldots, W_t^n)$ is the n-dimensional Brownian motion. As we have mentioned, solution of the parabolic equation (2.17) can be written as[1]

$$\phi(t,x) = \mathbb{E}(\phi_0(X(t)). \tag{2.23}$$

Let us consider for simplicity the case when $\phi_0(x)$ is the characteristic function of a set $A \subset \mathbb{R}^n$, then

$$\phi(t,x) = \mathbb{P}(X(t) \in A), \tag{2.24}$$

and (2.21) says that

$$\mathbb{P}(X(t) \in A) \leq \frac{C}{(\lambda t)^{n/2}}|A|. \tag{2.25}$$

If the flow $u(x)$ is divergence free then the solution map $S_t : x \to y(t)$ of the ODE without diffusion,

$$\dot{y} = -u(y), \quad y(0) = x, \tag{2.26}$$

is measure preserving and thus "mixing things around". "Therefore", it is unable to keep the particle in any given set in the presence of a diffusion, and that is reflected in estimate (2.25) – the probability to visit a given set A tends to zero as $t \to +\infty$ uniformly in the flow u.

Proof. The proof follows that of Theorem 2.1 with one modification. We multiply (2.17) by ϕ and integrate. As u is divergence-free, the term involving the drift vanishes:

$$\int (u \cdot \nabla\phi)\phi dx = \frac{1}{2}\int u \cdot \nabla(\phi^2)dx = -\frac{1}{2}\int \phi(\nabla \cdot u)dx = 0. \tag{2.27}$$

This cancellation means that we still have the identity (2.9):

$$\frac{1}{2}\frac{d}{dt}\int |\phi(t,x)|^2 dx = -\int (a(x)\nabla\phi(t,x) \cdot \nabla\phi(t,x))dx \leq -\lambda \int |\nabla\phi(t,x)|^2 dx. \tag{2.28}$$

[1] As $a(x)$ is the identity matrix, there is no difference between the divergence and non-divergence forms of the equation, hence we may use the probabilistic interpretation directly.

Moreover, as

$$\int (u \cdot \nabla \phi) dx = -\int \phi (\nabla \cdot u) dx = 0, \tag{2.29}$$

the integral of ϕ is still preserved:

$$\int \phi(t,x) dx = \int \phi_0(x) dx. \tag{2.30}$$

Therefore, using the Nash inequality we may proceed as in the proof of Theorem 2.1 to obtain

$$\|\phi(t)\|_{L^2} \le \frac{C}{(\lambda t)^{n/4}} \|\phi_0\|_{L^1}, \tag{2.31}$$

with a constant $C > 0$ that depends only on dimension n and not on the flow u. Once again, that means that the solution operator S_t for (2.17) satisfies the bound

$$\|S_t\|_{L^1 \to L^2} \le \frac{C}{(\lambda t)^{n/4}}, \tag{2.32}$$

and its adjoint satisfies

$$\|S_t^*\|_{L^2 \to L^\infty} \le \frac{C}{(\lambda t)^{n/4}}, \tag{2.33}$$

However, S_t is not self-adjoint when $u \not\equiv 0$. Rather, the adjoint operator S_t^* is the solution operator for the Cauchy problem

$$\begin{aligned} &\psi_t - u \cdot \nabla \psi = \nabla \cdot (a(x) \nabla \psi), \\ &\psi(0,x) = \psi_0(x). \end{aligned} \tag{2.34}$$

To verify this, set

$$B(s) = \int \phi(s,x) \psi(t-s,x) dx,$$

then

$$\begin{aligned} \frac{dB}{ds} &= \int [\nabla \cdot (a(x) \nabla \phi(s,x)) - u(x) \cdot \nabla \phi(s,x)] \psi(t-s,x) dx \\ &- \int \phi(s,x) [\nabla \cdot (a(x) \nabla \psi(t-s,x)) + u(x) \cdot \nabla \psi(t-s,x)] dx \\ &= \int [(a(x) \nabla \phi(s,x) \cdot \nabla \psi(t-s,x)) - \psi(t-s,x)(u(x) \cdot \nabla \phi(s,x))] dx \\ &- \int [(a(x) \nabla \psi(t-s,x) \cdot \nabla \phi(s,x)) - \psi(t-s,x)(u(x) \cdot \nabla \phi(s,x))] dx \\ &+ \int \psi(t-s,x) \phi(s,x) \nabla \cdot u(x) dx = 0, \end{aligned}$$

since $\nabla \cdot u = 0$. Therefore, $B(0) = B(t)$, that is,

$$\int \phi(0,x) \psi(t,x) dx = \int \phi(t,x) \psi(0,x),$$

which means exactly that S_t^* is the solution operator for (2.34). However, (2.34) has the same form as our original problem (2.17), with the flow u replaced by $(-u)$, which is also incompressible. Hence, from what we have already proved we know that

$$\|S_t^*\|_{L^1\to L^2} \le \frac{C}{(\lambda t)^{n/4}}. \tag{2.35}$$

This, in turn, implies that

$$\|S_t\|_{L^2\to L^\infty} \le \frac{C}{(\lambda t)^{n/4}}. \tag{2.36}$$

The rest is as in the proof of Theorem 2.1: the semigroup property implies that $S_t = S_{t/2}\circ S_{t/2}$ whence

$$\|S_t\|_{L^1\to L^\infty} \le \|S_{t/2}\|_{L^1\to L^2}\|S_t\|_{L^2\to L^\infty} \le \frac{C}{(\lambda t)^{n/2}}.$$

Therefore, (2.21) holds. □

3 Elliptic problems with an incompressible drift

Another application of the Nash inequality is to elliptic problems with an incompressible drift.

Theorem 3.1 *Let the flow $u(x)$ be divergence-free and let $\phi(x)$ be the solution of the elliptic problem*

$$\begin{aligned} &-\nabla\cdot(a(x)\nabla\phi) + u\cdot\nabla\phi = f(x) \ \textit{in}\ \Omega, \\ &\phi = 0 \ \textit{on}\ \partial\Omega, \end{aligned} \tag{3.1}$$

with $f(x)\in L^p(\Omega)$, $p > n/2$. There exists a constant $C(\Omega,n,p) > 0$ which depends on p, the ellipticity constant λ of the matrix a, and the domain Ω but not on the flow $u(x)$, so that

$$\|\phi\|_{L^\infty(\Omega)} \le C\|f\|_{L^p(\Omega)}. \tag{3.2}$$

The spirit of this theorem is very close to that of Theorem 2.2. Estimate (3.2) always holds for any flow u, whether divergence free or not – this is a standard elliptic regularity bound [69] – but with a constant C that depends on u in an uncontrolled way. The point is that the same constant in (3.2) works for all divergence free flows.

Exercise 3.2 Construct a flow u in the unit ball $B = \{|x|\le 1\}\subset\mathbb{R}^n$ which is not divergence-free, so that for the functions $\phi_A(x)$ which satisfy

$$\begin{aligned} &-\Delta\phi_A + Au\cdot\nabla\phi_A = 1 \text{ in } B, \\ &\phi_A = 0 \text{ on } \partial B, \end{aligned} \tag{3.3}$$

we have

$$\lim_{A\to+\infty}\phi_A(0) = +\infty. \tag{3.4}$$

The reason why estimate (3.1) holds, can be seen, once again, from the probabilistic interpretation of the solutions of (3.1). Let X_t be the solution of the SDE

$$dX_t = -u(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x, \tag{3.5}$$

with a matrix σ such that $\sigma(x)\sigma^T(x) = 2a(x)$, and let τ be the first exit time from the domain Ω for the process X_t. Then solution of the boundary value problem (3.1) can be written as

$$\phi(x) = \mathbb{E}_x\left(\int_0^\tau f(X(s))ds\right). \tag{3.6}$$

Exercise 3.3 A reader not familiar with the stochastic differential equations may think of a discrete equation on the on-dimensional lattice

$$u_n = \frac{1}{2}(u_{n-1} + u_{n+1}) + f_n, \quad 0 \le n \le N, \tag{3.7}$$

with the boundary condition $u_0 = u_N = 0$. Express the solution u_n in terms of a standard random walk and get an analog of (3.6).

Let us assume, once again, that $f(x) = \chi_A(x)$ is the characteristic function of a set A. Then (3.6) takes the form:

$$\phi(x) = \mathbb{E}_x\left(T_A\right), \tag{3.8}$$

where T_A is the total time the process X_t spends in the set A before exiting from Ω, and (3.2) says that

$$\mathbb{E}_x\left(T_A\right) \le C_p|A|^{1/p}, \tag{3.9}$$

for any $p > n/2$. This means that a combination of an incompressible flow and a diffusion can not keep a particle in any given set for too long time – the expected value of the exit time is bounded from above by a constant that depends only on the Lebesgue measure of A. If the flow is not divergence free then this is clearly not true – if a very strong flow points radially toward a given point then the particle will take a very long time to escape a small ball centered at that point, as shows Exercise 3.2.

One may wonder if we have some sort of a uniform lower bound for ϕ also: whether we can say, for instance, that if a ball $B(x_0, r)$ is contained strictly inside Ω, then solutions of (taking $a(x) = \mathrm{Id}$)

$$\begin{aligned} &-\Delta\phi + u\cdot\nabla\phi = 1 \text{ in } \Omega, \\ &\phi = 0 \text{ on } \partial\Omega, \end{aligned} \tag{3.10}$$

obey a uniform lower bound: $\phi(x) \ge C$ on $B(x_0, r)$ with the constant $C > 0$ that does not depend on u as long as u is divergence-free. The probabilistic interpretation for solutions of (3.10) is simple:

$$\phi(x) = \mathbb{E}_x(\tau), \tag{3.11}$$

where τ is the time the process X_t, solution of

$$dX_t = -u(X_t)dt + \sqrt{2}dW_t, \quad X_0 = x, \tag{3.12}$$

spends inside Ω before it exits this domain. It turns out that there is no lower bound on $\phi(x)$ that would be uniform in u – the reason is, roughly, that if u is very fast and very mixing then the particle will exit Ω very quickly with a very high probability – see [16, 32] for various results of this kind.

Proof of Theorem 3.1

We may assume that $f(x) \geq 0$ without loss of generality – if not, we decompose $f = f^+ - f^-$ and $\phi = \phi^+ - \phi^-$, where ϕ^+ and ϕ^- are solutions of (3.1) with f replaced by f^+ and f^-, respectively. We write $\phi(x)$, the solution of (3.1), as

$$\phi(x) = \int_0^\infty \psi(t,x)dt. \tag{3.13}$$

The function $\psi(t,x)$ satisfies the parabolic initial value problem

$$\begin{aligned} &\psi_t - \nabla \cdot (a(x)\nabla\psi) + u \cdot \nabla\psi = 0 \text{ in } \Omega, \\ &\psi(t,x) = 0 \text{ on } \partial\Omega, \\ &\psi(0,x) = f(x) \text{ in } \Omega. \end{aligned} \tag{3.14}$$

We will now show that there exists a pair of constants $C > 0$ and $\alpha > 0$ so that for any incompressible flow u and any solution of (3.14) with initial data $f(x)$, we have a uniform bound

$$|\psi(t,x)| \leq \frac{Ce^{-\alpha t}}{t^r}\|f\|_{L^1}, \tag{3.15}$$

with any $r > n/2$. The proof is close to that of Theorem 2.2, with a slight modification, we present the details for the convenience of the reader. First, multiplying (3.14) by ψ and integrating by parts we obtain

$$\frac{1}{2}\frac{d}{dt}\|\psi\|_2^2 = -\int_\Omega (a(x)\nabla\psi \cdot \nabla\psi)dx \leq -\lambda\|\nabla\psi\|_{L^2}^2. \tag{3.16}$$

Using the Poincaré inequality

$$\|\psi\|_2 \leq C_p\|\nabla\psi\|_{L^2}, \tag{3.17}$$

for all functions $\psi \in H_0^1(\Omega)$, we conclude that there exists a constant $\alpha > 0$ so that

$$\|\psi(t_2)\|_2 \leq e^{-\alpha(t_2-t_1)}\|\psi(t_1)\|_2 \tag{3.18}$$

for any pair of times $t_2 \geq t_1 \geq 0$.

In order to estimate the dissipation term in (3.16) we will use the following Nash-type inequality in Ω.

Lemma 3.4 *For all $0 < s < 4/n$ there exists a constant C that depends on Ω and s so that for all smooth functions ϕ such that $\phi = 0$ on $\partial\Omega$, we have*

$$\|\nabla\phi\|_{L^2}^2 \geq C\frac{\|\phi\|_{L^2}^{s+2}}{\|\phi\|_{L^1}^s}. \tag{3.19}$$

Proof. The Poincaré inequality implies that we have

$$\|\phi\|_{L^q} \leq C_q\|\nabla\phi\|_{L^2}, \text{ for all } 1 < q < \frac{2n}{n-2}.$$

Next, using the Hölder inequality, with $1/\alpha + 1/\beta = 1$ we obtain:

$$\|\phi\|_{L^2}^2 = \int |\phi|^2 \le \left(\int |\phi|\right)^{1/\alpha} \left(\int |\phi|^{(2-1/\alpha)\beta}\right)^{1/\beta} \le C\|\phi\|_{L^1}^{1/\alpha}\|\nabla\phi\|_{L^2}^{2-1/\alpha},$$

provided that

$$\left(2-\frac{1}{\alpha}\right)\beta = \left(2-\frac{1}{\alpha}\right)\frac{\alpha}{\alpha-1} = \frac{2\alpha-1}{\alpha-1} < \frac{2n}{n-2},$$

or, equivalently:

$$\alpha > (n+2)/4. \tag{3.20}$$

Therefore, we have

$$\|\nabla\phi\|_{L^2}^2 \ge C\frac{\|\phi\|_{L^2}^{4\alpha/(2\alpha-1)}}{\|\phi\|_{L^1}^{2/(2\alpha-1)}} = C\frac{\|\phi\|_{L^2}^{s+2}}{\|\phi\|_{L^1}^{s}},$$

with $s = 2/(2\alpha-1)$, that is, for $s < 4/n$. □

We continue the proof of Theorem 3.1. Using Lemma 3.4 we may rewrite (3.16) as

$$\frac{1}{2}\frac{d}{dt}\|\psi\|_{L^2}^2 \le -C\frac{\|\psi\|_{L^2}^{s+2}}{\|\psi\|_{L^1}^{s}}. \tag{3.21}$$

In order to estimate the L^1-norm above we integrate (3.14) over Ω:

$$\frac{d}{dt}\int_\Omega \psi dx = \int_{\partial\Omega}(a(x)\nabla\psi\cdot\nu)dy, \tag{3.22}$$

as

$$\int_\Omega (u\cdot\nabla\psi)dx = 0$$

because u is divergence-free and $\psi = 0$ on $\partial\Omega$. Here, ν is the outward normal to Ω. The parabolic maximum principle implies that $\psi(t,x) > 0$ for $x \in \Omega$ and $t > 0$, hence $\nabla\psi\cdot v < 0$ for any vector v such that $v\cdot\nu > 0$. The matrix $a(x)$ is positive-definite, hence $v = a(x)\nu$ satisfies this condition for all $x \in \partial\Omega$, thus

$$(a(x)\nabla\psi\cdot\nu) < 0 \text{ on } \partial\Omega.$$

We conclude from (3.22) that

$$\|\psi(t)\|_{L^1} = \int_\Omega \psi(t,x)dx \le \int_\Omega f(x)dx. \tag{3.23}$$

Using this inequality in (3.21) gives

$$\frac{1}{2}\frac{d}{dt}\|\psi\|_{L^2}^2 \le -C\frac{\|\psi\|_{L^2}^{s+2}}{\|f\|_{L^1}^{s}}, \tag{3.24}$$

hence $M(t) = \|\psi(t)\|_{L^2}$ satisfies

$$\frac{1}{M^{s+1}(t)}\frac{dM}{dt} \le -\frac{C}{\|f\|_{L^1}^{s}}.$$

Integrating in time we obtain

$$\frac{1}{M^s(0)} - \frac{1}{M^s(t)} \leq -\frac{Ct}{\|f\|_{L^1}^s}.$$

Setting $q = 1/s$, we conclude that

$$\|\psi(t)\|_{L^2} \leq \frac{C}{t^q}\|f\|_{L^1}, \quad \text{for any } q > n/4. \tag{3.25}$$

Consider now the solution operator $\mathcal{P}_t: \ \psi_0 \to \psi(t)$. We have shown in (3.18) that

$$\|\mathcal{P}_t\|_{L^2\to L^2} \leq Ce^{-\alpha t}, \tag{3.26}$$

while (3.25) says that

$$\|\mathcal{P}_t\|_{L^1\to L^2} \leq \frac{C}{t^q}, \quad \text{for any } q > n/4. \tag{3.27}$$

Once again, using the semi-group property we can write $\mathcal{P}_t = \mathcal{P}_{t/2} \circ \mathcal{P}_{t/2}$, and deduce that

$$\|\mathcal{P}_t\|_{L^1\to L^2} \leq \|\mathcal{P}_{t/2}\|_{L^1\to L^2}\|\mathcal{P}_{t/2}\|_{L^2\to L^2} \leq \frac{Ce^{-\alpha t/2}}{t^q}. \tag{3.28}$$

As we have already discussed, the adjoint operator $\mathcal{P}_t^*$ is simply the solution operator corresponding to the (also incompressible) flow $(-u)$. Therefore, we have the dual bound

$$\|\mathcal{P}_t^*\|_{L^1\to L^2} \leq \frac{Ce^{-\alpha t/2}}{t^{1/s}},$$

which in turn implies that

$$\|\mathcal{P}_t\|_{L^2\to L^\infty} \leq \frac{Ce^{-\alpha t/2}}{t^{1/s}}.$$

Putting these bounds together we obtain

$$\|\psi(t)\|_\infty = \|\mathcal{P}_t f\|_\infty = \|\mathcal{P}_{t/2} \circ \mathcal{P}_{t/2} f\|_\infty \leq \|\mathcal{P}_{t/2}\|_{L^2\to L^\infty}\|\mathcal{P}_{t/2}\|_{L^1\to L^2}\|f\|_1 \leq \frac{C_q e^{-\alpha t/2}}{t^{2q}}\|f\|_1,$$

which is (3.15).

The maximum principle also implies that we have a trivial bound

$$\|\psi\|_{L^\infty} \leq \|f\|_{L^\infty}. \tag{3.29}$$

Interpolating between these two bounds[2] we get the estimate

$$\|\psi(t)\|_{L^\infty} \leq \frac{C_\varepsilon e^{-\alpha_p t}}{t^{n/(2p)+\varepsilon}}\|f\|_{L^p}, \tag{3.30}$$

[2] We use here the Riesz-Thorin interpolation theorem [75]. The corollary that we need says that if an operator A is a bounded linear operator from L^1 to L^∞ and also from L^∞ to L^∞ with $\|A\|_{L^\infty\to L^\infty} \leq 1$, then A is also a bounded operator from L^p to L^∞ for any $p \in (1,\infty)$, with the norm bounded by $\|A\|_{L^p\to L^\infty} \leq \|A\|_{L^1\to L^\infty}^{1/p}$.

for any $\varepsilon > 0$. Now, (3.13) implies that

$$\|\phi\|_{L^\infty} \le C_\varepsilon \|f\|_{L^p} \int_0^\infty \frac{e^{-\alpha_p t}}{t^{n/(2p)+\varepsilon}} dt. \tag{3.31}$$

Note that for any $p > n/2$ we may choose $\varepsilon > 0$ sufficiently small so that the time integral in (3.31) is finite. It follows that

$$\|\phi\|_{L^\infty} \le C\|f\|_{L^p},$$

and the constant $C > 0$ is independent of the incompressible flow u. This finishes the proof of Theorem 3.1. □

4 The Gaussian bounds

The upper bounds in Theorems 2.1 and 2.2 are sharp – they have the correct decay in time as $t \to +\infty$ – but have no information about the spatial decay of solutions. One may not expect "uniform in an incompressible flow" bounds that would include the spatial decay – the spreading rates in space do depend on the presence of a flow. Thus, for the sake of simplicity we only obtain Gaussian bounds on solutions of the parabolic equations in the divergence form. An interested reader should consult [98] for an extension of these bounds to equations with an advection term when the coefficients are periodic. The matrix $a(x)$ is assumed to satisfy the usual uniform ellipticity condition:

$$\lambda|\xi|^2 \le (a(x)\xi \cdot \xi) \le \Lambda|\xi|^2, \tag{4.1}$$

for all $t > 0$, $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^n$. As we have mentioned, we will follow the proof of Fabes and Stroock [56]. Consider the Cauchy problem:

$$\begin{aligned} &\phi_t = \nabla \cdot (a(x)\nabla\phi), \quad t > 0, \quad x \in \mathbb{R}^n, \\ &\phi(0,x) = g(x). \end{aligned} \tag{4.2}$$

Solution of (4.2) can be written in terms of Green's function $\Gamma(t,x,y)$ for (4.2) as

$$\phi(t,x) = \int_{\mathbb{R}^n} \Gamma(t,x,y) g(y) dy. \tag{4.3}$$

Recall that $\Gamma(t,x,y)$ is the solution (in the sense of distributions) of the initial value problem

$$\begin{aligned} &\frac{\partial\Gamma(t,x,y)}{\partial t} = \nabla_x \cdot (a(x)\nabla\Gamma(t,x,y)), \quad t > 0, \quad x \in \mathbb{R}^n \\ &\Gamma(0,x,y) = \delta(x-y), \quad x \in \mathbb{R}^n. \end{aligned} \tag{4.4}$$

We will prove the following theorem.

Theorem 4.1 *There exists a constant $C > 0$ that depends only on the ellipticity constants λ and Λ of the matrix $a(t,x)$, and dimension n so that*

$$\frac{1}{Ct^{n/2}} e^{-C|x-y|^2/t} \le \Gamma(t,x,y) \le \frac{C}{t^{n/2}} e^{-|x-y|^2/(Ct)}, \tag{4.5}$$

for all $0 \le s < t$.

This result generalizes essentially verbatim to equations (4.1) with a diffusivity matrix $a(t,x)$ that depends both on t and x as long as the ellipticity condition (4.1) still holds for all $t \geq 0$ and all $x \in \mathbb{R}^n$. Then solution of the Cauchy problem starting at $t = s$:

$$\begin{aligned} &\phi_t = \nabla \cdot (a(t,x)\nabla\phi), \quad t > s, \quad x \in \mathbb{R}^n, \\ &\phi(s,x) = g(x) \end{aligned} \tag{4.6}$$

is expressed via Green's function (which depends now both on t and s) as

$$\phi(t,x) = \int_{\mathbb{R}^n} \Gamma(t,s,x,y)g(y)dy. \tag{4.7}$$

One can show that

$$\frac{1}{C(t-s)^{n/2}} e^{-C|x-y|^2/(t-s)} \leq \Gamma(t,s,x,y) \leq \frac{C}{(t-s)^{n/2}} e^{-|x-y|^2/[C(t-s)]}. \tag{4.8}$$

The proof is nearly identical to what we will present except for somewhat more cumbersome notation so we will stick to the case when the diffusivity matrix $a(x)$ is time-independent.

4.1 The proof of the upper bound

We first prove the upper bound.

Theorem 4.2 *There exists a constant $C > 0$ that depends on the dimension n so that*

$$\Gamma(t,x,y) \leq \frac{C}{(\lambda t)^{n/2}} e^{-|x-y|^2/8\Lambda t)}, \tag{4.9}$$

for all $t > 0$.

The constant 8 in the exponent in (4.9) is, of course, not optimal, we will point out the place in the proof where we lose the optimality: it can be replaced by any constant larger than 4, giving the upper bound

$$\Gamma(t,x,y) \leq \frac{C_\delta}{(\lambda t)^{n/2}} e^{-|x-y|^2/((4+\delta)\Lambda t)}, \tag{4.10}$$

with any $\delta > 0$, which is nearly optimal, as can be seen in the special case $a(x) \equiv \Lambda$.

The test case: the heat equation and exponential solutions

Let us first explain how the Gaussian decay in the upper bound on the heat kernel for the standard heat equation

$$u_t = \Delta u, \quad t > 0, \ x \in \mathbb{R}^n, \tag{4.11}$$

can be obtained with a minimum of explicit formulas. Note that for any $\alpha \in \mathbb{R}^n$ the function

$$\phi(t,x;\alpha) = \exp\{\alpha \cdot x + \alpha^2 t\} \tag{4.12}$$

is a solution of the heat equation (4.11). Let us think of it as a super-solution (and not as a solution). Then, the function

$$v(t,x) = \inf_{\alpha \in \mathbb{R}^n} \phi(t,x;\alpha) \tag{4.13}$$

is a super-solution:

$$v_t - \Delta v \geq 0. \tag{4.14}$$

This is a reflection of a general principle: infimum of super-solutions is a super-solution but we may also compute it directly. Indeed, we can evaluate the infimum explicitly and write

$$v(t,x) = \phi(t,x;\bar{\alpha}(t,x)) = e^{-|x|^2/(4t)}, \quad \bar{\alpha}(t,x) = -\frac{x}{2t}. \tag{4.15}$$

This gives an explicit super-solution that obeys a Gaussian bound but it does not give the factor $t^{n/2}$ yet. In order to improve the super-solution (make it smaller) and incorporate the Gaussian bound, we note that (4.14) is a strict inequality. In order to see that, we compute:

$$\frac{\partial v}{\partial x_k} = \frac{\partial \phi}{\partial x_k} + \sum_{j=1}^{n} \frac{\partial \phi}{\partial \alpha_j} \frac{\partial \bar{\alpha}_j}{\partial x_k}$$

and

$$\Delta v = \sum_{k=1}^{n} \frac{\partial^2 \phi}{\partial {x_k}^2} + 2 \sum_{k,j=1}^{n} \frac{\partial^2 \phi}{\partial x_k \partial \alpha_j} \frac{\partial \bar{\alpha}_j}{\partial x_k} + \sum_{k,j,m=1}^{n} \frac{\partial^2 \phi}{\partial \alpha_j \partial \alpha_m} \frac{\partial \bar{\alpha}_m}{\partial x_k} \frac{\partial \bar{\alpha}_j}{\partial x_k} + \sum_{j=1}^{n} \frac{\partial \phi}{\partial \alpha_j} \frac{\partial^2 \bar{\alpha}_j}{\partial {x_k}^2}. \tag{4.16}$$

This expression can be simplified. First, note that

$$\frac{\partial \phi}{\partial \alpha_j} \frac{\partial^2 \bar{\alpha}_j}{\partial {x_k}^2} = 0,$$

simply because $\bar{\alpha}(t,x)$ is linear in x. To deal with the other terms, note that, as $\bar{\alpha}(t,x)$ minimizes $\phi(t,x,\alpha)$, we have

$$\nabla_\alpha \phi(t,x,\alpha)\Big|_{\alpha=\bar{\alpha}(t,x)} = 0, \quad D^2_\alpha \phi(t,x,\alpha)\Big|_{\alpha=\bar{\alpha}(t,x)} > 0, \tag{4.17}$$

with the second inequality holding in the sense positive-definite matrices. Differentiating the first identity above in x_k we get

$$\frac{\partial^2 \phi(t,x,\bar{\alpha})}{\partial x_k \partial \alpha_i} + \sum_{m=1}^{n} \frac{\partial^2 \phi(t,x,\bar{\alpha})}{\partial \alpha_i \partial \alpha_m} \frac{\partial \bar{\alpha}_m}{\partial x_k} = 0, \text{ for all } 1 \leq i,k \leq n. \tag{4.18}$$

Using this in (4.16) gives

$$\Delta v = \sum_{k=1}^{n} \frac{\partial^2 \phi}{\partial {x_k}^2} - \sum_{k,j,m=1}^{n} \frac{\partial^2 \phi}{\partial \alpha_j \partial \alpha_m} \frac{\partial \bar{\alpha}_m}{\partial x_k} \frac{\partial \bar{\alpha}_j}{\partial x_k}, \tag{4.19}$$

so that

$$v_t - \Delta v = \sum_{k,j,m=1}^{n} \frac{\partial^2 \phi}{\partial \alpha_j \partial \alpha_m} \frac{\partial \bar{\alpha}_m}{\partial x_k} \frac{\partial \bar{\alpha}_j}{\partial x_k} > 0, \tag{4.20}$$

because of the inequality in (4.17). Thus, $v(t,x)$ is a strict-subsolution, and the "extent of its strictness" is given explicitly by the right side in (4.20). Using the explicit expression (4.15) for $\bar{\alpha}(t,x)$ we may simplify (4.20) to

$$v_t - \Delta v = \frac{1}{4t^2}\Delta_\alpha \phi(t,x,\bar{\alpha}) = \frac{1}{4t^2}[(x+2\bar{\alpha}t)^2 + 2nt]\phi(t,x,\bar{\alpha}) = \frac{n}{2t}\phi(t,x,\bar{\alpha}). \tag{4.21}$$

Therefore, if we define

$$\bar{v}(t,x) = a(t)\phi(t,x,\bar{\alpha}(t,x)),$$

the function $\bar{v}(t,x)$ is still a super-solution, as long as

$$\dot{a}(t) + \frac{n}{2t} \geq 0. \tag{4.22}$$

Therefore, we may take $a(t) = t^{-n/2}$, obtaining a super-solution

$$\bar{v}(t,x) = \frac{e^{-|x|^2/(4t)}}{t^{n/2}}, \tag{4.23}$$

which happens to be also a solution of the heat equation.

The periodic case: trying to use super-solutions

Let us generalize this approach to the problem

$$u_t = \nabla \cdot (a(x)\nabla\phi), \tag{4.24}$$

with a periodic uniformly elliptic matrix $a(x)$. We look for supersolutions in the form

$$\phi(t,x;\alpha) = \psi(x;\alpha)e^{\alpha\cdot x+\mu(\alpha)t}, \tag{4.25}$$

with a periodic function $\psi_\alpha(x)$. The rate $\mu(\alpha)$ comes from the eigenvalue problem for $\psi_\alpha(x)$. We insert the ansatz (4.25) into (4.24) and get

$$e^{-\alpha\cdot x}\nabla\cdot(a(x)e^{\alpha\cdot x}(\nabla\psi(x;\alpha) + \alpha\psi(x;\alpha))) = \mu(\alpha)\psi(x;\alpha), \tag{4.26}$$

or, equivalently:

$$\nabla\cdot(a(x)\nabla\psi(x;\alpha)) + \alpha\cdot(a(x)\nabla\psi(x;\alpha)) + \nabla\cdot(a(x)\alpha\psi(x;\alpha)) + (a(x)\alpha\cdot\alpha)\psi(x;\alpha) = \mu(\alpha)\psi(x;\alpha). \tag{4.27}$$

We obtain a super-solution by setting

$$\bar{v}(t,x) = \inf_{\alpha\in\mathbb{R}^n}\phi(t,x;\alpha) = \inf_\alpha \psi(t,x;\alpha)e^{\alpha\cdot x+\mu(\alpha)t}. \tag{4.28}$$

The general case: trying to use the subsolutions

For the general problem we may consider a family of solutions with exponential initial data: for each $\alpha \in \mathbb{R}^n$ let $\phi(t,x;\alpha)$ be the solution of the Cauchy problem

$$\frac{\partial \phi(t,x;\alpha)}{\partial t} = \nabla \cdot (a(x)\nabla \phi(t,x;\alpha)), \quad t>0, \quad x \in \mathbb{R}^n, \tag{4.29}$$
$$\phi(0,x;\alpha) = e^{\alpha \cdot x}.$$

Let us get an upper bound for $\phi(t,x,\alpha)$: set

$$w(t,x;\alpha) = e^{-\alpha\cdot x}\phi(t,x,\alpha).$$

The function $w(t,x;\alpha)$ satisfies

$$\frac{\partial w}{\partial t} = e^{-\alpha\cdot x}\nabla\cdot\Big(a(x)\nabla(e^{\alpha\cdot x}w)\Big) = e^{-\alpha\cdot x}\nabla\cdot\Big(a(x)(e^{\alpha\cdot x}(\nabla w + \alpha w))\Big) \tag{4.30}$$
$$= \frac{\partial a_{kj}}{\partial x_k}\Big(\frac{\partial w}{\partial x_j} + \alpha_j w\Big) + a_{kj}\alpha_k(2\frac{\partial w}{\partial x_j} + \alpha_j w\Big) + a_{kj}\frac{\partial^2 w}{\partial x_j \partial x_k} \tag{4.31}$$
$$w(0,x;\alpha) = 1.$$

The functions

$$\phi_i(t,x;\alpha) = \frac{\partial \phi(t,x;\alpha)}{\partial \alpha_i}, \quad \phi_{ij}(t,x;\alpha) = \frac{\partial^2 \phi(t,x;\alpha)}{\partial \alpha_i \partial \alpha_j}$$

satisfy the Cauchy problems

$$\frac{\partial \phi_i(t,x;\alpha)}{\partial t} = \nabla \cdot (a(x)\nabla \phi_i(t,x;\alpha)), \quad t>0, \quad x \in \mathbb{R}^n, \tag{4.32}$$
$$\phi_i(0,x;\alpha) = x_i e^{\alpha\cdot x},$$

and

$$\frac{\partial \phi_{ij}(t,x;\alpha)}{\partial t} = \nabla \cdot (a(x)\nabla \phi_{ij}(t,x;\alpha)), \quad t>0, \quad x \in \mathbb{R}^n, \tag{4.33}$$
$$\phi_{ij}(0,x;\alpha) = x_i x_j e^{\alpha\cdot x}.$$

The function $\phi(t,x;\alpha)$ is convex in α: for any vector $\xi \in \mathbb{R}^n$ the function

$$\psi(t,x,\alpha) = \sum_{i,j=1}^{n} \phi_{ij}(t,x;\alpha)\xi_i\xi_j$$

satisfies

$$\frac{\partial \psi(t,x;\alpha)}{\partial t} = \nabla \cdot (a(x)\nabla \phi(t,x;\alpha)), \quad t>0, \quad x \in \mathbb{R}^n, \tag{4.34}$$
$$\psi(0,x;\alpha) = (x\cdot\xi)^2 e^{\alpha\cdot x}.$$

Thus, $\psi(t,x;\alpha) > 0$, which means that the function $\phi(t,x;\alpha)$ is convex in α for each $t>0$ and $x \in \mathbb{R}^n$.

Exercise 4.3 Show that $\phi(t,x;\alpha)$ has a unique minimum, as a function of $\alpha \in \mathbb{R}^n$, for each $t > 0$ and $x \in \mathbb{R}^n$.

Let us denote this minimum by $\bar\alpha(t,x)$. The function

$$\bar v(t,x) = \phi(t,x;\bar\alpha(t,x)) = \inf_{\alpha\in\mathbb{R}^n} \phi(t,x;\alpha) \tag{4.35}$$

is a super-solution:

$$\frac{\partial \bar v}{\partial t} \geq \nabla\cdot(a(x)\nabla \bar v), \quad t>0, \quad x\in\mathbb{R}^n, \tag{4.36}$$

as it is an infimum of a family of super-solutions. Let us check how much room this gives us – how strict of a super-solution $\bar v(t,x)$ is. We compute:

$$\frac{\partial \bar v}{\partial t} = \frac{\partial \phi(t,x;\bar\alpha)}{\partial t} + \frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial t} = \frac{\partial \phi(t,x;\bar\alpha)}{\partial t},$$

as

$$\frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j} = 0, \quad 1\leq j\leq n. \tag{4.37}$$

Differentiating the above in x_k gives

$$\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial x_k\partial \alpha_j} + \frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial \alpha_j\partial \alpha_m}\frac{\partial \bar\alpha_m}{\partial x_k} = 0, \quad 1\leq j,k\leq n. \tag{4.38}$$

We also have

$$a_{mk}(x)\frac{\partial \bar v}{\partial x_k} = a_{mk}(x)\frac{\partial \phi(t,x;\bar\alpha)}{\partial x_k} + a_{mk}(x)\frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial x_k},$$

so that

$$\begin{aligned}
\nabla\cdot(a(x)\nabla\bar v) &= \frac{\partial}{\partial x_m}\Big(a_{mk}(x)\frac{\partial \bar v}{\partial x_k}\Big) = \frac{\partial}{\partial x_m}\Big(a_{mk}(x)\frac{\partial \phi(t,x;\bar\alpha)}{\partial x_k} + a_{mk}(x)\frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial x_k}\Big)\\
&= \nabla\cdot(a(x)\nabla\phi)(t,x;\bar\alpha) + a_{mk}(x)\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial x_k\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial x_m} + \frac{\partial a_{mk}(x)}{\partial x_m}\frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial x_k}\\
&+ a_{mk}(x)\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial x_m\partial \alpha_j}\frac{\partial \bar\alpha_j}{\partial x_k} + a_{mk}(x)\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial \alpha_j\partial \alpha_l}\frac{\partial \bar\alpha_l}{\partial x_m}\frac{\partial \bar\alpha_j}{\partial x_k} + a_{mk}(x)\frac{\partial \phi(t,x;\bar\alpha)}{\partial \alpha_j}\frac{\partial^2 \bar\alpha_j}{\partial x_k\partial x_m}.
\end{aligned}$$

Using (4.37), (4.38) and the symmetry of a_{mk}, this simplifies to

$$\nabla\cdot(a(x)\nabla\bar v) = \nabla\cdot(a(x)\nabla\phi)(t,x;\bar\alpha) - a_{mk}(x)\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial \alpha_j\partial \alpha_l}\frac{\partial \bar\alpha_l}{\partial x_k}\frac{\partial \bar\alpha_j}{\partial x_m}. \tag{4.39}$$

Therefore, we have

$$\bar v_t - \nabla\cdot(a(x)\nabla\bar v) = a_{mk}(x)\frac{\partial^2 \phi(t,x;\bar\alpha)}{\partial \alpha_j\partial \alpha_l}\frac{\partial \bar\alpha_l}{\partial x_k}\frac{\partial \bar\alpha_j}{\partial x_m} > 0. \tag{4.40}$$

Exercise 4.4 Show that $\phi(t,x;\alpha)$ satisfies an estimate

$$\phi(t,x;\alpha)\le e^{\alpha\cdot x+C\alpha^2 t}, \tag{4.41}$$

with a constant $C>0$ that depends only on Λ.

With this result in hand, we would already know that

$$\bar v(t,x)\le\phi(t,x;\alpha_1),$$

where

$$\alpha_1=-\frac{x}{2Ct},$$

thus we have a Gaussian super-solution:

$$\bar v(t,x)\le e^{\alpha_1\cdot x+C\alpha_1^2 t}=e^{-|x|^2/(4Ct)}. \tag{4.42}$$

In order to improve, so that we could include the decay $t^{-n/2}$, we need to find a good function $\beta(t)$ so that

$$w(t,x)=\beta(t)\bar v(t,x)$$

is still a super-solution. That is, we need to ensure that

$$\dot\beta(t)\phi(t,x;\bar\alpha)+\beta(t)a_{mk}(x)\frac{\partial^2\phi(t,x;\bar\alpha)}{\partial\alpha_j\partial\alpha_l}\frac{\partial\bar\alpha_l}{\partial x_k}\frac{\partial\bar\alpha_j}{\partial x_m}\ge 0. \tag{4.43}$$

This seems tough because $a_{mk}(x)$ may be very small.

The general case: the Nash approach

The general strategy of the proof is similar to that of the uniform bound without the Gaussian factor in Theorem 2.2 with several important modifications. An important role will be played again by the exponentials: rather than consider only the function $\phi(t,x)$ we will use the exponential moments of $\phi(t,x)$. Fix $\alpha\in\mathbb{R}^n$ and consider the function

$$\phi_\alpha(t,x)=e^{-\alpha\cdot x}\psi_\alpha(t,x).$$

Here $\psi_\alpha(t,x)$ is the solution of the initial value problem with exponentially weighted initial data:

$$\begin{aligned}&\frac{\partial\psi_\alpha}{\partial t}=\nabla\cdot(a(x)\nabla\psi_\alpha),\quad t>0,\quad x\in\mathbb{R}^n,\\&\psi_\alpha(0,x)=g(x)e^{\alpha\cdot x}.\end{aligned} \tag{4.44}$$

The key point is that L^∞ bounds for ϕ_α will give us decay estimates on the function $\phi(t,x)$ itself with a judiciously chosen α. We will show the following proposition.

Proposition 4.5 *There exists a constant $C>0$ that depends only on the dimension n so that*

$$\|\phi_\alpha(t)\|_{L^\infty}\le\frac{C}{(\lambda t)^{n/2}}e^{2\alpha^2 t\Lambda}\|g\|_{L^1}. \tag{4.45}$$

Exercise 4.6 Verify by a direct computation that the conclusion of Proposition 4.5 holds for the standard heat equation.

Let us explain how Theorem 4.2 follows from Proposition 4.5. Consider the operator P_t^α that maps $g(x)$ to $\phi_\alpha(t,x)$. It is given explicitly by

$$P_t^\alpha g(x) = e^{-\alpha\cdot x}\int \Gamma(t,x,y)g(y)e^{\alpha\cdot y}dy = \int K(t,x,y)g(y)dy, \tag{4.46}$$

with the integral kernel

$$K(t,x,y) = \Gamma(t,x,y)e^{\alpha\cdot(y-x)}. \tag{4.47}$$

Proposition 4.5 says that the operator P_t^α obeys the bound

$$\|P_t^\alpha g\|_{L^\infty} \le \frac{C}{(\lambda t)^{n/2}}e^{2\alpha^2 t\Lambda}\|g\|_{L^1}. \tag{4.48}$$

However, we know that an integral operator of the form

$$[\mathcal{I}g](x) = \int M(x,y)g(y)dy, \tag{4.49}$$

considered as a mapping $L^1 \to L^\infty$, has the norm

$$\|\mathcal{I}\|_{L^1\to L^\infty} = \sup_{x,y}|M(x,y)|. \tag{4.50}$$

Therefore, the (non-negative) integral kernel $K(t,x,y)$ of the operator P_t^α satisfies the L^∞-bound

$$K(t,x,y) \le \frac{C}{(\lambda t)^{n/2}}e^{2\alpha^2 t\Lambda}, \tag{4.51}$$

and Green's function itself satisfies

$$\Gamma(t,x,y) \le \frac{C}{(\lambda t)^{n/2}}e^{2\alpha^2 t\Lambda}e^{\alpha\cdot(x-y)}. \tag{4.52}$$

As this estimate holds for all $\alpha \in \mathbb{R}^n$, given any $x, y \in \mathbb{R}^n$, we can take, in particular,

$$\alpha = \frac{1}{4t\Lambda}(y-x) \tag{4.53}$$

and get the desired Gaussian upper bound

$$\Gamma(t,x,y) \le \frac{C}{(\lambda t)^{n/2}}e^{-|x-y|^2/(8\Lambda t)}. \tag{4.54}$$

Thus, the Gaussian bound on the function $\Gamma(t,x,y)$ is a consequence of the L^∞ bound (4.45) on the functions ϕ_α. Our task, therefore, is to prove the $L^1 \to L^\infty$ decay estimate (4.48) for the operator P_t^α. In the proof of Theorem 2.2 we have obtained such bound for the solution operator S_t for the original Cauchy problem (4.2):

$$S_t g(x) = \int \Gamma(t,x,y)g(y)dy. \tag{4.55}$$

This was done by first establishing the $L^1 \to L^2$ bound on S_t using the Nash inequality, and then using the fact that S_t is self-adjoint, and duality to deduce the $L^2 \to L^\infty$ bound on S_t. The final step was to use the semi-group property

$$S_t = S_{t/2} \circ S_{t/2},$$

that gives the $L^1 \to L^\infty$ estimate for S_t as the product of $L^1 \to L^2$ and $L^2 \to L^\infty$ bounds for $S_{t/2}$. Here, the strategy is reversed: we will first show the $L^2 \to L^\infty$ bound and then use duality and semi-group property of P_t^α to obtain the $L^1 \to L^\infty$ bound for P_t^α.

The operators P_t^α share a lot of common properties with the solution operator S_t. They are not symmetric like S_t but the adjoint operator $P_t^{\alpha *}$ is obtained by simply switching the sign of α:

$$P_t^{\alpha *} = P_t^{-\alpha}. \tag{4.56}$$

Indeed, recall that, as we have seen in the proof of Theorem 2.1, the operator S_t is symmetric, meaning that

$$\Gamma(t, x, y) = \Gamma(t, y, x). \tag{4.57}$$

Therefore, the operator $P_t^{\alpha *}$ has the form

$$\begin{aligned} P_t^{\alpha *} f(x) &= \int K(t, y, x) f(y) dy = e^{\alpha \cdot x} \int \Gamma(t, y, x) f(y) e^{-\alpha \cdot y} dy \\ &= e^{\alpha \cdot x} \int \Gamma(t, x, y) f(y) e^{-\alpha \cdot y} dy. \end{aligned} \tag{4.58}$$

In other words, (4.56) holds.

Continuing our analogy with S_t, the operators P_t^α form a semi-group:

$$P_t^\alpha = P_{t-s}^\alpha \circ P_s^\alpha, \quad 0 \le s \le t. \tag{4.59}$$

In order to verify (4.59) we will use the semigroup property of Green's function:

$$\Gamma(t, x, z) = \int \Gamma(t-s, x, y) \Gamma(s, y, z) dy. \tag{4.60}$$

We deduce from this property that

$$\begin{aligned} (P_{t-s}^\alpha \circ P_s^\alpha) g(x) &= \int e^{-\alpha \cdot x} \Gamma(t-s, x, y) [P_s^\alpha g](y) e^{\alpha \cdot y} dy \\ &= \int e^{-\alpha \cdot (x-y)} \Gamma(t-s, x, y) \Gamma(s, y, z) e^{-\alpha \cdot (y-z)} g(z) dz dy \\ &= \int e^{-\alpha \cdot (x-z)} \left(\int \Gamma(t-s, x, y) \Gamma(s, y, z) dy \right) g(z) dz \\ &= \int e^{-\alpha \cdot (x-z)} \Gamma(t, x, y) g(z) dz = P_t g(x), \end{aligned} \tag{4.61}$$

which is (4.59).

As we have mentioned, we will prove directly the $L^2 \to L^\infty$ bound rather than the $L^1 \to L^2$ bound as we did in the proof of Theorem 2.2.

Lemma 4.7 *There exists a constant $C > 0$ that depends only on the dimension n so that*

$$\|\phi_\alpha(t)\|_{L^\infty} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda} \|g\|_{L^2}, \tag{4.62}$$

that is,

$$\|P_t^\alpha\|_{L^2 \to L^\infty} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda}. \tag{4.63}$$

Here is how the conclusion of Proposition 4.5 follows from Lemma 4.7. As $P_t^{\alpha *} = P_t^{-\alpha}$, the adjoint operator also satisfies the $L^2 \to L^\infty$ estimate (4.63) (with α replaced by $(-\alpha)$ which makes no difference):

$$\|P_t^{\alpha *} g\|_{L^\infty} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda} \|g\|_{L^2}. \tag{4.64}$$

Therefore, for any function $g \in L^1$ and $f \in L^2$ we have

$$\int (P_t^\alpha g(x)) f(x) dx = \int g(x) P_t^{\alpha *} f(x) dx \le \|g\|_{L^1} \|P_t^{\alpha *} f\|_{L^\infty} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda} \|g\|_{L^1} \|f\|_{L^2}, \tag{4.65}$$

hence

$$\|P_t^\alpha g\|_{L^2} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda} \|g\|_{L^1}, \tag{4.66}$$

or

$$\|P_t^\alpha\|_{L^1 \to L^2} \le \frac{C}{(\lambda t)^{n/4}} e^{2\alpha^2 t\Lambda}. \tag{4.67}$$

As the operators P_t^α form a semi-group, we have

$$P_t^\alpha = P_{t/2}^\alpha \circ P_{t/2}^\alpha. \tag{4.68}$$

Hence, as in the proof of Theorem 2.1 we get the bound

$$\|P_t^\alpha\|_{L^1 \to L^\infty} \le \|P_{t/2}\|_{L^1 \to L^2} \|P_{t/2}\|_{L^2 \to L^\infty} \le \frac{C}{(\lambda t)^{n/2}} e^{2\alpha^2 t\Lambda}, \tag{4.69}$$

which proves Proposition 4.5.

The proof of Lemma 4.7

The most technical part of the proof of the upper bound in Theorem 4.2 is the $L^2 \to L^\infty$ bound for the operators P_t^α in Lemma 4.7. We will get a family of differential inequalities for the norms

$$M_p(t) = \|\phi_\alpha(t)\|_{L^{2p}}, \quad 1 \le p < +\infty, \tag{4.70}$$

of the form

$$\frac{dM_p}{dt} \le -\frac{C\lambda}{2p} \frac{M_p^{1+4p/n}}{M_{p/2}^{4p/n}} + \alpha^2 p \Lambda M_p, \tag{4.71}$$

together with the "boundary condition" at $p = 1$:

$$M_1(t) = \|\phi_\alpha(t)\|_{L^2} \le e^{\alpha^2 \Lambda t} \|g\|_{L^2}, \quad t \ge 0. \tag{4.72}$$

The second step will be to will use an ODE argument to get bounds on $M_p(t)$ in terms of $M_1(t)$ and finish the proof.

Let us show how (4.71) is obtained. The function ϕ_α satisfies the Cauchy problem

$$\begin{aligned}&\frac{\partial \phi_\alpha}{\partial t} = e^{-\alpha\cdot x}\nabla\cdot(a(x)\nabla(e^{\alpha\cdot x}\phi_\alpha)),\\&\phi_\alpha(0,x) = g(x).\end{aligned} \tag{4.73}$$

Multiplying this equation by ϕ_α^{2p-1} gives

$$\frac{1}{2p}\frac{d}{dt}\int |\phi_\alpha(t,x)|^{2p}dx = \int e^{-\alpha\cdot x}\phi_\alpha^{2p-1}\nabla\cdot(a(x)\nabla(e^{\alpha\cdot x}\phi_\alpha))dx. \tag{4.74}$$

Let us now rewrite the dissipation term in the right side as follows:

$$\begin{aligned}&D := \int e^{-\alpha\cdot x}\phi_\alpha^{2p-1}\nabla\cdot(a(x)\nabla(e^{\alpha\cdot x}\phi_\alpha))dx = \int e^{-\alpha\cdot x}\phi_\alpha^{2p-1}(a(x)\alpha\cdot\nabla(e^{\alpha\cdot x}\phi_\alpha))dx\\&-(2p-1)\int e^{-\alpha\cdot x}\phi_\alpha^{2p-2}(a(x)\nabla\phi_\alpha\cdot\nabla(e^{\alpha\cdot x}\phi_\alpha))dx = \int(a(x)\alpha\cdot\alpha)\phi_\alpha^{2p}dx\\&-(2p-2)\int\phi_\alpha^{2p-1}(a(x)\alpha\cdot\nabla\phi_\alpha)dx-(2p-1)\int\phi_\alpha^{2p-2}(a(x)\nabla\phi_\alpha\cdot\nabla\phi_\alpha)dx.\end{aligned} \tag{4.75}$$

We will use Young's inequality for the middle term in the last identity above:

$$|(a(x)\alpha\cdot\nabla\phi)| \le \frac{|\phi_\alpha|}{2}(a(x)\alpha\cdot\alpha)+\frac{1}{2|\phi_\alpha|}(a(x)\nabla\phi_\alpha\cdot\nabla\phi_\alpha). \tag{4.76}$$

This gives

$$\begin{aligned}D &\le p\int(a(x)\alpha\cdot\alpha)\phi_\alpha^{2p}dx - p\int\phi_\alpha^{2p-2}(a(x)\nabla\phi_\alpha\cdot\nabla\phi_\alpha)dx\\&\le p\Lambda|\alpha|^2\int|\phi_\alpha|^{2p}dx-\frac{\lambda}{p}\int|\nabla(\phi_\alpha^p)|^2dx.\end{aligned} \tag{4.77}$$

We will now use the Nash inequality for the function $\phi_\alpha^p(x)$:

$$\int|\nabla(\phi_\alpha)^p|^2dx \ge C_n\left(\int|\phi_\alpha|^{2p}dx\right)^{1+2/n}\left(\int|\phi_\alpha|^pdx\right)^{-4/n}. \tag{4.78}$$

Using this in (4.77) leads to the dissipation bound:

$$D \le p\Lambda|\alpha|^2\int|\phi_\alpha|^{2p}dx - \frac{C\lambda}{p}\left(\int|\phi_\alpha|^{2p}dx\right)^{1+2/n}\left(\int|\phi_\alpha|^pdx\right)^{-4/n}. \tag{4.79}$$

Going back to identity (4.74) and writing it in terms of the moments M_p gives

$$\frac{1}{2p}\frac{d}{dt}(M_p^{2p}) \le p\Lambda|\alpha|^2M_p^{2p}-\frac{C\lambda}{p}\frac{M_p^{2p(1+2/n)}}{M_{p/2}^{p(4/n)}}, \tag{4.80}$$

or

$$\frac{dM_p}{dt} \le p\Lambda|\alpha|^2 M_p - \frac{C\lambda}{p}\frac{M_p^{1+4p/n}}{M_{p/2}^{4p/n}}, \tag{4.81}$$

which is the differential inequality we were looking for. It is not closed as the right side involves not only M_p but also on $M_{p/2}$. The constant C here depends only on dimension n. One consequence of (4.81) is an exponentially growing in time bound

$$M_p(t) \le e^{p\Lambda|\alpha|^2 t}\|g\|_{L^{2p}}, \tag{4.82}$$

which we will use later with $p = 1$.

We now use the differential inequalities (4.81) to bound the moments $M_p(t)$ in terms of $M_{p/2}(t)$. Let us first take out the exponential factor: set

$$G_p(t) = M_p(t)e^{-p\Lambda|\alpha|^2 t}. \tag{4.83}$$

Then (4.81) implies that G_p satisfies

$$\frac{dG_p}{dt} \le -\frac{C\lambda}{p}\frac{G_p^{1+4p/n}}{M_{p/2}^{4p/n}}e^{-p\Lambda|\alpha|^2 t}e^{p\Lambda|\alpha|^2(1+4p/n)t} = -\frac{C\lambda}{p}\frac{G_p^{1+4p/n}}{M_{p/2}^{4p/n}}e^{4p^2\Lambda|\alpha|^2 t/n}, \tag{4.84}$$

hence

$$\frac{n}{4p}\frac{d}{dt}(G_p^{-4p/n}) \ge \frac{C\lambda}{p}\frac{1}{M_{p/2}^{4p/n}}e^{4p^2\Lambda|\alpha|^2 t/n}. \tag{4.85}$$

It would be convenient to proceed if we knew that $M_{p/2}(t)$ were increasing in time. Let us see what we may expect in this regard: consider the standard heat kernel (corresponding to $g(x) = \delta(x)$)

$$G_0(t,x) = \frac{e^{-x^2/(4t)}}{(4\pi t)^{n/2}},$$

and compute

$$\begin{aligned} |M_p^{(0)}(t)|^{2p} &= \int e^{-2p\alpha\cdot x}G_0^{2p}(t,x)dx = \int e^{-2p\alpha\cdot x}\frac{e^{-px^2/(2t)}}{(4\pi t)^{pn}}dx \\ &= e^{2p\alpha^2 t}\int \exp\{-\frac{p}{2}(\frac{x}{\sqrt{t}} + 2\alpha\sqrt{t})^2\}\frac{dx}{(4\pi t)^{pn}} = C_p\frac{e^{2p\alpha^2 t}}{t^{pn-n/2}}. \end{aligned} \tag{4.86}$$

Observe that while $M_p^{(0)}(t)$ is not monotonic in time, it becomes monotonic if we multiply it by $t^{p(n-1)/(4p)}$. This motivates the following: set

$$\bar{M}_p(t) = \max_{0\le s\le t}\left[s^{(p-1)n/(4p)}M_p(s)\right], \tag{4.87}$$

so that $\bar{M}_p(t)$ is non-decreasing in time, and

$$\frac{1}{M_{p/2}(t)^{4p/n}} \ge \frac{t^{p-2}}{\bar{M}_{p/2}(t)^{4p/n}}. \tag{4.88}$$

Using this in inequality (4.85) gives, with another constant C that depends only on dimension n:

$$\frac{1}{G_p(t)^{4p/n}} \geq C\lambda \int_0^t \frac{s^{p-2}}{\bar{M}_{p/2}(s)^{4p/n}} e^{4p^2\Lambda|\alpha|^2 s/n} ds. \tag{4.89}$$

As the function $\bar{M}_p(s)$ is non-decreasing in s we deduce that

$$\frac{1}{G_p(t)^{4p/n}} \geq \frac{C\lambda}{\bar{M}_{p/2}(t)^{4p/n}} \int_0^t s^{p-2} e^{4p^2\Lambda|\alpha|^2 s/n} ds. \tag{4.90}$$

The integral in the right side can be evaluated explicitly for integer p but we will only estimate it:

$$\begin{aligned}
&\int_0^t s^{p-2} e^{4p^2\Lambda|\alpha|^2 s/n} ds = \left(\frac{nt}{4p^2\Lambda|\alpha|^2}\right)^{p-1} \int_0^{4p^2\Lambda|\alpha|^2/n} s^{p-2} e^{ts} ds \\
&\geq \left(\frac{nt}{4p^2\Lambda|\alpha|^2}\right)^{p-1} \int_{(1-1/p^2)4p^2\Lambda|\alpha|^2/n}^{4p^2\Lambda|\alpha|^2/n} s^{p-2} e^{ts} ds \\
&\geq \left(\frac{nt}{4p^2\Lambda|\alpha|^2}\right)^{p-1} e^{(1-1/p^2)4p^2\Lambda|\alpha|^2 t/n} \int_{(1-1/p^2)4p^2\Lambda|\alpha|^2/n}^{4p^2\Lambda|\alpha|^2/n} s^{p-2} ds \\
&= \frac{t^{p-1}}{p-1} e^{(1-1/p^2)4p^2\Lambda|\alpha|^2 t/n} (1 - (1 - \frac{1}{p^2})^{p-1}).
\end{aligned} \tag{4.91}$$

This estimate can be improved if we replace the lower limit of integration in (4.91) not by $(1-1/p^2)$ times the upper limit but by $(1-\delta/p^2)$ times the upper limit with an appropriately chosen $\delta > 0$. This improves the final constant in the estimates and gives the more precise version (4.10) of the Gaussian upper bound[3] but we will not pursue this avenue here as our hands are already full with technicalities. As a slight simplification, the last factor above satisfies

$$1 - (1 - \frac{1}{p^2})^{p-1} \leq \frac{K}{p}, \quad \text{for all } p \geq 1, \tag{4.92}$$

with a universal constant K. Going back to (4.90) we obtain

$$\frac{1}{G_p(t)^{4p/n}} \geq \frac{\lambda}{\bar{M}_{p/2}(t)^{4p/n}} \frac{Kt^{p-1}}{p^2} e^{(1-1/p^2)4p^2\Lambda|\alpha|^2 t/n}. \tag{4.93}$$

We re-write this inequality in terms of $M_p(t)$:

$$\begin{aligned}
M_p(t) &\leq C^{n/(4p)} \bar{M}_{p/2}(t) \left(\frac{p^2}{\lambda t^{p-1}}\right)^{n/(4p)} e^{p\Lambda\alpha^2 t} e^{-(1-1/p^2)4p^2\Lambda|\alpha|^2 t/(4p)} \\
&= C^{n/(4p)} \bar{M}_{p/2}(t) \left(\frac{p^2}{\lambda t^{p-1}}\right)^{n/(4p)} e^{\Lambda|\alpha|^2 t/p}.
\end{aligned} \tag{4.94}$$

[3]In particular, the constant 8 in (4.10) can be turned into $4 + \delta$ for any $\delta > 0$ which is nearly optimal.

Multiplying both sides by $t^{(p-1)n/(4p)}$ gives

$$M_p(t)t^{(p-1)n/(4p)} \le C^{n/(4p)}\bar{M}_{p/2}(t)\left(\frac{p^2}{\lambda}\right)^{n/(4p)} e^{\Lambda|\alpha|^2 t/p}. \tag{4.95}$$

Therefore, for any $0 \le s \le t$ we have

$$M_p(s)s^{(p-1)n/(4p)} \le C^{n/(4p)}\bar{M}_{p/2}(s)\left(\frac{p^2}{\lambda}\right)^{n/(4p)} e^{\Lambda|\alpha|^2 s/p} \le C^{n/(4p)}\bar{M}_{p/2}(t)\left(\frac{p^2}{\lambda}\right)^{n/(4p)} e^{\Lambda|\alpha|^2 t/p}. \tag{4.96}$$

Taking the supremum over all $0 \le s \le t$ we arrive at

$$\bar{M}_p(t) \le C^{n/(4p)}\bar{M}_{p/2}(t)\left(\frac{p^2}{\lambda}\right)^{n/(4p)} e^{\Lambda|\alpha|^2 t/p}. \tag{4.97}$$

Taking $p = 2^k$ we deduce that for all $k \ge 1$ we have

$$\bar{M}_{2^k}(t) \le \frac{C}{\lambda^{n/4}}\bar{M}_1(t)e^{\Lambda|\alpha|^2 t} = CM_1(t)e^{\Lambda|\alpha|^2 t}, \tag{4.98}$$

since

$$\prod_{k=1}^{\infty}(2^k)^{1/2^k} = \exp\left[(\log 2)\sum_{k=1}^{\infty}\frac{k}{2^k}\right] < +\infty. \tag{4.99}$$

Now, we are almost done: (4.98) means that

$$\|\phi_\alpha\|_{L^{2^k}} \le \frac{Ce^{\alpha^2\Lambda t}e^{\Lambda|\alpha|^2 t}}{\lambda^{n/4}t^{(2^k-1)n/(4\cdot 2^k)}}\|g\|_{L^2}, \tag{4.100}$$

for all $k \ge 1$. Passing to the limit $k \to +\infty$, it follows that

$$\|\phi_\alpha\|_{L^\infty} \le \frac{Ce^{2\alpha^2\Lambda t}}{(\lambda t)^{n/4}}\|g\|_{L^2}, \tag{4.101}$$

with a constant $C > 0$ that depends only on the dimension n, as we have claimed. This competes the proof of Lemma 4.5 and Theorem 2.2.

4.2 The proof of the lower bound

In this section we prove the lower Gaussian bound for Green's function.

Theorem 4.8 *There exists a constant $C > 0$ that depends only on the ellipticity constants λ, Λ and dimension n so that*

$$\Gamma(t,x,y) \ge \frac{1}{Ct^{n/2}}e^{-C(x-y)^2/t}. \tag{4.102}$$

We will not try to track the dependence of the constant on λ and Λ as we did in the proof of Theorem 4.2, though that can also be done albeit at the expense of rather long expressions. Thus, throughout this proof we will denote by C various constants that depend on λ, Λ and dimension n. The main ingredient in the proof of Theorem 4.8 is the uniform lower bound on $\Gamma(t,x,y)$ on the set

$$\{|x-y| \le \sqrt{t}\}.$$

Theorem 4.9 *There exists a constant C that depends only on the dimension n and ellipticity constants λ and Λ so that*

$$\Gamma(t,x,y) \geq \frac{1}{Ct^{n/2}}, \tag{4.103}$$

for all x, y such that $|x-y| \leq \sqrt{t}$.

The uniform lower bound in Theorem 4.9 is actually sufficient to produce the Gaussian decay in Theorem 4.8, and this is what we show first. Without loss of generality we may assume that $y = 0$. We need to show that

$$\Gamma(t,x,0) \geq \frac{1}{Ct^{n/2}} e^{-Cx^2/t} \tag{4.104}$$

Theorem 4.9 implies that we only need to consider $|x| \geq \sqrt{t}$. Let $x \in \mathbb{R}^n$, $t > 0$ and k be the smallest integer larger than $4|x|^2/t$:

$$k - 1 \leq \frac{4|x|^2}{t} < k. \tag{4.105}$$

Consider a sequence of balls

$$B_j = B\Big(\frac{jx}{k}, \frac{\sqrt{t}}{2\sqrt{k}}\Big) = \Big\{y:\ \big|y - \frac{j}{k}x\big| \leq \frac{\sqrt{t}}{2\sqrt{k}}\Big\}, \quad j = 1, \dots, k-1. \tag{4.106}$$

As

$$\frac{|x|}{k} < \frac{\sqrt{t}}{2\sqrt{k}}, \tag{4.107}$$

each pair of consecutive balls B_j and B_{j+1} overlap, and, moreover, the center of B_{j+1} lies inside B_j and vice versa. In particular, the origin $y = 0$ lies inside B_1, and the point x lies inside B_{k-1}. Then, given any collection of points $\xi_l \in B_l$ they satisfy the following properties:

$$|\xi_1| \leq \frac{\sqrt{t}}{\sqrt{k}}, \quad |x - \xi_{k-1}| \leq \frac{\sqrt{t}}{\sqrt{k}}, \tag{4.108}$$

and

$$|\xi_{l+1} - \xi_l| \leq \frac{\sqrt{t}}{\sqrt{k}}, \text{ for all } 1 \leq k \leq k-2. \tag{4.109}$$

The semigroup property of Green's function $\Gamma(t,x,y)$ implies that

$$\begin{aligned}\Gamma(t,x,0) &= \int_{\mathbb{R}^n} \dots \int_{\mathbb{R}^n} \Gamma\big(\frac{t}{k}, x, \xi_{k-1}\big)\Gamma\big(\frac{t}{k}, \xi_{k-1}, \xi_{k-2}\big) \dots \Gamma\big(\frac{t}{k}, \xi_1, 0\big) d\xi_1 \dots d\xi_{k-1} \quad (4.110)\\ &\geq \int_{B_{k-1}} d\xi_{k-1} \int_{B_{k-2}} d\xi_{k-2} \dots \int_{B_1} d\xi_1 \Gamma\big(\frac{t}{k}, x, \xi_{k-1}\big)\Gamma\big(\frac{t}{k}, \xi_{k-1}, \xi_{k-2}\big) \dots \Gamma\big(\frac{t}{k}, \xi_1, 0\big).\end{aligned}$$

The uniform bound in Theorem 4.9 together with the bounds on distances (4.108) and (4.109) imply that

$$\Gamma\big(\frac{t}{k}, \xi_j, \xi_{j-1}\big) \geq \frac{k^{n/2}}{Ct^{n/2}}, \quad \text{for all } 2 \leq j \leq k-1, \tag{4.111}$$

as well as

$$\Gamma\big(\frac{t}{k}, x, \xi_{k-1}\big) \geq \frac{k^{n/2}}{Ct^{n/2}}, \quad \Gamma\big(\frac{t}{k}, \xi_1, 0\big) \geq \frac{k^{n/2}}{Ct^{n/2}}. \tag{4.112}$$

Using these estimates in (4.110) leads to

$$\Gamma(t,x,0) \geq |B_1|^{k-1} \left(\frac{k^{n/2}}{Ct^{n/2}}\right)^k = C_n^{k-1} \left(\frac{\sqrt{t}}{\sqrt{k}}\right)^{n(k-1)} \left(\frac{k^{n/2}}{Ct^{n/2}}\right)^k = \frac{C_n^{k-1}k^{n/2}}{C^k t^{n/2}} = DK_0^k \left(\frac{k}{t}\right)^{n/2}, \tag{4.113}$$

with constants K_0 and D that depend on λ and Λ. As $4x^2/t < k < 8x^2/t$, we conclude that

$$\Gamma(t,x,0) \geq \frac{1}{Ct^{n/2}} e^{-Cx^2/t}, \tag{4.114}$$

with a constant C that depends on λ, Λ and dimension n. Therefore, Theorem 4.8 is a consequence of Theorem 4.9.

Proof of the uniform lower bound in Theorem 4.9

We now prove the lower bound in Theorem 4.9, that is, we show that

$$\Gamma(t,x,y) \geq \frac{1}{Ct^{n/2}}, \tag{4.115}$$

for all x, y such that $|x-y| \leq \sqrt{t}$. The reason why this estimate holds is, roughly speaking, the following. Solution of the Cauchy problem

$$\phi_t = \nabla \cdot (a(x)\nabla\phi), \tag{4.116}$$

with $\phi(0,x) = \phi_0(x) \geq 0$ conserves mass:

$$\int \phi(t,x)dx = \int \phi_0(x)dx. \tag{4.117}$$

The Gaussian upper bound in Theorem 4.2 means that the total mass outside of the ball $B_N = \{|x| \geq N\sqrt{t}\}$ is small for large N, so that

$$\int_{|x|\leq N\sqrt{t}} \phi(t,x)dx \geq \frac{1}{2}\int \phi_0(x)dx, \tag{4.118}$$

for a sufficiently large N. If we imagine that $\phi(t,x)$ is more or less equally distributed over the ball B_N, we obtain the lower bound (4.115).

In order to simplify slightly the notation let us make the following observation. Let $\phi(t,x)$ be the solution of

$$\frac{\partial\phi}{\partial t} = \nabla \cdot (a(x)\nabla\phi), \tag{4.119}$$

and set $\phi_L(t,x) = \phi(L^2t), Lx)$. The function $\phi_L(t,x)$ satisfies

$$\frac{\partial\phi_L(t,x)}{\partial t} = L^2\frac{\partial\phi(L^2t,Lx)}{\partial t} = L^2\nabla\cdot(a\nabla\phi)(L^2t,Lx) = \nabla\cdot(a_L(x)\nabla\phi_L(t,x)), \tag{4.120}$$

which is an equation

$$\frac{\partial \phi_L(t,x)}{\partial t} = \nabla \cdot (a_L(x)\nabla\phi_L(t,x)), \tag{4.121}$$

of the same form as (4.119) but with the diffusion matrix $a(x)$ replaced[4] by $a_L(x) = a(Lx)$. Let us investigate how the Green's functions $\Gamma(t,x,y)$ and $\Gamma_L(t,x,y)$ for (4.119) and (4.121) are related. We know from the above that if $\phi(t,x)$ solves (4.119) with the initial data

$$\phi(0,x) = \phi_0(x),$$

and $\phi_L(t,x)$ solves (4.121) with the initial data

$$\phi_L(0,x) = \phi_0(Lx),$$

then

$$\phi_L(t,x) = \phi(L^2t, Lx).$$

In other words, we have the identity

$$\int \Gamma(L^2t, Lx, y)\phi_0(y)dy = \int \Gamma_L(t,x,y)\phi_0(Ly)dy = \frac{1}{L^n}\int \Gamma_L(t,x,\frac{y}{L})\phi_0(y)dy, \tag{4.122}$$

for all initial data $\phi_0 \in L^1$. As ϕ_0 is an arbitrary function, we deduce the following scaling relation

$$\Gamma(L^2t, Lx, Ly) = \frac{1}{L^n}\Gamma_L(t,x,y). \tag{4.123}$$

Therefore, in order to show that

$$\Gamma(t,x,y) \geq \frac{1}{Ct^{n/2}}, \text{ for all } |x-y| \leq \sqrt{t}, \tag{4.124}$$

it is sufficient to show that there exists a constant C that does not depend on L so that

$$\Gamma_L(1,x,y) \geq \frac{1}{C}, \text{ for all } |x-y| \leq 1. \tag{4.125}$$

The matrices $a(x)$ and $a_L(x)$ have the same ellipticity constants. Hence, we can reformulate Theorem 4.9 as the statement that there exists a constant $C > 0$ that depends only on the ellipticity constants and dimension so that

$$\Gamma(1,x,y) \geq \frac{1}{C}, \text{ for all } |x-y| \leq 1, \tag{4.126}$$

and this is what we will prove.

The key ingredient in the proof is, once again, an integral bound.

[4]As a slight digression, we mention an important question of what happens when is L large, meaning that we observe the original solution $\phi(t,x)$ after long times $t \sim L^2$ and on large scales $x \sim L$. This is the scope of the homogenization theory [12] that is particularly well developed when $a(x)$ is either periodic or random in x.

Lemma 4.10 *For every $\tau > 0$ there exists a constant B that depends only on λ, Λ, τ and n so that we have*

$$\int e^{-\pi|y|^2} \log \Gamma(\tau, x, y) dy \geq -B_\tau, \tag{4.127}$$

for all x such that $|x| \leq 1$.

Let us explain why this lemma is sufficient to prove the lower bound (4.126). Take any x and y so that $|x - y| \leq 1$. Without loss of generality we may assume that $|x| \leq 1$ and $y = 0$. The semi-group property implies that

$$\Gamma(1, x, 0) = \int \Gamma(1/2, x, \xi)\Gamma(1/2, \xi, 0) d\xi \geq \int \Gamma(1/2, x, \xi)\Gamma(1/2, \xi, 0) e^{-\pi|\xi|^2} d\xi \tag{4.128}$$

Applying Jensen's inequality, recalling that $\Gamma(t, \xi, 0) = \Gamma(t, 0, \xi)$, and using Lemma 4.10 gives

$$\log \Gamma(1, x, 0) \geq \int e^{-\pi|\xi|^2} \log \Gamma(1/2, x, \xi) d\xi + \int e^{-\pi|\xi|^2} \log \Gamma(1/2, \xi, 0) d\xi \geq -2B_{1/2}, \tag{4.129}$$

so that (4.126) holds.

The proof of Lemma 4.10

The very last step in the proof of Theorem 4.1 is to prove Lemma 4.10. Fix x such that $|x| \leq 1$, take any $\varepsilon > 0$, and set $u(t, y) = \Gamma(t, y, x) + \varepsilon$, and

$$G(t) = \int e^{-\pi|y|^2} \log u(t, y) dy. \tag{4.130}$$

The role of ε here is simply to make the integral above "clearly convergent" – otherwise, there may be a hypothetical problem as $|y| \to +\infty$, where $\Gamma(t, y, x)$ is very small. All our bounds will be uniform in ε. If we momentarily set $\varepsilon = 0$ then

$$\int u(t, y) dy = 1,$$

for all $t > 0$, Jensen's inequality implies then

$$0 = \log \Big(\int u(t, y) dy \Big) \geq \log \Big(\int u(t, y) e^{-\pi|y|^2} dy \Big) \geq \int (\log u(t, y)) e^{-\pi|y|^2} dy = G(t), \tag{4.131}$$

so $G(t) \leq 0$ when $\varepsilon = 0$, showing it is the lower bound that is non-trivial. Our goal is to show that, with $\varepsilon > 0$, $G(t)$ is bounded from below for each $t > 0$, uniformly in $|x| \leq 1$, and $\varepsilon > 0$. Note that if $\varepsilon = 0$ then $G(0)$ is not very well defined but $G(s) \to -\infty$ as $s \downarrow 0$ since $u(0, y) = \delta(y - x)$. Therefore an estimate from below for $G(s)$ that is uniform in ε is not an a priori obvious fact. Let us obtain a differential inequality for $G(t)$:

$$\begin{aligned} \frac{dG}{dt} &= \int \frac{1}{u(t, y)} \nabla \cdot (a(y) \nabla u(t, y)) e^{-\pi|y|^2} dy = - \int a(y) \nabla u(t, y) \cdot \nabla \Big(\frac{e^{-\pi|y|^2}}{u(t, y)} \Big) dy \\ &= 2\pi \int a(y) \nabla (\log u(t, y)) \cdot y e^{-\pi|y|^2} dy + \int (a(y) \nabla (\log u(t, y)) \cdot \nabla (\log u(t, y)) e^{-\pi|y|^2} dy. \end{aligned}$$

Let us rewrite the integrands above:

$$a(y)\nabla(\log u)\cdot\nabla(\log u)+2\pi a(y)\nabla(\log u)\cdot y=\frac{1}{2}a(y)\nabla(\log u)\cdot\nabla(\log u) \quad (4.132)$$
$$+\frac{1}{2}a(y)(\nabla(\log u)+2\pi y)\cdot(\nabla(\log u)+2\pi y)-2\pi^2(a(y)y\cdot y).$$

The first term in the right side is positive, which is good for us. Dropping the first term in the second line above gives

$$\frac{dG}{dt}\geq -2\pi^2\int(a(y)y\cdot y)e^{-\pi|y|^2}dy+\frac{1}{2}\int a(y)\nabla(\log u(t,y))\cdot\nabla(\log u(t,y))e^{-\pi|y|^2}dy$$
$$\geq -A+\frac{\lambda}{2}\int|\nabla(\log u(t,y))|^2e^{-\pi|y|^2}dy, \quad (4.133)$$

with a constant A that depends only on the ellipticity constants of the matrix $a(x)$. Therefore, the function $G(t)+At$ is non-decreasing for $t>0$, which is the right direction. It is not, however, sufficient since at the moment we do not have an ε-independent lower bound for $G(t)$ at any time, so saying that, for instance, $G(1)>-A+G(0)$ will not be of much use. What we will use is that the positive term in the right side of (4.133) is quadratic in $\log u$.

We will need the result of the following exercise.

Exercise 4.11 Let $d\mu(x)=e^{-\pi|x|^2}dx$. Show that there exists a constant $C>0$ so that for any function $\phi\in H^1(\mathbb{R};d\mu)$ we have

$$\int|\phi(x)-\langle\phi\rangle|^2d\mu(x)\leq C\int|\nabla\phi|^2d\mu(x), \quad (4.134)$$

with

$$\langle\phi\rangle=\int\phi(x)d\mu(x).$$

Thus, have the Poincaré inequality in the whole space

$$\int_{\mathbb{R}^n}(\log w(y)-\langle\log w\rangle_\mu)^2d\mu(y)\leq C\int_{\mathbb{R}}|\nabla(\log w(y))|^2d\mu(y). \quad (4.135)$$

A good reference for such generalized Poincaré inequalities is the book [84].

In our situation, this inequality takes the form

$$\int(\log u(t,y)-G(t))^2e^{-\pi|y|^2}dy\leq C\int|\nabla(\log u(t,y))|^2e^{-\pi|y|^2}dy. \quad (4.136)$$

Therefore, we have

$$\frac{dG}{dt}\geq -A+B\int(\log u(t,y)-G(t))^2e^{-\pi|y|^2}dy, \quad (4.137)$$

with the constants A and B that depend only on the ellipticity constants of the matrix $a(x)$. Given any $D\in\mathbb{R}$, the function

$$p(u)=\frac{(\log u-D)^2}{u}$$

is decreasing in u for $u > e^{2+D}$. In addition, we know from the upper bound on $\Gamma(t,x,y)$ that there exists a constant K_τ so that $u(s,y) \le K_\tau$ for all $y \in \mathbb{R}^n$ and all $\tau/2 \le t \le \tau$. Therefore, for all $\tau/2 \le t \le \tau$ we have

$$\frac{dG}{dt} \ge -A + B\int_{S_t} \frac{(\log u(t,y) - G(t))^2}{u(t,y)} u(t,y)e^{-\pi|y|^2}dy \qquad (4.138)$$
$$\ge -A + B\frac{(\log K - G(t))^2}{K}\int_{S_t} u(t,y)e^{-\pi|y|^2}dy.$$

Here S_t is the set

$$S_t = \{u(t,y) \ge e^{2+G(t)}\}.$$

If $G(t)$ is very negative (which is what we are trying to avoid), the set S_t is very large. The integral over S_t may be estimated as follows, using the fact that $u(t,y) \le e^{2+G(t)}$ for $y \notin S_t$, and $u(t,y) \ge e^{2+G(t)}$ for $y \in S_t$:

$$\int_{S_t} u(t,y)e^{-\pi|y|^2}dy \ge \int_{S_t} (u(t,y) - e^{2+G(t)})e^{-\pi|y|^2}dy \ge \int_{\mathbb{R}^n} (u(t,y) - e^{2+G(t)})e^{-\pi|y|^2}dy$$
$$= \int_{\mathbb{R}^n} u(t,y)e^{-\pi|y|^2}dy - e^{2+G(t)}. \qquad (4.139)$$

Next, as

$$\int_{\mathbb{R}^n} \Gamma(t,x,y)dy = 1, \qquad (4.140)$$

the upper Gaussian bounds on $\Gamma(t,x,y)$ imply that for any $\tau > 0$ there exists a constant c_0 (that depends on τ) so that

$$\int_{\mathbb{R}^n} \Gamma(t,x,y)e^{-\pi|y|^2/2}dy \ge c_0, \qquad (4.141)$$

for all $\tau/2 \le t \le \tau$. The same upper bound on $\Gamma(t,x,y)$, together with (4.141) implies that there exists R (that also depends on τ) so that

$$\int_{|y|\ge R} u(t,y)e^{-\pi|y|^2/2}dy \le \frac{c_0}{2}, \qquad (4.142)$$

also for all $\tau/2 \le t \le \tau$. This is the crucial step in the proof: the upper bound necessitates the lower bound on $\Gamma(t,x,y)$. Returning to (4.139) we get

$$\int_{S_t} u(t,y)e^{-\pi|y|^2}dy \ge e^{-\pi|R|^2/2}\int_{|y|\le R} u(t,y)e^{-\pi|y|^2/2}dy - e^{2+G(t)} \ge \frac{c_0e^{-\pi|R|^2}}{2} - e^{2+G(t)}. \qquad (4.143)$$

Inequality (4.138) now becomes

$$\frac{dG}{dt} \ge -A + B\frac{(\log K - G(t))^2}{K}\left[\frac{c_0e^{-\pi|R|^2}}{2} - e^{2+G(t)}\right], \quad \frac{\tau}{2} \le t \le \tau. \qquad (4.144)$$

Assume now that $G(\tau) < -M$ for some large M. The function $G(s) + As$ is non-decreasing in time, hence

$$G(t) \le G(\tau) + A\tau - At < -M/2, \tag{4.145}$$

for all $t \in [\tau/2, \tau]$ provided that $M > 100A\tau$. Suppose that M is so large that (4.145) implies that

$$e^{-\pi R^2} > 10e^{2+G(s)},$$

for all $\tau/2 \le s \le \tau$. Then, still under the assumption $G(\tau) < -M$, (4.144) implies that

$$\frac{dG}{dt} \ge -A + B\frac{(\log K - G(t))^2}{K}\frac{c_0 e^{-\pi|R|^2}}{5}, \quad \text{for } \frac{\tau}{2} \le t \le \tau. \tag{4.146}$$

However, if M is much larger than all other constants appearing in (4.146), and $G(t) \le -M/2$ for all $t \in [\tau/2, \tau]$, it follows from the last inequality that

$$\frac{dG}{dt} \ge cG(t)^2, \quad \text{for } \frac{\tau}{2} \le t \le \tau, \tag{4.147}$$

with the constant c that still depends only on τ and the ellipticity constants of the matrix $a(x)$. However, this quadratic inequality blows up in a finite "backward" time, so if $G(t)$ satisfies (4.147), and $G(\tau/2) > -\infty$, it is impossible that $G(\tau) < -M$ for too large M (that depends explicitly on constant c). This gives an a priori lower bound on $G(\tau)$ that depends only on τ and the ellipticity constants of the matrix $a(x)$ and is uniform in $\varepsilon > 0$. In order to remove the need for the regularization $\varepsilon > 0$ note that we have shown

$$\int e^{-\pi|y|^2} \log(\Gamma(\tau, x, y) + \varepsilon)dy > -B_\tau. \tag{4.148}$$

As the function $\Gamma(\tau, x, y)$ is uniformly bounded from above by a constant $K(\tau)$, it follows from (4.148) that

$$\int e^{-\pi|y|^2} \log_-(\Gamma(\tau, x, y) + \varepsilon)dy > -B'_\tau, \tag{4.149}$$

with some constant B'_τ. Here $\log_- u = 0$ if $u > 1$ and $\log_- u = \log u$ if $u \in (0, 1)$. Fatou's lemma now shows that

$$\int e^{-\pi|y|^2} \log_- \Gamma(\tau, x, y)dy > -B'_\tau. \tag{4.150}$$

This completes the proof of Theorem 4.1!

5 Gaussian bounds on the heat kernel imply everything

We will now show, once again following Fabes and Stroock [56], that the bounds on the heat kernel imply "all" classical regularity results on the parabolic equations in the divergence form. The physical reason for this implication is simple. Parabolic and elliptic equations tend to equilibrate locally, as can be seen, for instance, from the mean value property for harmonic functions. A potential enemy of this tendency is the "outside influence" – for example, if the solution is wild outside a ball, it may spoil the equilibrating properties inside the ball too, The heat kernel bounds provide two remedies against the outside influence – first, the upper Gaussian bounds impose limits on the influence of what happens outside the ball $B(x,R)$ on the solution inside a slightly smaller ball $B(x,\delta R)$ with $\delta<1$. On the other hand, the lower bounds on the heat kernel show that local influence is quite strong – the combination of the two allows to prove local regularity results.

5.1 A lower bound for Green's functions on a bounded domain

As before, we will denote by $\Gamma(t,x,y)$ the Green's function for the Cauchy problem

$$\begin{aligned}&\phi_t=\nabla\cdot(a(x)\nabla\phi),\quad t>0,\quad x\in\mathbb{R}^n,\\&\phi(0,x)=\phi_0(x).\end{aligned}\tag{5.1}$$

That is, solution of (5.1) can be written as

$$\phi(t,x)=\int_{\mathbb{R}^n}\Gamma(t,x,y)\phi_0(y)dy.\tag{5.2}$$

In order to bound the "outside influence" on $\phi(t,x)$ in a ball $B(\xi,R)$ we will consider the worst case scenario (assume for the sake of intuition that $\phi_0(x)\geq 0$), setting the Dirichlet boundary condition on the boundary $\partial B(\xi,R)$. The maximum principle implies that the true solution satisfies $\phi(t,x)>0$ on $\partial B(\xi,R)$, hence in that way we will account for the "strongest outside cooling influence".

To formalize this idea, we will make use of the Green's function for the Cauchy problem on bounded domains. Let $B(\xi,R)\subset\mathbb{R}^n$ be a ball of radius R centered at a point $\xi\in\mathbb{R}^n$. We will denote by $\Gamma_{\xi,R}(t,x,y)$ the Green's function for the problem

$$\begin{aligned}&\psi_t=\nabla\cdot(a(x)\nabla\psi),\quad t>0,\quad x\in B(\xi,R),\\&\psi(0,x)=\psi_0(x),\quad x\in B(\xi,R)\\&\psi(t,y)=0\text{ for }y\in\partial B(\xi,R).\end{aligned}\tag{5.3}$$

The function $\psi(t,x)$ has a representation

$$\psi(t,x)=\int_{B(\xi,R)}\Gamma_{\xi,R}(t,x,y)\psi_0(y)dy.\tag{5.4}$$

Our immediate task will be to find uniform lower bounds for $\Gamma_{\xi,R}(t,x,y)$ strictly inside the ball $B(\xi,R)$, that is, in a slightly smaller ball $B(\xi,\delta R)$ – we can not possibly expect such bounds all the way to the boundary as $\Gamma_{\xi,R}$ vanishes there.

It is instructive to write an equation for the function $\psi(t,y)$ in the whole space, in the sense of distributions. For any smooth test function $\eta\in\mathcal{S}(\mathbb{R}^n)$ (the Schwartz class), we have

$$\begin{aligned}\int_{\mathbb{R}^n}\psi(t,x)\nabla\cdot(a(x)\nabla\eta(x))dx&=\int_{B(\xi,R)}\psi(t,x)\nabla\cdot(a(x)\nabla\eta(x))dx\\&=\int_{\partial B(x,R)}\psi(t,y)(a(y)\nabla\eta(y)\cdot\nu)dy-\int_{B(\xi,R)}(a(x)\nabla\eta(x)\cdot\nabla\psi(t,x))dx.\end{aligned}\tag{5.5}$$

Here, ν is the outward normal to the sphere $\partial B(\xi, R)$. The first term in the right side vanishes because of the Dirichlet boundary conditions, leading to

$$\int_{\mathbb{R}^n} \psi(t,x)\nabla\cdot(a(x)\nabla\eta(x))dx = -\int_{B(\xi,R)} (a(x)\nabla\eta(x)\cdot\nabla\psi(t,x))dx \tag{5.6}$$
$$= -\int_{\partial B(x,R)} \eta(y)(a(y)\nabla\psi(t,y)\cdot\nu)dy + \int_{B(\xi,R)} \eta(x)\nabla\cdot(a(x)\nabla\psi(t,x))dx$$
$$= -\int_{\partial B(x,R)} \eta(y)(a(y)\nabla\psi(t,y)\cdot\nu)dy + \int_{B(\xi,R)} \eta(x)\psi_t(t,x)dx.$$

Therefore, $\psi(t,x)$ satisfies the following problem in the whole space:

$$\psi_t = \nabla\cdot(a(x)\nabla\psi) + (a(x)\nabla\psi(t,x)\cdot\nu)\delta_{\partial B(\xi,R)}(x), \quad t>0, \quad x\in\mathbb{R}^n, \tag{5.7}$$
$$\psi(0,x) = \psi_0(x), \quad x\in\mathbb{R}^n.$$

Recall that if $\psi_0(x)\geq 0$ then $\psi(t,x)\geq 0$ inside $B(\xi,R)$. Therefore, as $\psi(t,x)=0$ on $\partial B(\xi,R)$, we have

$$(a(x)\nabla\psi\cdot\nu)\leq 0 \text{ for } x\in\partial B(\xi,R).$$

Hence, the source in (5.6) is negative, as it should be from the physical considerations – the boundary has the cooling effect. Duhamel's principle implies that $\psi(t,x)$ can be written as

$$\psi(t,x) = \int_{B(\xi,R)} \Gamma(t,x,y)\psi_0(y)dy + \int_0^t\int_{\partial B(\xi,R)} \Gamma(t-s,x,y)(a(y)\nabla\psi(s,y)\cdot\nu(y))dy. \tag{5.8}$$

If we take the initial data $\psi_0(y)=\delta(y-z)$ with some $z\in B(\xi,R)$, we get from (5.8) an integral equation for $\Gamma_{\xi,R}$:

$$\Gamma_{\xi,R}(t,x,z) = \Gamma(t,x,z) + \int_0^t\int_{\partial B(\xi,R)} \Gamma(t-s,x,y)(a(y)\nabla\Gamma_{\xi,R}(s,y,z)\cdot\nu(y))dyds. \tag{5.9}$$

The maximum principle implies, once again, that

$$(a(y)\nabla\Gamma_{\xi,R}(s,y,z)\cdot\nu(y)) < 0,$$

so we may rewrite (5.10) as

$$\Gamma_{\xi,R}(t,x,z) = \Gamma(t,x,z) - \int_0^t\int_{\partial B(\xi,R)} \Gamma(t-s,x,y)d\mu(y)ds. \tag{5.10}$$

Here, we have defined the measure

$$d\mu(s,y) = -(a(y)\nabla\Gamma_{\xi,R}(s,y,z)\cdot\nu(y))dy.$$

In order to estimate the effect of the cold boundary in (5.10), we go back to the equation for $\Gamma_{\xi,R}$:

$$\frac{\partial\Gamma_{\xi,R}}{\partial t} = \nabla\cdot(a(x)\nabla\Gamma_{\xi,R}), \quad t>0, \quad x\in B(\xi,R), \tag{5.11}$$
$$\Gamma_{\xi,R}(0,x) = \delta(x-z), \quad x\in B(\xi,R)$$
$$\Gamma_{\xi,R}(t,y) = 0 \text{ for } y\in\partial B(\xi,R),$$

and integrate over the ball $B(\xi,R)$ and in time:

$$\int_{B(\xi,R)} \Gamma_{\xi,R}(t,x,z)dx - 1 = \int_0^t \int_{\partial B(\xi,R)} (a(y)\nabla\Gamma_{\xi,R}(s,y,z)\cdot\nu(y))dyds = -\int_0^t \int_{\partial B(\xi,R)} d\mu(s,y)ds. \tag{5.12}$$

We conclude that

$$\int_0^t \int_{\partial B(\xi,R)} d\mu(s,y)ds < 1, \tag{5.13}$$

for all $t > 0$. This is a very important point – we have a bound on the total effect of the cold boundary over time.

A lower bound on $\Gamma_{\xi,R}(t,x,z)$ for nearby points and short times

Take now any $\delta \in (0,1)$ and assume that both x and z lie in the "slightly smaller" ball $B(\xi,\delta R)$. Then the upper Gaussian bound on $\Gamma(t,x,y)$ that we have already proved, together with (5.10) and (5.13) imply that

$$\Gamma_{\xi,R}(t,x,z) \geq \Gamma(t,x,z) - \sup_{0\leq\tau\leq t} \frac{C}{\tau^{n/2}} e^{-(1-\delta)^2R^2/(C\tau)}. \tag{5.14}$$

Note that the upper bound on $\Gamma(t,x,y)$ gives a lower bound on the Dirichlet Green's function away from the boundary – this is exactly the phenomenon we have mentioned – the upper bounds limit the influence of the cold boundary strictly inside the domain. Next, the lower Gaussian bound on $\Gamma(t,x,z)$ gives

$$\Gamma_{\xi,R}(t,x,z) \geq \frac{1}{Ct^{n/2}} e^{-C|x-z|^2/t} - \sup_{0\leq\tau\leq t} \frac{C}{\tau^{n/2}} e^{-(1-\delta)^2R^2/(C\tau)}. \tag{5.15}$$

Here, we see the competition between the "heating from inside" given by the first term in the right side, and the cooling by the boundary expressed by the second term in the right side of (5.15). We deduce from (5.15) that there exists $r \in (0, 1-\delta)$, which depends only on δ, so that for all $0 < t \leq r^2R^2$ and $|z-x| < rR$, with $z, x \in B(\xi,\delta R)$, we have

$$\Gamma_{\xi,R}(t,x,z) \geq \frac{1}{2Ct^{n/2}} e^{-C|x-z|^2/t}. \tag{5.16}$$

The constant C depends only on δ but not on R.

The fact that we obtained first a lower bound on $\Gamma_{\xi,R}(t,x,z)$ only for nearby points x and z and at short times is very natural – by virtue of being inside $B(\xi,\delta R)$, these points are separated from the boundary of $B(x,R)$ (where the Dirichelt boundary condition is imposed) by the distance $(1-\delta)R$ which is larger than $|x-z|$. Therefore, it is reasonable to expect that the "warming" influence of z at x at short times is stronger than the combined "cooling" influence of all boundary points on $\partial B(x,R)$ that are too far away to compete.

Extension of the lower bound to all of $B(\xi, \delta R)$ and larger times

In order to extend this estimate to all of the smaller ball $B(\xi, \delta R)$ and all times in an interval of the form $\gamma R^2 \le t \le R^2$ with some $\gamma > 0$, we use a simpler version of the argument in the proof of Theorem 4.9. Take any $x, z \in B(\xi, \delta R)$ and $t \le R^2$. Consider a sequence of balls

$$B_j = B(\xi_j, rR/3),\ j = 1, \dots, k-1,$$

such that $\xi_1 = z$, $x \in B_{k-1}$ and each next center $\xi_{j+1} \in B_j$. By possibly increasing the number of balls we can also ensure that $t/k \le r^2 R^2$, our threshold for a "short time" in (5.16). The total number k of the required balls is bounded by

$$k \le 1 + \max\left[\frac{|x-z|}{10rR}, \frac{t}{R^2 r^2}\right] \le 1 + \max\left[C(\delta), \frac{t}{R^2 r^2}\right]. \tag{5.17}$$

Therefore, as $t \le R^2$, we conclude that k is bounded by a constant K_δ that only depends on δ:

$$k \le K_\delta \text{ for } t \le R^2. \tag{5.18}$$

The semigroup property of Green's function $\Gamma_{\xi,R}(t, x, z)$ implies that we may iterate:

$$\Gamma_{\xi,R}(t,x,z) = \int\limits_{B(\xi,R)} \dots \int\limits_{B(\xi,R)} \Gamma_{\xi,R}\big(\frac{t}{k}, x, \xi_{k-1}\big)\Gamma_{\xi,R}\big(\frac{t}{k}, \xi_{k-1}, \xi_{k-2}\big) \dots \Gamma_{\xi,R}\big(\frac{t}{k}, \xi_1, z\big) d\xi_1 \dots d\xi_{k-1}$$

$$\ge \int\limits_{B_{k-1}} d\xi_{k-1} \int\limits_{B_{k-2}} d\xi_{k-2} \dots \int\limits_{B_1} d\xi_1 \Gamma_{\xi,R}\big(\frac{t}{k}, x, \xi_{k-1}\big)\Gamma_{\xi,R}\big(\frac{t}{k}, \xi_{k-1}, \xi_{k-2}\big) \dots \Gamma_{\xi,R}\big(\frac{t}{k}, \xi_1, z\big). \tag{5.19}$$

If $t \ge \gamma R^2$ then for any $z, z' \in B(\xi, \delta R)$ we have

$$\frac{(z-z')^2}{t} \le \frac{\delta^2 R^2}{\gamma R^2} = \frac{\delta^2}{\gamma}.$$

Moreover, our choice of the balls ensures that if $\xi_j \in B_j$ and $\xi_{j+1} \in B_{j+1}$, then estimate (5.16) applies to $\Gamma_{\xi,R}(t/k, \xi_j \xi_{j+1})$. Hence, in the range $\gamma R^2 \le t \le R^2$, the above bound, implies that

$$\Gamma_{\xi,R}\big(\frac{t}{k}, \xi_j, \xi_{j-1}\big) \ge \frac{1}{C' t^{n/2}}, \quad \text{for all } 2 \le j \le k, \tag{5.20}$$

with a constant C' that depends only on γ and δ. Similarly, we have

$$\Gamma_{\xi,R}\big(\frac{t}{k}, x, \xi_k\big) \ge \frac{1}{C t^{n/2}}, \quad \Gamma_{\xi,R}\big(\frac{t}{k}, \xi_1, z\big) \ge \frac{1}{C t^{n/2}}. \tag{5.21}$$

Using these estimates in (5.19) leads to

$$\Gamma(t,x,z) \ge |B_1|^{k-1}\left(\frac{1}{Ct^{n/2}}\right)^k = C_n^{k-1}(rR)^{n(k-1)}\left(\frac{1}{Ct^{n/2}}\right)^k = \frac{DK_0^k}{R^n}\left(\frac{r^2R^2}{t}\right)^{nk/2}, \tag{5.22}$$

with the constants K_0 and D that only depend on δ. As $\gamma R^2 < t \le R^2$, and k obeys the upper bound (5.18) we simply get

$$\Gamma(t,x,z) \ge \frac{C}{R^n}, \quad \gamma R^2 < t \le R^2, \quad x, z \in B(\xi, \delta R). \tag{5.23}$$

Let us summarize this result.

Theorem 5.1 *For each $\delta \in (0,1)$ and $\gamma > 0$ there exists c_0 that depends only on the ellipticity constants of the matrix $a(x)$, dimension n, δ and γ so that for all $\xi \in \mathbb{R}^n$ and all $R > 0$ we have a lower bound*

$$\Gamma_{\xi,R}(t,x,z) \geq \frac{c_0}{R^n}, \tag{5.24}$$

for all $x, z \in B(\xi, \delta R)$ and all $\gamma R^2 \leq t \leq R^2$.

It is this corollary of the Gaussian heat kernel bounds that will be crucial in the proof of parabolic regularity properties below: it limits the outside influence!

5.2 A decay of oscillation estimate

The lower bound on the Green's function in a ball that we have obtained above implies a decay of oscillation estimate. Consider a parabolic cylinder

$$D(s,\xi;R) = \{s - R^2 \leq t \leq s, \ \ |x-\xi| \leq R\}, \tag{5.25}$$

and the corresponding oscillation of a function u over $D(s,\xi;R)$:

$$\mathrm{Osc}(u;s,\xi,R) = \sup\{|u(t,x,) - u(t',x')| : \ \ (t,x),(t',x') \in D(s,\xi;R)\}. \tag{5.26}$$

We will now deduce from the Green function bounds the following decay of oscillation estimate.

Theorem 5.2 *For each $\delta \in (0,1)$ there exists $\rho < 1$ that depends only on dimension n, the ellipticity constants of the matrix $a(x)$ and δ, so that any solution $u \in C^\infty([s-R^2,s]\times\bar{B}(\xi,R))$ of*

$$u_t = \nabla\cdot(a(x)\nabla u), \quad s - R^2 < t < s, \quad x \in B(\xi,R), \tag{5.27}$$

satisfies

$$\mathrm{Osc}(u;s,\xi,\delta R) \leq \rho \mathrm{Osc}(u;s,\xi,R). \tag{5.28}$$

Proof. Let $m(r)$ and $M(r)$ denote the minimum and maximum of u over the parabolic cylinder $D(s,\xi;r)$. Consider the set

$$S = \Big\{x \in B(\xi,\delta R) : \ \ u(s-R^2,x) \geq \frac{M(R)+m(R)}{2}\Big\},$$

where the solution is "large" at time $s - R^2$. Assume first that this set itself is relatively large: $|S| \geq |B(\xi,\delta R)|/2$. Together with the bounds on the Dirichlet Green's function this will be enough to show that at any time separated from $s - R^2$:

$$s - \delta^2 R^2 < t \leq s,$$

solution inside the ball $B(\xi,\delta R)$ will be "not too small". That is, since $u(s-R^2,x)$ is large on a big set inside $B(x,\delta R)$, after a short time it will be "large enough" on all of $B(\xi,\delta R)$. To this end, note that the function

$$u_1(t,x) = u(t,x) - m(R)$$

satisfies the same equation as $u(t,x)$, and is positive for all $x \in D(s,\xi,R)$. In particular, we know that $u_1(t,x) \geq 0$ for all $s-R^2 \leq t \leq s$ and all $x \in \partial B(\xi,R)$. The function

$$u_2(x) = \int_{B(\xi,R)} (u(s-R^2,y) - m(R))\Gamma_{\xi,R}(t-(s-R^2),x,y)dy$$

satisfies the same parabolic equation as $u_1(t,x)$, but $u_2(t,x) = 0$ for all $s-R^2 \leq t \leq s$ and all $x \in \partial B(\xi,R)$, and, in addition, $u_1(s-R^2,x) = u_2(s-R^2,x)$ for all $x \in B(\xi,R)$. It follows from the maximum principle that $u_1(t,x) \geq u_2(t,x)$ for all $(t,x) \in D(s,\xi,R)$. Therefore, for $(t,x) \in D(s,\xi,\delta R)$ we have

$$\begin{aligned} u(t,x) - m(R) &\geq \int_{B(\xi,R)} (u(s-R^2,y) - m(R))\Gamma_{\xi,R}(t-(s-R^2),x,y)dy \\ &\geq \frac{M(R)-m(R)}{2}\int_S \Gamma_{\xi,R}(t-(s-R^2),x,y)dy \geq \frac{M(R)-m(R)}{2}\frac{c_\delta}{R^n}|S| \\ &= \varepsilon(M(R)-m(R)). \end{aligned} \tag{5.29}$$

We used Theorem 5.1 in the last step above, since $t-(s-R^2) \geq (1-\delta^2)R^2$, and also the assumption $|S| \geq |B(\xi,\delta R)|/2$. The constant ε does not depend on R or u. It follows that

$$m(\delta R) \geq m(R) + \varepsilon(M(R)-m(R)),$$

and thus

$$M(\delta R) - m(\delta R) \leq M(R) - m(\delta R) \leq (1-\varepsilon)(M(R)-m(R)). \tag{5.30}$$

On the other hand, if S is small: $|S| \leq |B(\xi,\delta R)|/2$, we would simply consider the difference $M(R) - u(t,x)$ for any $s-\delta^2R^2 < t \leq s$, and $x \in B(\xi,\delta R)$:

$$\begin{aligned} M(R) - u(t,x) &\geq \int_{B(\xi,R)} (M(R) - u(s-R^2,y))\Gamma_{\xi,R}(t-(s-R^2),x,y)dy \\ &\geq \frac{M(R)-m(R)}{2}\int_{S^c} \Gamma_{\xi,R}(t-(s-R^2),x,y)dy \geq \frac{M(R)-m(R)}{2}\frac{c_\delta}{R^n}|S^c| \\ &= \varepsilon(M(R)-m(R)), \end{aligned} \tag{5.31}$$

which also implies (5.30). $\Box$

5.3 The Hölder regularity

The decay of oscillations implies the Hölder regularity of solutions – we will use an iterative local blow-up argument. Consider some $s > 0$, $\xi \in \mathbb{R}^n$ and R such that $s - R^2 > 0$, and assume that $u(t,x)$ satisfies the parabolic equation in the parabolic cylinder $D(s,\xi,R)$:

$$u_t = \nabla \cdot (a(x)\nabla u), \quad x \in B(\xi,R), \quad s-R^2 \leq t \leq s. \tag{5.32}$$

We are going to show that if $u(t,x)$ is bounded in $D(s,\xi,R)$ then $u(t,x)$ has to satisfy Hölder a priori bounds in a smaller set

$$D_\delta(s,\xi) = \{s-(1-\delta^2)R^2 \leq t \leq s, \quad |x-\xi| \leq (1-\delta)R\},$$

for any $\delta \in (0,1)$. The main point is that if we step slightly inside $D(s,\xi,R)$, away from the boundary $|x-\xi| = R$, and from the initial time $t = s - R^2$, then the Hölder norm of u depends only[5] on the L^∞ norm of u in the slightly bigger set $D(s,\xi,R)$. Accordingly, take some t, t' so that

$$s - (1-\delta^2)R^2 \le t' \le t \le s,$$

and $x, x' \in B(\xi, (1-\delta)R)$. Let us also denote

$$l = \sqrt{t-t'} + |x-x'|,$$

so that

$$(t', x') \in [t - l^2, t] \times B(x, l), \tag{5.33}$$

and set

$$M = \sup\{|u(r,y)| : \quad s - R^2 \le r \le s, \quad |y - \xi| \le R\}.$$

Note that

$$|u(t,x) - u(t',x')| \le \mathrm{Osc}(u; t, x, l), \tag{5.34}$$

because of (5.33). Iterating (5.34), going to larger and larger parabolic cylinders, with the help of Theorem 5.2 gives

$$|u(t,x) - u(t',x')| \le \mathrm{Osc}(u; t, x, l) \le \rho \mathrm{Osc}(u; t, x, \frac{l}{\delta}) \le \cdots \le \rho^{m-1}\mathrm{Osc}(u; t, x, \frac{l}{\delta^m}) \le 2M\rho^{m-1}. \tag{5.35}$$

We may iterate (blow-up the cylinder) as long as we stay inside $D(s,\xi,R)$:

$$t - \frac{l^2}{\delta^{2m}} > s - R^2, \tag{5.36}$$

and

$$B(x, \frac{l}{\delta^m}) \subset B(\xi, R). \tag{5.37}$$

For (5.36) to hold, as $t > s - (1-\delta^2)R^2$, it suffices to have

$$s - (1-\delta^2)R^2 - \frac{l^2}{\delta^{2m}} > s - R^2, \tag{5.38}$$

that is,

$$\frac{l}{\delta^{m+1}} < R. \tag{5.39}$$

On the other hand, as $|x - \xi| \le (1-\delta)R$, for (5.37) to hold it is enough to ensure

$$(1-\delta)R + \frac{l}{\delta^m} < R, \tag{5.40}$$

which is nothing but (5.39) again. Let us choose the largest m so that (5.39) holds, that is

$$\delta^{m+2} \le \frac{l}{R} < \delta^{m+1},$$

[5] And of course, also on δ – on how far away from the boundary $|x-\xi| = R$ and from the the initial time $s - R^2$ we are.

then (5.35) gives

$$|u(t,x)-u(t',x')| \le 2M\rho^{m-1} \le C(\delta,\rho)M\exp\big\{\frac{\log\rho}{\log\delta}\log\big(\frac{l}{R}\big)\big\} \le CM\big(\frac{l}{R}\big)^{\beta}, \tag{5.41}$$

with the constants C and β that depend only on δ and λ (recall that ρ itself depends only on δ and λ). Therefore, we have shown that there exist a constant $C>0$ and $\beta>0$ that depend only on δ and λ so that if $u(t,x)$ satisfies

$$u_t = \nabla\cdot(a(x)\nabla u), \quad x\in B(\xi,R), \quad s-R^2\le t\le s, \tag{5.42}$$

then for any t,t' so that $s-(1-\delta^2)R^2 \le t' \le t \le s$ and $x,x'\in B(\xi,(1-\delta)R)$ we have

$$|u(t,x)-u(t',x')| \le C(\lambda,\delta)M\left(\frac{|x-x'|+\sqrt{t-t'}}{R}\right)^{\beta}, \tag{5.43}$$

which is the desired Hölder estimate. Of course, the constant $C(\delta,\lambda)$ blows up as $\delta\downarrow 0$, as expected – the initial condition is assumed to be only locally bounded, not Hölder! But for any $t>0$ any solution is Hölder continuous both in time and space.

5.4 The Harnack inequality

The last step in milking the heat kernel bounds is to prove the Harnack inequality.

Theorem 5.3 *(The Harnack inequality) Let $0<\alpha<\beta<1$ and $0<\delta<1$ be given, and let $u(t,x)\ge 0$ be the solution of*

$$u_t = \nabla\cdot(a(x)\nabla u), \quad x\in \bar B(x,R), \quad s-R^2\le t\le s. \tag{5.44}$$

There exist a constant M that depends on the dimension n, the ellipticity constants of the matrix $a(x)$, and α, β, and δ, but not on R and u such that for all

$$s-\beta R^2 \le t \le s-\alpha R^2 \text{ and } y\in B(x,\delta R),$$

we have

$$u(t,y)\le Mu(s,x). \tag{5.45}$$

The non-negativity of $u(t,x)$ is absolutely essential for the Harnack inequality to hold.

Physically, this means that a hot point at distance r away will heat $u(s,x)$ after a time of the order r^2 passes. The reason are the Hölder bounds – if $u(t,y)$ is large, it is "not too small" in a neighborhood $B(y,r_0)$ of y. The fact that $u\ge 0$ everywhere means that we may bound $u(s,x)$ from below if we simply restrict u to be zero away from $B(y,r_0)$ at time t, and consider the corresponding Cauchy problem starting at time t with this cut-off initial data and the Dirichlet boundary condition at $\partial B(x,R)$. The lower bounds on the Dirichlet Green's function will imply a lower bound on $u(s,x)$.

The above was the perspective that "u at an earlier time bounds u at later time from below". Alternatively, we may think that the Harnack inequality says that "u at a later time bounds u at an earlier time from above". To see it from that point of view, it is convenient to set $(s,x)=(0,0)$ and $R=1$ – the general case follows by the usual shifting and scaling argument. We may also assume that $u(0,0)=1$. The general strategy is as follows: as $u\geq 0$, we may get from the bounds in Theorem 5.1 on the Dirichlet Green's function $\Gamma_{0,1}$ in the unit ball[6] that for any $t<0$ the measure of the set of points y such that $u(t,y)>M$ can not be too large for large M – otherwise, we would have $u(0,0)>1$. On the other hand, if this set is small, then the oscillation of $u(t,y)$ around any point where $u(t,y)>2M$ is large. Iterating backward in time will produce larger and larger oscillation and show that u is unbounded on the time interval $[-1,0]$ which would be a contradiction.

Let us now formalize the above argument. Given any $r\in[-1,-\alpha]$ let $v(t,x)$ satisfy

$$v_t=\nabla\cdot(a(x)\nabla v),\quad x\in\bar B(0,1),\quad r\leq t\leq 0, \tag{5.46}$$

with the initial condition $v(r,x)=u(r,x)$ and the Dirichlet boundary condition $v(t,y)=0$ for $|y|=1$. As $u(t,y)\geq 0$ on the boundary $\partial B(0,1)$, we know that

$$1=u(0,0)\geq\int_{B(0,1)}\Gamma_{0,1}(-r,0,y)u(r,y)dy. \tag{5.47}$$

Theorem 5.1 implies that there exists $\varepsilon>0$, which depends on α, so that for all $M>0$, and all $-1\leq r\leq-\alpha$, we have

$$1=u(0,0)\geq\int_{B(0,1)}\Gamma_{0,1}(-r,0,y)u(r,y)dy\geq\varepsilon M|S(r,M)|, \tag{5.48}$$

where

$$S(t,M)=\{y\in B(0,(1+\delta)/2):\ u(t,y)\geq M\}.$$

We conclude that

$$|S(t,M)|\leq\frac{1}{\varepsilon M}, \tag{5.49}$$

for all $t\in[-1,-\alpha]$ and $M>0$. Suppose that $y\in S(t,M)$, and l is such that $t-4l^2\geq-1$, and $B(y,2l)\in\bar B(0,(1+\delta)/2)$. If $B(y,l)$ is contained in the set $S(t,\sigma M)$ with some $\sigma<1$, then

$$c_nl^n\leq|S(t,\sigma M)|\leq\frac{1}{\varepsilon\sigma M}. \tag{5.50}$$

Let us choose $\sigma=(1-\rho)/2<1$, with ρ as in the decay of oscillation estimate (5.28) in Theorem 5.2, and

$$l=\left(\frac{2}{c_n\varepsilon\sigma M}\right)^{1/n}.$$

Then (5.50) is false, meaning that $B(y,l)$ is not contained inside $S(t,\sigma M)$, and there exists a point $y_1\in B(y,l)$ such that $u(t,y_1)<\sigma M$. It follows that

$$\mathrm{Osc}(u;t,y,l)\geq u(t,y)-u(t,y_1)\geq(1-\sigma)M. \tag{5.51}$$

[6]Here $\Gamma_{0,1}$ denotes the Green's function for the parabolic Dirichlet problem on the unit ball $B(0,1)$, in accordance with the notation of Theorem 5.1.

Applying Theorem 5.2 we deduce that

$$\mathrm{Osc}(u; t, y, 2l) \geq \frac{1}{\rho}\mathrm{Osc}(u; t, y, l) \geq \frac{(1-\sigma)}{\rho} M = KM. \tag{5.52}$$

In particular, there exists $t' \in [t - 4l^2, t]$ and $y' \in B(y, 2l)$ such that

$$u(t', y') \geq KM.$$

Note that if we set

$$\sigma = \frac{1-\rho}{2},$$

then

$$K = \frac{1-\sigma}{\rho} = \frac{1+\rho}{2\rho} > 1.$$

Let us now proceed inductively using the above argument. Assume that there is $t_0 \in [-\beta, -\alpha]$ and $y_0 \in B(0, \delta)$ such that $u(t_0, y_0) \geq M_0$. Then we may find a point t_1, y_1 such that (with a constant c that does not depend on M_0 but rather ε, ρ, etc.)

$$t_0 - \frac{4c}{M_0^{2/n}} \leq t_1 \leq t_0, \quad |y_1 - y_0| \leq \frac{2c}{M_0^{2/n}}$$

so that $u(t_1, y_1) \geq M_1 = KM_0 > M_0$. Iterating, we obtain a sequence of points t_m, y_m so that

$$t_m - \frac{4c}{M_m^{2/n}} \leq t_{m+1} \leq t_m, \quad |y_{m+1} - y_m| \leq \frac{2c}{M_m^{2/n}}$$

so that $u(t_{m+1}, y_{m+1}) \geq M_{m+1} = KM_m = K^m M_0$. Since $K > 1$, if M_0 is sufficiently large (depending on ε, ρ, α and β), the sequence t_m, y_m converges to a point $\bar{t}, \bar{y}$ in $[-\beta, -\alpha] \times B(0, \delta)$. But then $u(\bar{t}, \bar{y})$ is unbounded which is contradiction. This gives a bound on M_0, proving Theorem 5.3.

Chapter 4

KPP invasions in periodic media

In this chapter we will consider equations of the form

$$u_t - \Delta u = \mu(x)u - u^2, \quad t > 0, \quad x \in \mathbb{R}^n, \tag{0.1}$$

with a smooth function $\mu(x)$ that is 1-periodic in all variables x_j, $j = 1, \dots, n$. The question we will focus on is what happens to solutions of (0.1) with compactly supported initial data $u(0, x) = u_0(x)$ such that $0 \le u_0(x) \le 1$. A crucial role in the final result will be played by the periodic eigenvalue problem:

$$\begin{aligned} &-\Delta\phi - \mu(x)\phi = \lambda\phi, \\ &\phi(x) \text{ is 1-periodic in all its variables.} \end{aligned} \tag{0.2}$$

A classical result of the spectral theory for second order elliptic operators (a good basic reference is, as usual, [52]) is that this eigenvalue problem is self-adjoint, has a purely discrete spectrum λ_k, $k \in \mathbb{N}$, with

$$\lim_{k \to +\infty} \lambda_k = +\infty,$$

and all eigenvalues of (0.2) are real. The Krein-Rutman theorem [43], together with the comparison principle, implies that there is a unique eigenvalue λ_1 that corresponds to a positive eigenfunction ϕ_1 (all other eigenfunctions change sign). Moreover, λ_1 is a simple eigenvalue, and it is is the smallest eigenvalue of (0.2). It is called the principal eigenvalue of (0.2), and has a variational characterization in terms of the Rayleigh quotient:

$$\lambda_1 = \inf_{\psi \in H^1(\mathbb{T}^n)} \frac{\displaystyle\int_{\mathbb{T}^n} (|\nabla\psi|^2 - \mu(x)\psi^2)dx}{\displaystyle\int_{\mathbb{T}^n} |\psi(x)|^2 dx}. \tag{0.3}$$

Here $\mathbb{T}^n = [0, 1]^n$ is the n-dimensional torus (the unit period cell of $\mu(x)$), and $H^1(\mathbb{T}^n)$ is the set of all 1-periodic functions in the Sobolev space H^1. Our main assumption about the function $\mu(x)$ will be that

$$\lambda_1 < 0. \tag{0.4}$$

This condition holds, for instance, if the (continuous) function $\mu(x)$ is non-negative and not identically equal to zero in $\mathbb{T}^n$: this can be seen by simply taking the test function $\psi(x) \equiv 1$ in (0.4) Assuming (0.4), we will show the following: first, the steady equation

$$-\Delta u = \mu(x)u - u^2, \quad x \in \mathbb{R}^n, \tag{0.5}$$

posed in the whole space, has a unique positive bounded solution $u_+(x)$. Moreover, $u_+(x)$ is 1-periodic in all variables. Second, any solution of the Cauchy problem for (0.1) with a nonnegative, bounded and compactly supported initial data $u_0(x)$ (that is positive on some open set) will tend to $u_+(x)$ as $t \to +\infty$, uniformly on every compact subset of $\mathbb{R}^n$. For example, when $\mu(x) \equiv 1$ (or any other constant) then $u_+(x) \equiv 1$, and this result says that $u(t,x) \to 1$ as $t \to +\infty$, uniformly on compact sets in x.

Finally, and this is the core of this chapter, we will prove the following propagation result. For each unit vector $e \in \mathbb{R}^n$, $|e| = 1$, consider the solution of the linear equation

$$v_t - \Delta v = \mu(x)v, \quad x \in \mathbb{R}^n, \tag{0.6}$$

of the form

$$v(t,x) = e^{-\lambda(x\cdot e - ct)}\phi(x), \tag{0.7}$$

with a positive 1-periodic function $\phi(x)$. Such exponential solutions are extremely important in the theory for the nonlinear problem. It will be not hard to see that for each direction $e \in \mathbb{S}^{n-1}$ they exist only for $c \geq c_*(e)$, where $c_*(e)$ is the smallest possible propagation speed of such exponential. If we set

$$w_*(e) = \inf_{|e'|=1,(e\cdot e')>0} \frac{c_*(e')}{(e \cdot e')}, \tag{0.8}$$

then the following holds for solutions of the nonlinear problem (0.1) with a nonnegative bounded and compactly supported initial data $u_0(x)$: for each $w \in (0, w_*(e))$ we have

$$\lim_{t\to+\infty} \sup_{r\in[0,w]} |u(t, rte) - u_+(rte)| = 0, \tag{0.9}$$

and for each $w \in (w_*(e), +\infty)$ we have

$$\lim_{t\to+\infty} \sup_{r\geq wt} u(t, re) = 0. \tag{0.10}$$

That is, if we observe the solution $u(t,x)$ along the ray in the direction e, $u(t,x)$ is close to $u_+(x)$ at distances much smaller than $w_*(e)t$ and $u(t,x)$ is close to zero at distances much larger than $w_*(e)t$. The remarkable fact is that the invasion speed $w_*(e)$ is completely determined by the linear problem (0.6)!

This propagation bound was discovered by Freidlin and Gärtner [66] who proved it with probabilistic tools. Since then its scope was considerably extended and at least four additional methods of proof are known:
(i) Probabilistic proofs using large deviation methods, due to Freidlin [67].
(ii) Viscosity solution methods (Evans and Souganidis [54, 55]).
(iii) Monotone dynamical systems methods (Weinberger [115]).
(iv) PDE methods (Berestycki, Hamel and Nadin [16]).

The goal of this chapter is to explain the Freidlin-Gärtner formula for the propagation speed as well as its applications to biology.

1 Origins of the model

The original motivation for considering the Fisher-KPP equation (0.1) (KPP stands for Kolmogorov, Petrovskii and Piskunov) in [63] and [82] was by problems in genetics, while Freidlin and Gärtner motivated their study of (0.1) as a model for concentration waves in a periodic medium. There is also a nice interpretation of this equation in terms of population dynamics. Let a population of animals, or bacteria, or even some flora be described in terms of its local density $u(t,x)$. That is, $u(t,x)dx$ is the number of individuals present at time t in an infinitesimal volume dx around a point x – the total number of individuals present in a given domain Ω at a time t is

$$\int_\Omega u(t,x)dx.$$

This description assumes implicitly that the number of individuals is large, or equivalently, they are not too sparse – one would not be able to describe the animals in a desert in this way. The individuals multiply and disappear. In other words, in the absence of a spatial displacement, the population density evolves as

$$\frac{du}{dt} = \mu(x)u - u^2 = (\mu(x) - u)u. \tag{1.1}$$

Here, x is the spatial position, and $\mu(x)$ is the local growth rate at x for small u. These equations are uncoupled at different points x. The negative term in the right side of (1.1) accounts for the fact that there are limited resources – too many individuals present at one point prevent population growth due to competition. The threshold value at which the growth becomes negative in this model is $u = \mu(x)$. Hence, $\mu(x)$ can be both interpreted as the growth rate for small u and as the carrying capacity of the population.

An aspect missing in (1.1) is movement of the individuals, displacements and migrations. Assume for the moment that there is no growth of the population but the species may disperse. If the chances of entering a small volume dx around x from position y are $k(x,y)$ then the balance equation for the population density is

$$\frac{\partial u(t,x)}{\partial t} = \int k(x,y)u(t,y)dy - \left(\int k(y,x)dy\right) u(t,x). \tag{1.2}$$

The first term on the right accounts for individuals entering the volume dx from all other positions y and the negative term accounts for those leaving dx. Assume now that the transition kernel $k(x,y)$ is localized and radially symmetric:

$$k(x,y) = \frac{1}{\varepsilon^n} r\left(\frac{|x-y|}{\varepsilon}\right),$$

and the mean drift is zero:

$$\int xr(x)dx = 0.$$

Then, expanding (1.2) in ε we obtain, in the leading order:

$$\frac{\partial u}{\partial t} = D\varepsilon^2 \Delta u, \tag{1.3}$$

with the diffusion coefficient

$$D = \int |x|^2 r(x) dx.$$

Exercise 1.1 This formal procedure is not difficult to make rigorous – this limit is, essentially, the (much simpler) PDE analog of the probabilistic result showing that a discrete time random walk on a lattice converges to a Brownian motion if we scale the lattice step and the time step appropriately. Make this connection in a careful fashion.

Putting (1.1) and (1.3) together (with the appropriate time rescaling in (1.3) to get rid of the ε^2 factor and setting $D = 1$) gives the Fisher-KPP equation

$$\frac{\partial u}{\partial t} = \Delta u + \mu(x)u - u^2, \tag{1.4}$$

that we will study in this chapter. A much more detailed explanation of the modeling issues is given in Murray's books [94, 95].

2 The steady solution as the long time limit for the Cauchy problem

It is reasonable to expect that if solutions of (1.4) converge as $t \to +\infty$ to a certain limit $p(x)$, this function should satisfy the steady problem[1]

$$\begin{aligned} &-\Delta p = \mu(x)p - p^2, \quad x \in \mathbb{R}^n, \\ &p(x) > 0 \text{ for all } x \in \mathbb{R}^n \text{ and } p(x) \text{ is bounded.} \end{aligned} \tag{2.1}$$

In this section we will investigate existence of such steady solutions. One of the main points here is that we impose neither periodicity nor any decay conditions on $p(x)$ as $|x| \to +\infty$, but only require that $p(x)$ is positive and bounded. Let us recall that we denote by λ_1 the principal eigenvalue of

$$\begin{aligned} &-\Delta\phi - \mu(x)\phi = \lambda_1\phi, \\ &\phi(x) \text{ is 1-periodic in all its variables, } \phi(x) > 0, \end{aligned} \tag{2.2}$$

and that the requirement that the eigenfunction $\phi(x)$ is positive identifies λ_1 uniquely. The next theorem explains the role of the principal eigenvalue rather succinctly.

Theorem 2.1 *The problem (2.1) has a unique solution if $\lambda_1 < 0$ and no solutions if $\lambda_1 \geq 0$.*

The existence part of Theorem 2.1 has been known for a long time now but the uniqueness part is recent [18]. This result is important for two reasons: (1) it classifies all solutions to the steady problem, and (2) is the key to understanding the long time behavior of the solutions the corresponding Cauchy problem, as shown by the following theorem.

[1]Another reasonable possibility is that the limit is a solution of the time-dependent problem that is defined for all times, positive and negative, of which a steady solution is just one example.

Theorem 2.2 *Let $u(t,x)$ be the solution of the initial value problem*

$$\frac{\partial u}{\partial t} = \Delta u + \mu(x)u - u^2, \quad t > 0, \ x \in \mathbb{R}^n, \qquad (2.3)$$
$$u(0,x) = u_0(x),$$

with a bounded non-negative function $u_0(x)$ such that $u_0(x) \not\equiv 0$, and let λ_1 be the principal eigenvalue of (2.2). Then if $\lambda_1 < 0$ we have

$$u(t,x) \to p(x) \text{ as } t \to +\infty, \qquad (2.4)$$

uniformly on compact sets $K \subset \mathbb{R}^n$. On the other hand, if $\lambda_1 \geq 0$ then

$$u(t,x) \to 0 \text{ as } t \to +\infty, \qquad (2.5)$$

uniformly in $\mathbb{R}^n$.

In the rest of this section we will prove these two theorems – the proof of Theorem 2.1, in particular, is not short but it utilizes various tools that are interesting in their own right.

Triviality of the steady solutions when $\lambda_1 \geq 0$

Let us first explain what happens if $\lambda_1 > 0$. Let $\phi(x)$ be the corresponding (periodic) eigenfunction of (2.2), and let $u(t,x)$ satisfy the time-dependent problem (2.3). As $\phi(x)$ is periodic, its minimum is positive. Hence, as $u_0(x)$ is bounded, we can find $M > 0$ so that at $t = 0$ we have

$$u(0,x) = u_0(x) \leq \sup_{x\in\mathbb{R}^n} u_0(x) \leq M \min_{x\in\mathbb{T}^n} \phi(x) \leq M\phi(x). \qquad (2.6)$$

The function

$$\psi(t,x) = Me^{-\lambda_1 t}\phi(x)$$

satisfies

$$\psi_t = \Delta\psi + \mu(x)\psi, \qquad (2.7)$$

which means that $\psi(t,x)$ is a super-solution to (2.3):

$$\psi_t > \Delta\psi + \mu(x)\psi - \psi^2. \qquad (2.8)$$

This, together with the inequality (2.6), by virtue of the parabolic maximum principle, implies that for all $t \geq 0$ we have

$$u(t,x) \leq \psi(t,x) = Me^{-\lambda_1 t}\phi(x) \leq M\|\phi\|_{L^\infty(\mathbb{T}^n)} e^{-\lambda_1 t}. \qquad (2.9)$$

It follows that

$$u(t,x) \to 0 \text{ as } t \to +\infty, \qquad (2.10)$$

uniformly in $\mathbb{R}^n$, and, in particular, precludes the existence of non-trivial bounded solutions to (2.1).

The argument in the case $\lambda_1 = 0$ is similar albeit with a nice additional step. In this situation, the eigenfunction is a periodic function $\phi(x) > 0$ such that

$$-\Delta\phi = \mu(x)\phi. \tag{2.11}$$

By the same token as before, we know that for any solution of (2.3) with a bounded initial data $u_0(x) \geq 0$ we can find a constant $M > 0$ so that

$$u(0, x) \leq M\phi(x). \tag{2.12}$$

The parabolic maximum principle[2] implies that this inequality holds for all $t \geq 0$:

$$u(t, x) \leq M\phi(x), \text{ for all } x \in \mathbb{R}^n.$$

Let now M_k be the smallest constant M so that we have

$$u(k, x) \leq M\phi(x) \text{ for all } x \in \mathbb{R}^n, \tag{2.13}$$

at the time $t = k$. The sequence M_k is non-increasing: since $M_k\phi(x)$ is a super-solution, (2.13) together with the strong maximum principle guarantees that (with the strict inequality)

$$u(k+1, x) < M_k\phi(x), \tag{2.14}$$

which implies that $M_{k+1} \leq M_k$. Let us now show that the strong maximum principle implies that this inequality is strict: $M_{k+1} < M_k$. It suffices to verify this for $k = 1$: assume that $M_2 = M_1$. Then there exists a sequence x_k such that

$$u(2, x_k) \geq \left(M_1 - \frac{1}{k}\right)\phi(x_k). \tag{2.15}$$

Let us define the translates

$$v_k(t, x) = u(t, x + x_k), \quad \phi_k(x) = \phi(x + x_k).$$

The parabolic regularity theory implies that the shifted functions $v_k(t, x)$ and $\phi_k(x)$ are uniformly bounded in $C^{2,\alpha}_{loc}$ for $1 \leq t \leq 2$, hence we may extract a subsequence $k_n \to +\infty$ so that the limits

$$\bar{v}(t, x) = \lim_{n\to+\infty} v_{k_n}(t, x), \quad \bar{\phi}(x) = \lim_{n\to+\infty} \phi_{k_n}(x)$$

exist. The shifted coefficients $\mu_k(x) = \mu(x + x_k)$ also converge after extracting a subsequence, locally uniformly to a limit $\bar{\mu}(x)$. The limits satisfy

$$\frac{\partial\bar{v}}{\partial t} = \Delta\bar{v} + \bar{\mu}(x)\bar{v} - \bar{v}^2, \quad 1 \leq t \leq 2, \ x \in \mathbb{R}^n, \tag{2.16}$$

and

$$-\Delta\bar{\phi} = \bar{\mu}(x)\bar{\phi}. \tag{2.17}$$

[2]Recall that $\phi(x)$ is a super-solution to the problem (2.3) that $u(t, x)$ satisfies.

In addition, we have $\bar{v}(t=1,x) \le M_1\bar{\phi}(x)$ for all $x \in \mathbb{R}^n$, and $\bar{v}(t=2,x=0) = M_1\bar{\phi}(0)$. This contradicts the strong maximum principle since $\bar{\phi}$ is a strict super-solution to (2.16). Therefore, the sequence M_n is strictly decreasing.

Let now

$$\bar{M} = \lim_{k\to+\infty} M_k. \tag{2.18}$$

We need to show that $\bar{M} = 0$, in order to conclude that $u(t,x) \to 0$ as $t \to +\infty$, uniformly in $x \in \mathbb{R}^n$. As in the previous step, choose x_k so that

$$v(k,x_k) \ge (M_k - \frac{1}{k})\phi(x_k),$$

and define the translates

$$v_k(t,x) = v(k+t, x_k + x), \quad \phi_k(x) = \phi(x + x_k), \tag{2.19}$$

as well as $\mu_k(x) = \mu(x + x_k)$. Once again, the parabolic regularity theory implies that the sequences $v_k(t,x)$, $\phi_k(x)$ and $\mu_k(x) = \mu(x+x_k)$ (after extraction of a subsequence) converge as $k \to +\infty$, locally uniformly, to the respective limits $\bar{v}(t,x)$, $\bar{\phi}(x)$ and $\bar{\mu}(x)$ that satisfy, in this case,

$$\frac{\partial \bar{v}}{\partial t} = \Delta\bar{v} + \bar{\mu}(x)\bar{v} - \bar{v}^2, \quad -\infty < t < +\infty,\ x \in \mathbb{R}^n, \tag{2.20}$$

and

$$-\Delta\bar{\phi} = \bar{\mu}(x)\bar{\phi}. \tag{2.21}$$

That is, $\bar{v}(t,x)$ is a global in time solution, defined for positive and negative t. In addition, the normalization (2.19) implies that

$$\bar{v}(0,0) = \bar{M}\bar{\phi}(0), \tag{2.22}$$

while we also have

$$\bar{v}(t,x) \le \bar{M}\bar{\phi}(x), \quad -\infty < t < +\infty, \quad x \in \mathbb{R}^n. \tag{2.23}$$

The parabolic strong maximum principle implies that then $\bar{v}(t,x) \equiv \bar{M}\bar{\phi}(x)$ which is only possible if $\bar{M} = 0$. Therefore, $\bar{M} = 0$, and

$$u(t,x) \to 0 \text{ as } t \to +\infty, \text{ uniformly in } x \in \mathbb{R}^n,$$

also when $\lambda_1 = 0$.

Existence of the periodic steady solutions when $\lambda_1 < 0$

We now turn to the most interesting case $\lambda_1 < 0$. We need to show that then a non-trivial steady solution $p(x)$ of (2.1) exists, and, moreover, solution of the parabolic problem converges to it as $t \to +\infty$, locally uniformly in x.

Let $\phi(x)$ be the positive periodic eigenfunction of

$$-\Delta\phi - \mu(x)\phi = \lambda_1\phi. \tag{2.24}$$

Consider the function $\phi_\varepsilon(x) = \varepsilon\phi(x)$. A simple but very important observation is that for $\varepsilon > 0$ sufficiently small we have

$$-\Delta\phi_\varepsilon - \mu(x)\phi_\varepsilon = \lambda_1\phi_\varepsilon \le -\phi_\varepsilon^2, \tag{2.25}$$

that is, $\phi_\varepsilon(x)$ is a sub-solution for the steady nonlinear problem. More precisely, this inequality holds as soon as

$$\varepsilon < -\frac{\lambda_1}{\max_{x\in\mathbb{T}^n}\phi(x)}, \tag{2.26}$$

and it is here that we need the assumption $\lambda_1 < 0$. On the other hand, the constant function $w(x) \equiv M$ satisfies

$$-\Delta w - \mu(x)w = -\mu(x)M \ge -M^2, \tag{2.27}$$

as soon as

$$M \ge \max_{x\in\mathbb{T}^n}\mu(x). \tag{2.28}$$

Therefore, we have both a sub-solution $\phi_\varepsilon(x)$ (with an ε that satisfies (2.26)) and a super-solution $w(x)$ (with M that satisfies (2.28)) for the steady problem (2.1). With these in hand, a true solution of (2.1) can be constructed using a standard iteration scheme. First, choose a number $N > -2\lambda_1$ and restate (2.1) as

$$-\Delta p(x) - \mu(x)p(x) + Np(x) = Np(x) - p^2. \tag{2.29}$$

The reason to add the term $Np(x)$ on the left is to make sure that all eigenvalues of the periodic problem

$$-\Delta\phi - \mu(x)\phi + N\phi(x) = \lambda\phi, \tag{2.30}$$

are strictly positive. In this case, the inhomogeneous elliptic problem

$$-\Delta\phi - \mu(x)\phi + N\phi(x) = f(x) \tag{2.31}$$

has a unique periodic solution $p(x)$ for any bounded periodic function $f(x)$. Moreover, Theorem 2.1 says that if $f(x) > 0$ for all $x \in \mathbb{T}^n$ then the solution of (2.31) is also positive.

We set up the iteration scheme as follows: let $p_0 = \phi_\varepsilon(x)$ and for $k \ge 1$ let $p_k(x)$ be the periodic solution of

$$-\Delta p_k - \mu(x)p_k + Np_k(x) = Np_{k-1}(x) - p_{k-1}^2(x). \tag{2.32}$$

We claim that the sequence $p_k(x)$ is increasing pointwise in x:

$$p_{k+1}(x) \ge p_k(x), \text{ for all } k \ge 0 \text{ and all } x \in \mathbb{T}^n, \tag{2.33}$$

and satisfies

$$p_k(x) \le \frac{N}{2} \text{ for all } k \ge 0 \text{ and all } x \in \mathbb{T}^n. \tag{2.34}$$

In order to prove the upper bound (2.34) we observe that $p_0(x) \le N/2$ if ε is sufficiently small, and then use induction: define $w_k(x) = N/2 - p_k(x)$, assume that $p_{k-1}(x) \le N/2$ for all $x \in \mathbb{T}^n$, and write

$$\begin{aligned}
-\Delta w_k - \mu(x) w_k + N w_k &= -\mu(x)\frac{N}{2} + \frac{N^2}{2} + \Delta p_k + \mu(x) p_k - N p_k \\
&= -\mu(x)\frac{N}{2} + \frac{N^2}{2} - N p_{k-1} + p_{k-1}^2 \ge -\bar{\mu}\frac{N}{2} + \frac{N^2}{2} - \frac{N^2}{4} > 0,
\end{aligned}$$

as long as $N > 2\bar{\mu}$. This proves that $w_k(x) > 0$, hence (2.34) holds. The reason for the pointwise monotonicity of the sequence $p_k(x)$ is that p_0 is a sub-solution for (2.30). The proof is by induction: set

$$z_k(x) = p_k(x) - p_{k-1}(x), \quad k \ge 1,$$

then z_1 satisfies

$$\begin{aligned}
-\Delta z_1 - \mu(x) z_1 + N z_1 &= -\Delta p_1 - \mu(x) p_1 + N p_1 + \Delta p_0 + \mu(x) p_0 - N p_0 \qquad (2.35)\\
&= N p_0 - p_0^2 - \lambda_1 p_0 - N p_0 = -\lambda_1 p_0 - p_0^2 > 0.
\end{aligned}$$

The last inequality above holds by virtue of (2.26), and, once again, requires that $\lambda_1 < 0$. Now, Theorem 2.1 implies that $z_1 \ge 0$ – as discussed above, just below (2.31). Next, assume that $z_j(x) \ge 0$ for all $x \in \mathbb{T}^n$ and all $j = 1, \ldots, k$. The function $z_{k+1}(x)$ satisfies

$$\begin{aligned}
-\Delta z_{k+1} - \mu(x) z_{k+1} + N z_{k+1} &= -\Delta p_{k+1} - \mu(x) p_{k+1} + N p_{k+1} + \Delta p_k + \mu(x) p_k - N p_k \\
&= N p_k - p_k^2 - N p_{k-1} + p_{k-1}^2 = N z_k - (p_{k-1} + p_k) z_k > 0. \qquad (2.36)
\end{aligned}$$

We used the induction assumption $z_k \ge 0$ and the upper bound (2.34) in the last step. Once again, Theorem 2.1 implies that $z_{k+1}(x) \ge 0$ for all $x \in \mathbb{T}^n$. Thus, the sequence $p_k(x)$ is, indeed, increasing. Therefore, the sequence $p_k(x)$ converges pointwise in x to a limit profile $p(x)$ that satisfies

$$\phi_\varepsilon(x) \le p(x) \le \frac{N}{2}, \qquad (2.37)$$

and

$$-\Delta p - \mu(x) p + N p = N p - p^2, \qquad (2.38)$$

which is nothing but (2.1). Condition (2.37) is very important – it ensures that $p(x) \not\equiv 0$. We have, thus, established that when $\lambda_1 < 0$ this equation has a non-trivial steady periodic solution, finishing the proof of the existence part of Theorem 2.1.

Uniqueness of a bounded solution when $\lambda_1 < 0$

Next, we show that the periodic solution of (2.1) that we have just constructed is unique in the class of bounded solutions. That is, if $s(x)$ is another bounded (not necessarily periodic) solution of

$$\begin{aligned}
&-\Delta s = \mu(x) s - s^2 \qquad (2.39)\\
&s(x) \text{ is bounded, and } s(x) > 0 \text{ for all } x \in \mathbb{R}^n,
\end{aligned}$$

then $s(x)$ coincides with the periodic solution $p(x)$ that we have constructed above. The crucial part in the proof of uniqueness is played by the following lemma.

Lemma 2.3 *Any solution of (2.39) is bounded from below by a positive constant:*

$$\inf_{x\in\mathbb{R}^n} s(x) > 0. \tag{2.40}$$

Let us first explain why uniqueness of the solution of (2.39) follows from this lemma. Let $p(x)$ and $s(x)$ be two solutions. Lemma 2.3 allows us to define r_0 as the smallest r such that $s(x) \le rp(x)$:

$$r_0 = \inf\{r : \quad s(x) \le rp(x), \quad \text{for all } x \in \mathbb{R}^n\}.$$

We claim that $r_0 \le 1$. Indeed, the difference

$$v(x) = r_0 p(x) - s(x)$$

satisfies

$$-\Delta v - \mu(x)v = -r_0 p^2 + s^2,$$

and a simple computation shows that

$$\begin{aligned}&-\Delta v + (-\mu(x) + r_0 p(x) + s(x))v = r_0 pv + sv - r_0 p^2 + s^2\\ &= r_0 p(r_0 p - s) + s(r_0 p - s) - r_0 p^2 + s^2 = r_0(r_0 - 1)p^2(x).\end{aligned}$$

Therefore, if $r_0 > 1$ the function $v(x)$ satisfies

$$\begin{aligned}&-\Delta v + (-\mu(x) + r_0 p(x) + s(x))v = r_0(r_0 - 1)p(x) > c_0 = r_0(r_0 - 1) \inf_{x\in\mathbb{R}^n} p(x) > 0,\\ &\quad v(x) \ge 0 \text{ for all } x \in \mathbb{R}^n.\end{aligned} \tag{2.41}$$

As $v(x) \ge 0$, the strong maximum principle implies that $v(x) > 0$ for all $x \in \mathbb{R}^n$. Furthermore, if there is a sequence x_k such that $|x_k| \to +\infty$ such and

$$\lim_{k\to\infty} v(x_k) = 0,$$

this is also a contradiction to the strong maximum principle. Indeed, as we have seen several times before, the elliptic regularity theory implies that we may extract a subsequence $n_k \to +\infty$ so that the shifted functions $v_k(x) = v(x_k + x)$, $p_k(x) = p(x + x_k)$, $s_k(x) = s(x + x_k)$, and $\mu_k(x) = \mu(x + x_k)$ converge to the respective limits $\bar{v}(x)$, $\bar{p}(x)$, $\bar{s}(x)$ and $\bar{\mu}(x)$ that satisfy

$$\begin{aligned}&-\Delta\bar{v} + (-\bar{\mu}(x) + r_0\bar{p}(x) + \bar{s}(x))\bar{v} > c_0 > 0,\\ &\quad v(x) \ge 0 \text{ for all } x \in \mathbb{R}^n,\end{aligned} \tag{2.42}$$

with $\bar{v}(0) = 0$, which is impossible.

We conclude that $r_0 \le 1$, meaning that $s(x) \le p(x)$. The only property of the solution $p(x)$ we have used above is that there exist two constants $c_{1,2} > 0$ so that

$$0 < c_1 < p(x) < c_2 < +\infty \text{ for all } x \in \mathbb{R}^n.$$

Lemma 2.3 asserts that "the other" solution $s(x)$ obeys same bounds (with different constants $c_{1,2}$). Hence, an identical argument implies that $p(x) \le s(x)$, and it follows that $p(x) = s(x)$ establishing uniqueness of the solutions of (2.39).

The uniform lower bound: the proof of Lemma 2.3

We now prove Lemma 2.3, the last ingredient in the proof of Theorem 2.1. An immediate trivial observation is that if $s(x)$ is a periodic solution of

$$\begin{aligned} &-\Delta s = \mu(x)s - s^2 \\ &s(x) \text{ is bounded and } s(x) > 0 \text{ for all } x \in \mathbb{R}^n, \end{aligned} \tag{2.43}$$

then, of course,

$$\inf_{x\in\mathbb{R}^n} s(x) > 0. \tag{2.44}$$

The main difficulty is, therefore, in dealing with general bounded solutions, that need not be periodic. To this end, we would like to get a nice subsolution for (2.43) that we would be able to put under $s(x)$ to give a lower bound for $s(x)$. As in the proof of existence of a solution to (2.43), a good candidate is $\phi_\varepsilon(x) = \varepsilon\phi(x)$, where $\phi(x)$ is the principal periodic eigenfunction of

$$\begin{aligned} &-\Delta\phi - \mu(x)\phi = \lambda_1\phi, \\ &\phi(x) > 0 \text{ for all } x \in \mathbb{T}^n. \end{aligned} \tag{2.45}$$

Recall that the function $\phi_\varepsilon(x)$ satisfies

$$-\Delta\phi_\varepsilon - \mu(x)\phi_\varepsilon + \phi_\varepsilon^2 = \lambda_1\phi_\varepsilon + \phi_\varepsilon^2 < 0, \tag{2.46}$$

provided that (compare to (2.26))

$$\varepsilon < -\frac{\lambda_1}{\max_{x\in\mathbb{T}^n}\phi(x)}. \tag{2.47}$$

The difficulty in using this subsolution is that it is periodic – how can we put it under $s(x)$ unless we know that $s(x)$ is uniformly positive? Instead, we are going to use the principal Dirichlet eigenfunction in a ball $B(m,R)$ where $m \in \mathbb{Z}^d$ is an integer point, and R is sufficiently large. Its advantage is that this eigenfunction is compactly supported so that a sufficiently small multiple of it can be put under any positive function. Let λ_R be the principal Dirichlet eigenvalue in such ball. It does not depend on m since the coefficient $\mu(x)$ is periodic, hence we set $m = 0$ for the moment:

$$\begin{aligned} &-\Delta\psi_R(x) - \mu(x)\psi_R = \lambda_R\psi_R(x), \quad |x| < R, \\ &\psi_R(x) > 0 \text{ for } |x| < R, \\ &\psi_R(x) = 0 \text{ on } \{|x| = R\}. \end{aligned} \tag{2.48}$$

This is where we use Proposition 3.1, that we restate as a

Lemma 2.4 *Let λ_1 be the principal periodic eigenvalue of the problem (2.45), and λ_R be the principal Dirichlet eigenvalue of the problem (2.48), then*

$$\lim_{R\to+\infty} \lambda_R = \lambda_1. \tag{2.49}$$

Here is how we use it. Let ψ_R be the eigenfunction of (2.48) normalized so that

$$\sup_{|x|\le R} \psi_R(x) = 1.$$

Set $\phi_{\varepsilon,R} = \varepsilon\psi_R(x)$, then, as in (2.46) we have

$$-\Delta\phi_{\varepsilon,R} - \mu(x)\phi_{\varepsilon,R} + \phi_{\varepsilon,R}^2 < 0, \tag{2.50}$$

as long as

$$\varepsilon < -\lambda_R. \tag{2.51}$$

Lemma 2.4 implies that there exists R so that for all $R' > R$ we have $\lambda_{R'} < -\lambda_1/2$. Therefore, the function $\phi_{\varepsilon,R}$ is a sub-solution to (2.46) on the ball $B(0,R)$ for any $\varepsilon < -\lambda_1/2$.

In addition, we know that for ε sufficiently small we have $\phi_{\varepsilon,R} < p(x)$ for all $x \in B(0,R)$ simply because $p(x) > 0$ for all $x \in \mathbb{R}^n$. Let us now start increasing ε until $p(x)$ and $\phi_{\varepsilon,R}$ touch:

$$\varepsilon_0 = \sup\{\varepsilon > 0 :\ p(x) \ge \varepsilon\phi_R(x) \text{ for all } x \in B(0,R)\}.$$

We claim that $\varepsilon_0 \ge -\lambda_R$. Indeed, otherwise $\phi_{\varepsilon,R}$ is a sub-solution and $p(x)$ is a solution, hence they can not touch without violating the maximum principle. Thus, we have $\varepsilon_0 \ge -\lambda_R$, that is, ε_0 is sufficiently large so that $\phi_{\varepsilon_0,R}$ is no longer a sub-solution. Therefore, we have shown that

$$p(x) \ge \left(-\frac{\lambda_1}{2}\right)\phi_R(x) \text{ for all } x \in B(0,R). \tag{2.52}$$

By considering a shifted ball $B(m,R)$ we see that, actually, we have a generalization of (2.52):

$$p(x) \ge \left(-\frac{\lambda_1}{2}\right)\phi_R(x-m) \text{ for all } x \in B(m,R), \text{ and all } m \in \mathbb{Z}^n. \tag{2.53}$$

It follows immediately that there exists a constant $c_0 > 0$ so that $p(x) > c_0$ for all $x \in \mathbb{R}^n$. Note that we may only shift ϕ_R by an integer m - otherwise it would cease being a solution since $\mu(x)$ is not a constant. Therefore, the proof of Lemma 2.3 is complete, as well as that f of Theorem 2.1.

Convergence of the solutions of the Cauchy problem

We now prove Theorem 2.2. Recall that we need to prove that if $\lambda_1 < 0$ then solutions of the Cauchy problem

$$\begin{aligned} &\frac{\partial u}{\partial t} = \Delta u + \mu(x)u - u^2, \quad t > 0,\ x \in \mathbb{R}^n, \\ &u(0,x) = u_0(x), \end{aligned} \tag{2.54}$$

with a bounded non-negative function $u_0(x)$ such that $u_0(x) \not\equiv 0$, have the long time limit

$$u(t,x) \to p(x) \text{ as } t \to +\infty, \tag{2.55}$$

uniformly on compact sets $K \subset \mathbb{R}^n$. Here, as before, $p(x)$ is the unique bounded positive solution of the steady problem

$$-\Delta p = \mu(x)p - p^2, \quad x \in \mathbb{R}^n.$$

Recall that we have already shown that if $\lambda_1 \geq 0$ then

$$u(t,x) \to 0 \text{ as } t \to +\infty, \tag{2.56}$$

uniformly in $\mathbb{R}^n$.

Let us first show that

$$\liminf_{t\to+\infty} u(t,x) \geq p(x). \tag{2.57}$$

To this end, we will use the sub-solution $\varepsilon\phi_R(x)$ we used in the proof of Lemma 2.3. First, we wait until time $t = 1$ to make sure that $u(t = 1, x) > 0$ in $\mathbb{R}^n$. Then, we may find $\varepsilon > 0$ sufficiently small, and R sufficiently large, so that $\phi_\varepsilon(x) = \varepsilon\phi_R(x)$ is a sub-solution:

$$-\Delta\phi_\varepsilon \leq \mu(x)\phi_\varepsilon - \phi_\varepsilon^2,$$

and $\phi_\varepsilon(x) < p(x)$ for all $x \in \mathbb{R}^n$ (we extend $\phi_\varepsilon(x) = 0$ outside the ball $B(0, R)$). We also take ε so small that $u(t = 1, x) > \phi_\varepsilon(x)$ for all $x \in \mathbb{R}^n$. Let now $v(t, x)$ be the solution of the Cauchy problem

$$\begin{aligned} &\frac{\partial v}{\partial t} = \Delta v + \mu(x)v - v^2, \quad t > 1, \ x \in \mathbb{R}^n, \\ &v(t = 1, x) = \phi_\varepsilon(x). \end{aligned} \tag{2.58}$$

The parabolic comparison principle implies immediately that $v(t, x) \leq u(t, x)$ for all $t > 1$.

Exercise 2.5 Use the fact that $\phi_\varepsilon(x)$ (which is the initial data for $v(t, x)$), is a sub-solution, to show that $v(t, x) \geq \phi_\varepsilon(x)$ for all $t \geq 1$ and $x \in \mathbb{R}^n$.

Exercise 2.6 Use the result of the previous exercise to show that $v(t, x)$ is strictly increasing in time. Hint: set, for all $h > 0$,

$$v_h(t, x) = v(t + h, x) - v(t, x),$$

and verify that $v_h(t, x)$ satisfies

$$\frac{\partial v_h}{\partial t} = \Delta v_h + \mu(x)v_h - (v(t + h, x) + v(t, x))v_h,$$

with $v_h(t = 1, x) \geq 0$ for all $x \in \mathbb{R}^n$. Use the parabolic comparison principle to deduce that $v_h(t, x) \geq 0$ for all $t \geq 1$, that is, the function $v(t, x)$ is monotonically increasing in t.

Exercise 2.7 Use the fact that $p(x)$ is a solution, while $\phi_\varepsilon(x)$ is a sub-solution to show that $v(t, x) \leq p(x)$ for all $t \geq 1$, if $\varepsilon > 0$ is sufficiently small.

A consequence of the above observations is that the limit

$$s(x) = \lim_{t\to+\infty} v(t,x)$$

exists and is a positive bounded steady solution:

$$-\Delta s = \mu(x)s - s^2.$$

Uniqueness of such solutions implies that $s(x) = p(x)$, and thus

$$\liminf_{t\to+\infty} u(t,x) \geq \lim_{t\to+\infty} v(t,x) = p(x), \tag{2.59}$$

as we have claimed. Moreover, if

$$u_0(x) \leq p(x) \text{ for all } x \in \mathbb{R}^n, \tag{2.60}$$

then by the same token we have $u(t,x) \leq p(x)$ for all $t \geq 0$, meaning that

$$\lim_{t\to+\infty} u(t,x) = p(x).$$

Let us finally see what happens if (2.60) does not hold. If we multiply $p(x)$ by a number $M > 1$ and set $p_M(x) = Mp(x)$, we get a super-solution:

$$-\Delta p_M - \mu(x)p_M + p_M^2 = -M\Delta p - M\mu(x)p + M^2p^2 = -Mp^2 + M^2p^2 > 0, \tag{2.61}$$

as $M > 1$. If we choose $M > 1$ sufficiently large so that $u_0(x) \leq p_M(x)$ then $u(t,x) \leq w(t,x)$, solution of

$$\begin{aligned} &\frac{\partial w}{\partial t} = \Delta w + \mu(x)w - w^2, \quad t > 0, \ x \in \mathbb{R}^n, \\ &w(t,x) = Mp(x). \end{aligned} \tag{2.62}$$

As $p_M(x)$ is a super-solution, the argument we used to show that $v(t,x)$ was increasing in time, shows that $w(t,x)$ is monotonically decreasing in time. In addition, as $M > 1$, we know from the comparison principle that $w(t,x) \geq p(x)$ for all $t > 0$. Its point-wise limit (as $t \to +\infty$) is therefore a non-trivial steady solution of our problem and thus equals to $p(x)$:

$$\lim_{t\to+\infty} w(t,x) = p(x).$$

As a consequence, we obtain

$$\limsup_{t\to+\infty} u(t,x) \leq p(x). \tag{2.63}$$

This, together with (2.59) proves that

$$\lim_{t\to+\infty} u(t,x) = p(x),$$

and the proof of Theorem 2.2 is complete.

3 The speed of invasion

We now tackle the heart of this chapter: finding the speed of invasion of the stable steady state $p(x)$ – in this section we assume that $\mu(x)$ is such that $\lambda_1 < 0$ so that the steady state does exist.

3.1 The homogeneous case

We first consider the uniform case $\mu(x) \equiv 1$, where the proof is much simpler, especially if we replace the nonlinearity $u - u^2$ by a function $f(u)$ which is linear close to zero:

$$f(u) = \begin{cases} u & \text{if } u \leq \theta, \\ u - u^2 & \text{if } u \text{ is close to 1.} \end{cases} \tag{3.1}$$

We also assume that $f(u)$ is smooth, and $f(u) \leq u$ for all $u \in [0,1]$ – this is a crucial assumption. Thus, we momentarily consider the problem

$$u_t = u_{xx} + f(u), \quad t > 0, x \in \mathbb{R}, \tag{3.2}$$

with a nonnegative initial condition $u(0,x) = u_0(x) \not\equiv 0$, and $f(u)$ as above. The unique stable steady state is $p(x) \equiv 1$, and we are interested in how fast it invades the areas where u is small at $t = 0$.

An upper bound for the spreading speed

The function $u(t,x)$ satisfies the inequality

$$u_t - u_{xx} \leq u. \tag{3.3}$$

Let us look for exponential super-solutions to (3.2) of the form

$$\bar{u}(t,x) = e^{-\lambda(x-ct)}.$$

Because of (3.3), the function $\bar{u}(t,x)$ is a super-solution if

$$\lambda^2 - c\lambda + 1 = 0. \tag{3.4}$$

As we need $\bar{u}(t,x)$ to be real, (3.4) means that we have to take $c \geq 2$, and that for $c = 2$ we can take $\lambda = 1$. We conclude that if the initial $u_0(x)$ satisfies

$$u(x) \leq Me^{-|x|}, \tag{3.5}$$

then $u(t,x)$ satisfies

$$u(t,x) \leq M \min\left(e^{-(x-2t)}, e^{x+2t}\right), \tag{3.6}$$

whence

$$\lim_{t\to+\infty} \sup_{|x|\geq ct} u(t,x) = 0, \tag{3.7}$$

for all $c > 2$. Therefore, the steady state $u \equiv 1$ can not invade with a speed larger than $c_* = 2$.

A lower bound for the spreading speed

Next, we show that the state $u \equiv 1$ invades with the speed at least equal to $c_* = 2$ (or, rather, faster than any speed smaller than c_*), matching the upper bound for the invasion speed. It will be slightly easier to devise the lower bound in a moving frame. Let us take some $0 < c < 2$ and write $v(t,x) = u(t, x + ct)$, so that

$$v_t - cv_y = v_{yy} + f(v). \tag{3.8}$$

Because of the simplifying assumption (3.1) on the nonlinearity, any function $\underline{u}(t,x)$ such that

$$\underline{u}_t - c\underline{u}_y \le \underline{u}_{yy} + \underline{u}, \tag{3.9}$$

and such that $\underline{u}(t,y) \le \theta$ for all $t > 0$ and $x \in \mathbb{R}$ is a sub-solution to (3.8). We consider a time-independent exponential sub-solution

$$\underline{u}(y) = e^{-\lambda y},$$

but, as we take $c < 2$, the number λ, which satisfies (3.4), will have to be complex. In order to keep the sub-solution real, we set, for $t > 1$:

$$\underline{u}(y) = \begin{cases} m \exp\{-\text{Re } \lambda x\} \cos(\text{Im } \lambda y) \text{ if } |y| \le \pi/(2\text{Im } \lambda), \\ \quad\quad\quad\quad 0 \text{ otherwise.} \end{cases} \tag{3.10}$$

The constant $m > 0$ is chosen so that $\underline{u}(y) \le \theta$, and, in addition, $\underline{u}(y) \le u_0(y)$ – we assume here that $u_0(y) > 0$ on the interval $[-\pi/(2\text{Im } \lambda), \pi/(2\text{Im } \lambda)]$, otherwise we may simply wait until time $t = 1$, when $v(t = 1, y) > 0$ for all $y \in \mathbb{R}$, and put a small multiple of $\underline{u}(y)$ below $v(t = 1, y)$. We conclude that $v(t,y) \ge \underline{u}(y)$, for all $t > 0$. As a consequence, we immediately obtain that

$$\limsup_{t\to+\infty} u(t,ct) \ge m, \text{ for all } 0 \le c < 2. \tag{3.11}$$

Exercise 3.1 *Use the function $\underline{u}(y)$ as the initial data for the Cauchy problem in the moving frame to bootstrap the above argument to*

$$\liminf_{t\to+\infty} u(t,ct) = 1, \tag{3.12}$$

for all $0 \le c < 2$. Hint: such solution will be monotonically increasing in time.

The above arguments, hopefully, convince the reader that with a little bit of simplification, the speed of invasion can be found very easily. In the remainder of this section we will drop the simplifying assumptions about the nonlinearity that we have used here – that will create some technical difficulties but not change the moral of the story.

3.2 The exponential solutions

As in the homogeneous case considered in the previous section, exponential solutions of the linearized problem play a crucial role in the general periodic case. These are solutions of the equation

$$v_t = \Delta v + \mu(x)v, \quad x \in \mathbb{R}^n, \tag{3.13}$$

of the form

$$v(t,x) = e^{-\lambda(x\cdot e - ct)}\phi(x), \tag{3.14}$$

with a fixed unit vector $e \in \mathbb{R}^n$, $|e| = 1$, and a 1-periodic (in all directions) function $\phi(x)$. As we will use $v(t,x)$ as a super-solution, we will require that $\phi(x) > 0$. It will be convenient to factor $\phi(x) = \bar\phi(x)\Phi(x)$. Here, $\bar\phi(x)$ is the principal (positive) periodic eigenfunction of the problem we have encountered before:

$$\begin{aligned} &-\Delta\bar\phi - \mu(x)\bar\phi = \lambda_1\bar\phi \\ &\bar\phi(x) \text{ is 1-periodic,} \\ &\bar\phi(x) > 0 \text{ for all } x \in \mathbb{R}^n. \end{aligned} \tag{3.15}$$

Recall that our main assumption in this section is that $\lambda_1 < 0$. The function $\Phi(x)$ is the solution of

$$\begin{aligned} &\widetilde{L}_\lambda\Phi = -(\lambda_1 + c\lambda)\Phi \\ &\Phi(x) \text{ is 1-periodic,} \\ &\Phi(x) > 0 \text{ for all } x \in \mathbb{R}^n, \end{aligned} \tag{3.16}$$

with the operator $\widetilde{L}_\lambda$ given by

$$\widetilde{L}_\lambda\Phi = -e^{\lambda x\cdot e}\left[\Delta(e^{-\lambda x\cdot e}\Phi) - 2\frac{\nabla\bar\phi}{\bar\phi}\cdot\nabla(e^{-\lambda x\cdot e}\Phi)\right]. \tag{3.17}$$

Therefore, the speed $c \in \mathbb{R}$ of an exponential solution and its decay rate λ are related by the equation

$$c\lambda = -\lambda_1 - \mu_1^{per}(\widetilde{L}_\lambda). \tag{3.18}$$

Here $\mu_1^{per}(\widetilde{L}_\lambda)$ is the principal periodic eigenvalue of the operator $\widetilde{L}_\lambda$, and, as such, is a function of λ. The main result of this section is the following.

Theorem 3.2 *For every $e \in \mathbb{R}^n$, with $|e| = 1$ there exists $c_*(e) > 0$ so that (i) if $c < c_*(e)$, equation (3.18) has no solution $\lambda > 0$, (ii) if $c > c_*(e)$, equation (3.18) has two solutions $\lambda > 0$, and (iii) if $c = c_*(e)$, equation (3.18) has exactly one solution $\lambda > 0$.*

The key step in the proof of Theorem 3.2 is the next observation.

Lemma 3.3 *The function $\mu_1^{per}(\widetilde{L}_\lambda)$ is concave in λ.*

Let us step back and see what this result means in dimension $n = 1$ and when $\mu(x) \equiv 1$. Then $\lambda_1 = -1$, and both $\bar\phi(x) \equiv 1$ and $\Phi(x) \equiv 1$, while $\mu_1(\widetilde{L}_\lambda) = -\lambda^2$, so that (3.18) is simply

$$c\lambda = 1 + \lambda^2.$$

We see that in this special case the claim of Theorem 3.2 is true with $c_* = 2$, and that $\mu_1(\widetilde{L}_\lambda) = -\lambda^2$ is, indeed, concave in λ. In the general case, the key to the proof of Lemma 3.3 is the following observation: set

$$E_\lambda = \{\psi \in C^2(\mathbb{R}^n) :\ e^{\lambda x\cdot e}\psi(x) \text{ is 1-periodic, } \psi(x) > 0 \text{ for all } x \in \mathbb{R}^n\}, \tag{3.19}$$

then $\mu_1^{per}(\widetilde{L}_\lambda)$ has the min-max characterization

$$k(\lambda) := \mu_1^{per}(\widetilde{L}_\lambda) = \max_{\psi\in E_\lambda} \inf_{x\in\mathbb{R}^n} \frac{\mathcal{L}\psi(x)}{\psi(x)}. \tag{3.20}$$

Here we have denoted

$$\mathcal{L}\psi = -\Delta\psi - 2\frac{\nabla\bar{\phi}}{\bar{\phi}} \cdot \nabla\psi.$$

With the above notation, we need to prove that for any $t \in [0,1]$ we have

$$tk(\lambda_1) + (1-t)k(\lambda_2) \le k(t\lambda_1 + (1-t)\lambda_2), \tag{3.21}$$

for all $\lambda_1, \lambda_2 > 0$. Let ϕ_1 and ϕ_2 be the principal eigenfunctions of the operators $\widetilde{L}_{\lambda_1}$ and $\widetilde{L}_{\lambda_2}$, respectively, and set

$$\psi_i(x) = e^{-\lambda_i x\cdot e}\phi_i(x), \quad i = 1,2, \quad \psi(x) = \psi_1^t(x)\psi_2^{1-t}(x).$$

Note that $\psi \in E_\lambda$, with $\lambda = t\lambda_1 + (1-t)\lambda_2$ and hence can be used as a test function in the max-min principle for $k(\lambda)$. We compute:

$$\frac{\nabla\psi}{\psi} = t\frac{\nabla\psi_1}{\psi_1} + (1-t)\frac{\nabla\psi_2}{\psi_2},$$

and

$$\frac{\Delta\psi}{\psi} = t\frac{\Delta\psi_1}{\psi_1} + (1-t)\frac{\Delta\psi_2}{\psi_2} + t(t-1)\left(\frac{\nabla\psi_1}{\psi_1} - \frac{\nabla\phi_2}{\phi_2}\right)^2.$$

It follows that

$$\begin{aligned}\frac{\mathcal{L}\psi(x)}{\psi(x)} &= -\frac{\Delta\psi(x)}{\psi(x)} - 2\frac{\nabla\bar{\phi}}{\bar{\phi}}\cdot\frac{\nabla\psi}{\psi} = t\frac{\mathcal{L}\psi_1(x)}{\psi_1(x)} + (1-t)\frac{\mathcal{L}\psi_2(x)}{\psi_2(x)} - t(t-1)\left(\frac{\nabla\psi_1}{\psi_1} - \frac{\nabla\phi_2}{\phi_2}\right)^2 \\ &\ge t\frac{\mathcal{L}\psi_1(x)}{\psi_1(x)} + (1-t)\frac{\mathcal{L}\psi_2(x)}{\psi_2(x)},\end{aligned}$$

and thus

$$\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi(x)}{\psi(x)} \ge t\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi_1(x)}{\psi_1(x)} + (1-t)\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi_2(x)}{\psi_2(x)},$$

We deduce that

$$\sup_{\psi\in E_\lambda}\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi(x)}{\psi(x)} \ge t\sup_{\psi_1\in E_{\lambda_1}}\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi_1(x)}{\psi_1(x)} + (1-t)\sup_{\psi_2\in E_{\lambda_2}}\inf_{x\in\mathbb{R}^n}\frac{\mathcal{L}\psi_2(x)}{\psi_2(x)},$$

which is nothing but (3.21). Hence, the function $\mu_1^{per}(\lambda)$ is, indeed, concave in λ.

Now, we can prove Theorem 3.2. Let us first summarize some basic properties of the function

$$s(\lambda) = -\lambda_1 - \mu_1^{per}(\widetilde{L}_\lambda).$$

We have just shown that it is convex and, in addition, by assumption we have $s(0) = -\lambda_1 > 0$.

Exercise 3.4 *Show that*

$$\lim_{\lambda\to+\infty}\frac{\mu_1^{per}(\widetilde{L}_\lambda)}{\lambda^2}=-1$$

This exercise implies that the function $s(\lambda)$ is super-linear at infinity:

$$\lim_{\lambda\to+\infty}\frac{s(\lambda)}{\lambda}=+\infty. \tag{3.22}$$

Exercise 3.5 *Use finite differences to show that the function $k(\lambda)=\mu_1^{per}(\widetilde{L}_\lambda)$ and the corresponding eigenfunction ϕ_λ of $\widetilde{\mathcal{L}}_\lambda$ are differentiable in λ (in fact, analytic).*

The last property of $s(\lambda)$ that we will need is

$$s'(0)=k'(0)=0. \tag{3.23}$$

To see this, recall that

$$-\Delta\phi_\lambda+2\lambda(e\cdot\nabla\phi_\lambda)-(\lambda^2+\frac{2\lambda}{\bar\phi}(e\cdot\nabla\bar\phi))\phi_\lambda+\frac{2\nabla\bar\phi}{\bar\phi}\cdot\nabla\phi_\lambda=k(\lambda)\phi_\lambda, \tag{3.24}$$

hence (with $\phi_0=\phi_{\lambda=0}$),

$$-\Delta\phi_0+\frac{2\nabla\bar\phi}{\bar\phi}\cdot\nabla\phi_0=k(0)\phi_0. \tag{3.25}$$

It follows that

$$k(0)=0 \text{ and } \phi_0=1. \tag{3.26}$$

Differentiating (3.24) in λ, we obtain, at $\lambda=0$:

$$-\Delta\psi_0+\frac{2\nabla\bar\phi}{\bar\phi}\cdot\nabla\psi_0-\frac{2}{\bar\phi}(e\cdot\nabla\bar\phi)\phi_0+2(e\cdot\nabla\phi_0)=k(0)\psi_0+k'(0)\phi_0,$$

with the function

$$\psi_0=\frac{d\phi_\lambda}{d\lambda}|_{\lambda=0}.$$

Taking (3.26) into account, this simplifies to

$$-\Delta\psi_0+\frac{2\nabla\bar\phi}{\bar\phi}\cdot\nabla\psi_0-\frac{2}{\bar\phi}(e\cdot\nabla\bar\phi)=k'(0). \tag{3.27}$$

The adjoint equation to (3.25) is

$$-\Delta\phi_0^*-2\nabla\cdot\left(\frac{\nabla\bar\phi}{\bar\phi}\phi_0^*\right)=0, \tag{3.28}$$

or

$$-\nabla\cdot\left(\nabla\phi_0^*+\frac{2\nabla\bar\phi}{\bar\phi}\phi_0^*\right)=0.$$

It is satisfied by $\phi_0^*(x) = 1/\bar\phi^2(x)$. Multiplying then (3.27) by $\bar\phi^{-2}(x)$ and integrating over the period cell gives

$$k'(0)\int_{\mathbb{T}^n}\frac{dx}{\phi_1^2(x)} = -2\int_{\mathbb{T}^n}\frac{e\cdot\nabla\bar\phi}{\bar\phi^3}dx = 0,$$

hence $k'(0) = 0$.

Let us summarize the above observations about the function $s(\lambda)$: we know that $s(\lambda)$ is convex, super-linear at infinity, $s(0) > 0$ and $s'(0) = 0$. It follows that there exists a threshold $c_*(e)$ so that the equation

$$s(\lambda) = c\lambda, \tag{3.29}$$

has no solutions for $0 < c < c_*(e)$, one solution for $c = c_*(e)$ and two solutions for $c > c_*(e)$ – this proves Theorem 3.2.

We will denote below by $\lambda_*(e)$ the unique solution of (3.29) at $c = c_*(e)$, and by $\lambda_e(c)$ the smaller of the two positive solutions for $c > c_*(e)$.

3.3 The Freidlin-Gärtner formula

As we have discussed above, in the introduction to this chapter, the speed of invasion in a direction $e \in \mathbb{S}^{n-1}$ is given not by $c_*(e)$ but by (0.8):

$$w_*(e) = \inf_{|e'|=1,(e\cdot e')>0}\frac{c_*(e')}{(e\cdot e')}. \tag{3.30}$$

In order to understand where (3.30) comes from, recall that for any $e \in \mathbb{S}^{n-1}$ the exponential solution

$$v_e(t,x) = e^{-\lambda_*(e)(x\cdot e - c_*(e)t)}\phi_e(x),$$

with $\phi_e(x) = \phi_{\lambda_*(e)}(x)$, is a super-solution to the Cauchy problem:

$$v_t \ge \Delta v + \mu(x)v - v^2. \tag{3.31}$$

Therefore, the function

$$\bar v(t,x) = \inf_{|e|=1} e^{-\lambda_*(e)(x\cdot e - c_*(e)t)}\phi_e(x)$$

is also a super-solution. Hence, any solution of the Cauchy problem

$$u_t = \Delta u + \mu(x)u - u^2, \tag{3.32}$$

with a compactly supported function $u_0(x)$, lies below a large multiple of $\bar v(t,x)$. In order to see how small $\bar v(t,x)$ is on a given line $L_e = \{re, r > 0\}$ with a fixed $e \in \mathbb{S}^{n-1}$, we need to understand when

$$\inf_{|e'|=1} e^{-\lambda_*(e')(r(e\cdot e') - c_*(e')t)}\phi_{e'}(re)$$

is exponentially small. Note that

$$\inf_{e'\in\mathbb{S}^{n-1},x\in\mathbb{T}^n}\phi_{e'}(x) > 0,$$

thus $\bar{v}(t, re)$ is small if

$$r(e \cdot e') \gg c_*(e')t,$$

for all $|e'| = 1$, and large t, that is, for $r \gg w_*(e)t$. This is where the formula (3.30) for $w_*(e)$ comes from. More precisely, the above argument shows that if we take any $c > w_*(e)$, then we have, for all $x \in \mathbb{R}^n$ fixed:

$$\lim_{t\to+\infty} u(t, x + cte) = 0. \qquad (3.33)$$

The next step is to prove that for each $c \in (0, w_*(e))$ we have

$$\lim_{t\to+\infty} \sup_{0\le r<c} |u(t, rte) - p(rte)| = 0. \qquad (3.34)$$

here $p(x)$ is the unique positive bounded steady solution to (3.31). In turn, the crucial step to establish (3.34) is to show that

$$\liminf_{t\to+\infty} \sup_{0\le r<c} u(t, rte) > 0. \qquad (3.35)$$

With this in hand, we will proceed as before – put a compactly supported sub-solution under $u(t, x)$ at a large time t and let the solution of the corresponding Cauchy problem "grow" to $p(x)$.

Thus, we take $c < c_*(e)$ and go into the moving frame: set $v(t, y) = u(t, y + cte)$:

$$v_t - ce \cdot \nabla v = \Delta v + \mu(y + cte)v - v^2. \qquad (3.36)$$

As in the homogeneous case, the proof of (3.34) boils down to the finding a compactly supported sub-solution to (3.36) that does not vanish as $t \to +\infty$. We will make a (slightly) simplifying assumption that

$$\mu(x) > 0, \text{ for all } x \in \mathbb{T}^n. \qquad (3.37)$$

We will establish the following.

Proposition 3.6 *Let $e \in \mathbb{S}^{n-1}$ and $0 \le c < w_*(e)$. There exists $R_0 > 0$ sufficiently large, $\gamma > 0$, and a positive bounded function $s_e(t, y)$ that satisfies*

$$\frac{\partial s_e}{\partial t} - ce \cdot \nabla s_e = \Delta s_e + \mu(y + cte)s_e - \gamma s_e, \quad t \in \mathbb{R}, \quad |x| \le R_0, \qquad (3.38)$$

with the Dirichlet boundary condition $s_e(t, x) = 0$ for $|x| = R$, and such that

$$\liminf_{t\to+\infty} \inf_{|x|\le R_0/2} s_e(t, x) > 0. \qquad (3.39)$$

Let us first explain how this result implies the Freidlin-Gärtner formula. The first step is the following exercise that uses the techniques we have seen several times.

Exercise 3.7 *Show that to prove the Freidlin-Gärtner formula it is sufficient to show that solution of the KPP problem in the moving ball:*

$$\begin{aligned} &\frac{\partial v}{\partial t} - ce \cdot \nabla v = \Delta v + \mu(y + cte)v - v^2, \quad t \in \mathbb{R}, \quad |x| \le 2R_0, \\ &v = 0 \text{ on } |x| = 2R_0, \end{aligned} \qquad (3.40)$$

satisfies

$$\liminf_{t\to+\infty} \inf_{|x|\le R/2} v(t, x) > 0. \qquad (3.41)$$

The second observation is that the function $s_\varepsilon(t,x) = \varepsilon s_e(t,x)$ with $s_e(t,x)$ as is in Proposition 3.6, is a sub-solution to (3.40), provided that $\varepsilon > 0$ is sufficiently small:

$$\frac{\partial s_\varepsilon}{\partial t} - \Delta s_\varepsilon - \mu(y + cte)s_\varepsilon + s_\varepsilon^2 = -\beta\varepsilon s_e^2 + \varepsilon^2 s_e^2 < 0.$$

This implies that there exists $\beta > 0$ sufficiently small so that $v(t,x) > \beta s_\varepsilon(t,x)$ (choosing β so that this inequality holds at $t = 1$). Now, (3.41) follows from (3.39). Thus, the proof of the Freidlin-Gärtner formula hinges on Proposition 3.6.

The proof of Proposition 3.6: rational angles

Let us first assume that $e \in \mathbb{Q}^n$ is a "rational angle", that is, all components of e are rationally dependent. Then the coefficient $a(t,y) = \mu(y + cte)$ is 1-periodic in y and is also periodic in time, with the period $T_c = M/c$. Here M is the smallest number so that all Me_j are integers. A key role is played by the principal eigenfunction for the problem

$$\begin{aligned} &z_t - \Delta z - ce\cdot\nabla z - a(t,y)z = \lambda_1(c,R)z, \quad t \in \mathbb{R},\ y \in B_R = \{|y| \le R\}, \\ &z(t,y) > 0 \text{ is } T_c\text{-periodic in } t, \\ &z(t,y) = 0 \text{ for } |y| = R. \end{aligned} \tag{3.42}$$

To simplify slightly the notation we do not show explicitly the dependence of $\lambda_1(c,R)$ on e.

Lemma 3.8 *There exists R_1 so that for all $R > R_1$ we can find $c_*(R)$ such that*

$$\lambda_1(c_*(R), R) = 0.$$

The principal periodic eigenvalue λ_1 of the operator

$$-\Delta - \mu(x)$$

is negative when $c = 0$ – this is our main assumption. Lemma 2.4 tells us that then the principal Dirichlet eigenvalue $\lambda_1(R)$ on the ball B_R of the same operator is also negative – in other words, in our curent notation, $\lambda_1(0,R) < 0$ for R sufficiently large – this sets R_1. The function $\lambda_1(c,R)$ is analytic in c, thus $\lambda_1(c,R) < 0$ for all $c > 0$ sufficiently small and R large enough. On the other hand, for all $c > 0$ sufficiently large we have $\lambda_1(c,R) > 0$. To see that, set

$$z(t,x) = e^{-c(x\cdot e)/2}\widetilde{z}(t,x).$$

The function $\widetilde{z}(t,x)$ satisfies

$$\widetilde{z}_t - \Delta\widetilde{z} + \frac{c^2}{4}\widetilde{z} - a(t,y)\widetilde{z} = \lambda_1(c,R)\widetilde{z}, \tag{3.43}$$

with the periodic boundary conditions in T and the Dirichlet boundary conditions on ∂B_R. It follows that $\lambda_1(c,R) > 0$ if

$$c > \sqrt{1 + 4\|a\|_\infty}. \tag{3.44}$$

Thus, there exist $c(R) > 0$ so that $\lambda_1(c(R), R) = 0$ and we will denote by $c_*(R)$ the smallest such $c > 0$ (once again, $c_*(R)$ depens also on e but we do not indicate this dependence explicitly in our notation). Note that $c_*(R)$ is bounded from above because of (3.44).

Exercise 3.9 *Show that $c_*(R)$ is uniformly bounded from below as $R \to +\infty$ (also uniformly in $e \in \mathbb{S}^{n-1}$).*

Here is the key lemma. We use the notation $T_*(R) = T_{c_*(R)}$.

Lemma 3.10 *We have, for all $e \in \mathbb{R}^n$ with $|e| = 1$,*

$$\liminf_{R\to+\infty} c_*(R) \geq w_*(e). \tag{3.45}$$

Let $z_R(t,x)$ be the Dirichlet eigenfunction:

$$\begin{aligned}&\frac{\partial z_R}{\partial t} - \Delta z_R - c_*(R)e\cdot\nabla z_R - a(t,y)z_R = 0, \quad t \in \mathbb{R},\ y \in B_R \\ &z_R(t,y) > 0 \text{ is } T_*(R)\text{-periodic in } t, \\ &z_R(t,y) = 0 \text{ for } |y| = R,\end{aligned} \tag{3.46}$$

normalized so that $z_R(0,0) = 1$. Because of the uniform bounds on $c_*(R)$, we can extract a sub-sequence $R_n \to +\infty$ so that $c_*(R_n) \to \bar{c}$ and the periods $T_*(R_n) = M/c_*(R) \to M/\bar{c}$, and, moreover the functions $z_{R_n}(t,x)$ converge (after possibly extracting another subsequence) locally uniformly to a positive T-periodic function $q(t,x)$ that solves

$$q_t - \Delta q - \bar{c}e\cdot\nabla q - a(t,y)q = 0, \quad t \in \mathbb{R},\ y \in \mathbb{R}^n, \tag{3.47}$$

and satisfies $q(0,0) = 1$. We will now construct an exponential solution to (3.47) starting with $q(t,y)$. Let $\hat{e}_1$ be the first coordinate vector. The Harnack inequality implies that there exists a constant m so that

$$mq(t,y+\hat{e}_1) \leq q(t+T,y), \tag{3.48}$$

for all $y \in \mathbb{R}$ and $t \in \mathbb{R}$. As the function $q(t,y)$ is T-periodic, we conclude that there exist $m, M > 0$ so that

$$mq(t,y+\hat{e}_1) \leq q(t+T,y) \leq Mq(t,y+\hat{e}_1), \tag{3.49}$$

Let M_1 be the smallest M so that this inequality holds. If there exists t_0, y_0 so that

$$q(t_0,y_0) = M_1 q(t_0,y_0+\hat{e}_1),$$

and $q(t,y) \leq M_1 q(t,y+\hat{e}_1)$ for all $t \in \mathbb{R}$ and $y \in \mathbb{R}^n$, the maximum principle would imply that

$$q(t,y) = Mq(t,y+\hat{e}_1), \text{ for all } t \in \mathbb{R} \text{ and } y \in \mathbb{R}^n. \tag{3.50}$$

On the other hand, if there exists a sequence of points t_n, y_n such that

$$q(t_n,y_n) \geq (M_1 - \frac{1}{n})q(t_n,y_n+\hat{e}_1),$$

then by considering the shifted functions $q_m(t,y) = q(t+t_m, y+[y_m])$ and passing to the limit $n \to +\infty$ we would construct a solution $\bar{q}(t,x)$ of (3.47) such that

$$\bar{q}(0,\bar{y}) = M_1\bar{q}(0,\bar{y}+\hat{e}_1),$$

with some $\bar{y} \in [0,1]^n$, and $\bar{q}(t,y) \le M_1 \bar{q}(t, y+e)$ for all $t \in \mathbb{R}$ and $y \in \mathbb{R}^n$. The maximum principle would, once again, imply that $\bar{q}(t,x)$ satisfies (3.33). In other words, $\bar{q}(t,y)$ is a solution of

$$\bar{q}_t - \Delta\bar{q} - \bar{c}e \cdot \nabla\bar{q} - a(t,y)\bar{q} = 0, \tag{3.51}$$

which, in addition, satisfies (with $\lambda_1 = \log M_1$)

$$\bar{q}(t,y) = e^{-\lambda_1 y_1}\Psi_1(t,y), \tag{3.52}$$

with a function $\Psi_1(t,y)$ which is 1-periodic in y_1 and T-periodic in t. Iterating this process we will construct a solution of (3.51) that is of the form

$$\bar{q}(t,y) = e^{-\sum_{i=1}^{n}\lambda_i y_i}\Psi(t,y). \tag{3.53}$$

Here, the function $\Psi(t,y)$ is T-periodic in t and 1-periodic in all y_i. In the original variables this corresponds to a solution of

$$r_t = \Delta r + \mu(x) r \tag{3.54}$$

of the form

$$r(t,x) = e^{-\sum_{i=1}^{n}\lambda_i(x_i - \bar{c}e_i t)}\Phi(t,x). \tag{3.55}$$

As $T = M/\bar{c}$, the function $\Phi(t,x) = \Psi(t, x - \bar{c}te)$ is T-periodic in time, 1-periodic in all x_i, and satisfies an autonomous equation

$$\Phi_t + \bar{c}(e \cdot \lambda)\Phi = \Delta\Phi - 2\lambda \cdot \nabla\Phi + |\lambda|^2\Phi + \mu(x)\Phi.$$

As in the previous step, it follows that for any τ there exists q so that $\Phi(t+\tau, y) = q\Phi(t,y)$. Periodicity of $\Phi(t,y)$ in t implies then that $q = 1$ so that Φ does not depend on t: $\Phi(x)$ is a 1-periodic function of x. Then, we set $e'_i = \lambda_i/|\lambda|$ and write

$$\sum_{i=1}^{n}\lambda_i(x_i - \bar{c}e_i t) = |\lambda|\sum_{i=1}^{n}(x_i e'_i - \bar{c}te'_i e_i) = |\lambda|[(x \cdot e') - \bar{c}'t],$$

with $\bar{c}' = \bar{c}(e \cdot e')$. Therefore, $r(t,x)$ is an exponential solution in the direction e' moving with the speed $\bar{c}'$. It follows that $\bar{c}(e \cdot e') \ge c_*(e')$, hence

$$\bar{c} \ge \frac{c_*(e')}{(e \cdot e')} \ge w_*(e),$$

and the proof of Lemma 3.10 is complete.

Returning to the proof of Proposition 3.6 for rational angles, we conclude that for any $c < w_*(e)$, we can find R sufficiently large so that $\lambda_1(R) < 0$. Taking the corresponding eigenfunction for (3.46) on B_R as the function $s_e(t,x)$ in (3.38), we deduce the claim of that proposition.

The proof of Proposition 3.6: irrational angles

We now consider an irrational direction e – the proof is essentially by a density argument. Here, we will use the simplifying assumption (3.37) that

$$m := \inf_{x \in \mathbb{T}^n} \mu(x) > 0. \tag{3.56}$$

Exercise 3.11 *It follows from (3.56) that the solution propagates in all directions at least at the speed* $2\sqrt{m} := \underline{c}$*: for any* $0 \le c < \underline{c}$ *and any* $x \in \mathbb{R}^n$ *we have*

$$\liminf_{t \to +\infty} u(t, x + ct) > 0. \tag{3.57}$$

Given $c < w_*(e)$, we take $\varepsilon > 0$ sufficiently small, and consider a rational direction e_ε at distance at most ε^2 from e:

$$|e - e_\varepsilon| \le \varepsilon^2,$$

and such that

$$|w_*(e) - w_*(e_\varepsilon)| \le \varepsilon^2,$$

and so that we would have

$$c_\varepsilon = c(1 + \varepsilon) < w_*(e_\varepsilon).$$

Consider now a large time $T > 0$ and the corresponding positions along the two rays:

$$X = c_\varepsilon T e \text{ and } X_\varepsilon = c_\varepsilon T e_\varepsilon.$$

As $c_\varepsilon < w_*(e_\varepsilon)$, if T is sufficiently large (possibly depending on ε), then, by what we have shown for the propagation in a rational direction, the function $u(T, x)$ is larger than some $\delta > 0$ in a large ball centred at X_ε. Therefore, as follows from Exercise 3.11, at the time $T' = T + \varepsilon T$, u will be larger than δ in a ball of radius at least $\underline{c}\varepsilon T$, centered at X_ε. However, for small ε, the point X is in this ball:

$$|X - X_\varepsilon| \le c\varepsilon^2 T \ll \underline{c}\varepsilon T.$$

It follows that, for all T sufficiently large we have

$$u(T(1 + \varepsilon), c(1 + \varepsilon)Te) \ge \delta,$$

which implies that

$$\liminf_{t \to +\infty} u(t, cte) > 0.$$

Chapter 5

Counting the solutions of Hamilton-Jacobi equations

1 Introduction

In this chapter, we will consider the Hamilton-Jacobi equations

$$u_t + H(x, \nabla u) = 0 \ (t > 0,\ x \in \mathbb{T}^n), \tag{1.1}$$

on the unit torus $\mathbb{T}^n \subset \mathbb{R}^n$. Such problems arise in problems in physics (front propagation), imaging (sharpening and other tools), optimal control theory, and finance, – this list may be continued to a nearly arbitrary length, and we encourage the reader to investigate his favorite applications. We will be interested both in the Cauchy problem, that is, (1.1) supplemented with the initial data

$$u(0, x) = u_0(x), \tag{1.2}$$

as well as in a stationary version of (1.1):

$$H(x, \nabla u) = c, \quad x \in \mathbb{T}^n. \tag{1.3}$$

It will soon become clear why (1.3) has a constant c in the right side – we will prove that under reasonable assumptions, solutions exists only for a unique value of c which has no reason to be equal to zero. Thus, the "standard" steady equation

$$H(x, \nabla u) = 0$$

typically would have no solutions.

We will ultimately make the assumption that the function $H(x, p)$ is smooth and uniformly strictly convex in its second variable, that is there exists $\alpha > 0$ such that the inequality

$$D_p^2 H(x, p) \geq \alpha I, \quad [D_p^2 H(x, p)]_{ij} = \frac{\partial^2 H}{\partial p_i \partial p_j}, \tag{1.4}$$

is true, in the sense of quadratic forms, for all $x \in \mathbb{T}^n$ and $p \in \mathbb{R}^n$.

A reader familiar with the theory of conservation laws, would see immediately the connection between them and the Hamilton-Jacobi equations: in one dimension, $n = 1$, differentiating (1.1) in x we get a conservation law for $v = u_x$:

$$v_t + (H(x, v))_x = 0. \tag{1.5}$$

The basic conservation law theory tells us that it is reasonable to expect that $v(t, x)$ becomes discontinuous in x at a finite time t, which means that the function $u(t, x)$ fails to be in C^1 though it will remain Lipschitz. In agreement with this intuition, it is well known that, for smooth initial data u_0 on $\mathbb{T}^n$, the Cauchy problem (1.1)-(1.2) has a unique *local* smooth solution. That is, there exists a time $t_0 > 0$, which depends on the initial data u_0, such that (1.1) has a C^1 solution $u(t, x)$ on the time interval $[0, t_0]$ such that $u(0, .) = u_0$. However, this solution is not global: in general, it is impossible to extend it in a smooth fashion to $t = +\infty$. This is described very nicely in [52]. On the other hand, if we relax the smoothness constraint: "u is C^1 on $\mathbb{R}_+ \times \mathbb{T}^n$", and replace it by: "$u$ is Lipschitz on $\mathbb{R}_+ \times \mathbb{T}^n$" – and require (1.1)-(1.2) to hold almost everywhere, there are, in general, several solutions to the Cauchy problem (this parallels the fact that the weak solutions to the conservation laws are not unique). See, for instance, [85] for a discussion of these issues. A natural question is, therefore, to know if an additional criterion, less stringent than the C^1 regularity, but stronger than the mere Lipschitz regularity, enables us to select a unique solution to the Cauchy problem – as the notion of the entropy solutions does for the conservation laws.

The above considerations have motivated the introduction, by Crandall and Lions [42], at the beginning of the 80's, of the notion of a *viscosity solution* to (1.1). This notion selects, among all the solutions of (1.1), "the one that has a physical meaning" – though understanding the connection to physics may require some thought from the reader. Being weaker than the notion of a classical solution, it introduces new difficulties to the existence and uniqueness issues. Note that even if there is a unique viscosity solution to the Cauchy problem (1.1)-(1.2), the stationary equation (1.3) has no reason to have a unique steady solution. The classification of all solutions to (1.3) is the problem that motivates the present chapter.

The unique viscosity solution of (1.1)-(1.2) has, as we shall remind the reader, an (almost) explicit expression, known as the *Lax-Oleinik formula*. At the end of the 90's, A. Fathi, in four notes [57], [58], [59], [60], shows how the exploration of this (at a first glance, intractable) formula, combined to (simple) ideas coming from the Mather theory [90] - could shed a new light on the qualitative properties of (1.1)-(1.2) and (1.3). The ideas present in these four notes have been further developed in [61], and culminate at an important theorem asserting the existence of smooth critical C^1 sub-solutions, presented in Fathi-Siconolfi [62]. This chapter is a self-contained introduction to the Fathi-Siconolfi theorem, and how it allows to understand the qualitative properties of the Hamilton-Jacobi equations.

Section 2 of this chapter is an introduction to viscosity solutions, ending up with an existence theorem (due to a preprint by Lions, Papanicolaou, Varardhan – perhaps, the most cited unpublished paper in PDEs) for (1.3). Section 3 recalls the Lax-Oleinik formula as an explicit solution to the Cauchy problem for (1.1)-(1.2), and outlines its main properties. The short Section 4 presents, in quite a simple way, the $C^{1,1}$ regularity properties that will turn out to be crucial for the sequel. Section 5 analyses the Fathi-Siconolfi theorem, and its applications to the uniqueness sets for (1.3), and existence of invariant regions for the underlying hamiltonian system.

We would be amiss not to mention that one could at no cost replace the torus $\mathbb{T}^n$ by a smooth compact Riemaniann manifold without boundary. This would, however, only result in a heavier presentation and would in no way lead to a deeper understanding.

2 Viscosity solutions

Here, we present the basic notions of the viscosity solutions for the first order Hamilton-Jacobi equations, and prove a uniqueness result which is typical in this theory. This brings us very quickly to the Lions-Papanicolaou-Varadhan fundamental theorem of existence of steady solutions; this theorem raises uniqueness issues that will be one of the motivations to go further. The reader interested in all the subtleties of the theory may enjoy reading Barles [6], or Lions [85].

2.1 The definition of a viscosity solution

Let us begin with more general equations than (1.1) – we will restrict the assumptions as the theory develops. Consider the Cauchy problem

$$u_t + F(x, u, \nabla u) = 0 \ (t > 0,\ x \in \mathbb{T}^n), \qquad u(0, x) = u_0(x), \tag{2.1}$$

with $F \in C(\mathbb{T}^n \times \mathbb{R} \times \mathbb{R}^n; \mathbb{R})$. The initial datum u_0 is assumed – this will always be the case – continuous on $\mathbb{T}^n$.

In order to motivate the notion of a viscosity solution, philosophically, one believes that (smooth) solutions of the regularized problem

$$u_t^\varepsilon + F(x, u^\varepsilon, \nabla u^\varepsilon) = \varepsilon \Delta u^\varepsilon \tag{2.2}$$

are a good approximation to $u(t, x)$. Hence, a natural attempt would be to pass to the limit $\varepsilon \downarrow 0$ in (2.2). This, however, is too blunt to succeed in general – we will comment more on this below. To motivate a different route, instead, consider a smooth sub-solution of (2.2):

$$u_t + F(x, u, \nabla u) \le \varepsilon \Delta u. \tag{2.3}$$

Next, take a smooth function $\phi(t, x)$ such that the difference $u - \phi$ attains its maximum at a point (t_0, x_0). Then, at this point we have

$$u_t(t_0, x_0) = \phi_t(t_0, x_0), \quad \nabla \phi(t_0, x_0) = \nabla u(t_0, x_0),$$

and

$$D^2 \phi(t_0, x_0) \ge D^2 u(t_0, x_0).$$

It follows that

$$\begin{aligned} &\phi_t(t_0, x_0) + F(x_0, u(t_0, x_0), \nabla \phi(t_0, x_0)) - \varepsilon \Delta \phi(t_0, x_0) \\ &\le u_t(t_0, x_0) + F(x_0, u(t_0, x_0), \nabla u(t_0, x_0)) - \varepsilon \Delta u(t_0, x_0) \le 0. \end{aligned} \tag{2.4}$$

In a similar vein, if $u(t, x)$ is a smooth super-solution to the regularized problem:

$$u_t + F(x, u, \nabla u) \ge \varepsilon \Delta u, \tag{2.5}$$

we consider a smooth function $\phi(t,x)$ such that the difference $u-\phi$ attains its minimum at a point (t_0,x_0). Then, at this point we have

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))-\varepsilon\Delta\phi(t_0,x_0)\geq 0. \tag{2.6}$$

Passing to the limit $\varepsilon\downarrow 0$ for a smooth test function $\phi(t,x)$ (rather than an uncontrolled approximation $u^\varepsilon(t,x)$) in (2.4) and (2.6) leads to the following.

Definition 2.1 *We say that $u(t,x)\in C([0,+\infty),\mathbb{T}^n)$ is a viscosity subsolution to* (2.1) *if, for all test functions $\phi\in C^1([0,+\infty)\times\mathbb{T}^n)$ and all $(t_0,x_0)\in(0,+\infty)\times\mathbb{T}^n$ such that (t_0,x_0) is a local maximum for $u-\phi$, we have:*

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))\leq 0. \tag{2.7}$$

Furthermore, we say that $u(t,x)$ is a viscosity supersolution to (2.1) *if, for all test functions $\phi\in C^1((0,+\infty)\times\mathbb{T}^n)$ and all $(t_0,x_0)\in(0,+\infty)\times\mathbb{T}^n$ such that the point (t_0,x_0) is a local minimum for the difference $u-\phi$, we have:*

$$\phi_t(t_0,x_0)+F(x_0,u(t_0,x_0),\nabla\phi(t_0,x_0))\geq 0. \tag{2.8}$$

Finally, $u(t,x)$ is a viscosity solution to (2.1) *if it is both a viscosity subsolution and a viscosity supersolution to* (2.1).

Definition 2.1 trivially extends to steady equations of the type $F(x,u,\nabla u)=0$ on $\mathbb{T}^n$.

Exercise 2.2 *A C^1 solution to* (2.1) *is a viscosity solution. The maximum of two viscosity subsolutions is a viscosity subsolution, and the minimum of two viscosity supersolutions is a viscosity supersolution.*

The reader may justly wonder whether such a seemingly weak definition has any selective power. This is the case, and we give below, without proof, a list of some basic properties of the viscosity solutions to (2.1), as exercises to the reader. These exercises are not so easy as Exercise 2.2, but the hints below should be helpful.

Exercise 2.3 *(Stability) Let F_j be a sequence of functions in $C(\mathbb{T}^n\times\mathbb{R}\times\mathbb{R}^n)$, which converges locally uniformly to $F\in C(\mathbb{T}^n\times\mathbb{R}\times\mathbb{R}^n)$. Let u_j be a viscosity solution to* (2.1) *with $F=F_j$, and assume that u_j converges locally uniformly to $u\in C(\mathbb{R}_+,\mathbb{T}^n)$. Then u is a viscosity solution to* (2.1).

Hint: this is not difficult.

Exercise 2.4 *Let u be a locally Lipschitz viscosity solution to* (2.1). *Then it satisfies* (2.1) *almost everywhere.*

Hint: if u is Lipschitz, then u is differentiable almost everywhere. Prove that, at a point of differentiability (t_0,x_0), one may construct a C^1 test function $\phi(t,x)$ such that (t_0,x_0) is a local maximum (respectively, a local minimum) of $u-\phi$. If you have no idea of how to do it, see [42].

Exercise 2.5 *(Maximum principle) Assume that the Hamiltonian* $F(x,u,p)$ *has the form* $H(x,p)$*, where* $H \in C(\mathbb{T}^n \times \mathbb{R}^n)$ *satisfies the following (coercivity) property:*

$$\lim_{|p|\to+\infty} H(x,p) = +\infty, \quad \textit{uniformly in } x \in \mathbb{T}^n. \tag{2.9}$$

Let $u_{10} \le u_{20}$ *be two continuous initial data for* (2.1)*, and assume that* u_{10} *(resp.* u_{20}*) generates at least one viscosity solution* u_1 *(resp.* u_2*) for* (2.1)*. Then* $u_1 \le u_2$*.*

Hint: try to reproduce the proof of Proposition 2.7 below.

Exercise 2.6 *(Weak contraction) If* $F(x,u,p) = H(x,p)$*, and under the coercivity assumption* (2.9)*, let* $u_{10} \le u_{20}$ *be two continuous initial data for* (2.1)*. Assume that* u_{10} *(respectively,* u_{20}*) generates at least one viscosity solution* u_1 *(respectively,* u_2*) for* (2.1)*. Then, we have* $\|u_1(t,.) - u_2(t,.)\|_\infty \le \|u_{10} - u_{20}\|_\infty$*.*

Hint: notice that the constants solve the equation and use Exercise 2.5.

Definition 2.1 has been introduced by Crandall and Lions in their seminal paper [42]. Let us notice one of the main advantages of the notion: Exercise 2.3 asserts that one may safely "pass to the limit" in equation (2.1), as soon as estimates on the moduli of continuity of the solutions are available (This, however, may be very difficult to prove, even impossible). Exercise 2.5 implies uniqueness of the solutions to the Cauchy problem - without, however, implying existence.

The name "viscosity solution" comes out of trying to identify a "physically meaningful" solution to (2.1). A natural idea is to regularize (2.1) by a second order dissipative term, and to solve the – apparently more complicated – problem (2.2):

$$u_t + F(x,u,\nabla u) = \varepsilon\Delta u \ (t>0,\ x \in \mathbb{T}^n), \quad u(0,x) = u_0(x), \tag{2.10}$$

Then one tries to pass to the limit $\varepsilon \to 0$. This can be carried out when the Hamiltonian $F(x,u,p)$ has, for instance, the form $H(x,p)$. It is possible to prove that there is a unique limiting solution and that one actually ends up with a nonlinear semigroup. In particular, one may show that, if we take this notion of solution as a definition, there are uniqueness and contraction properties analogous to above – see [85] for further details. Taking (2.10) as a definition is, however, not intrinsic: there is always the danger that the solution depends on the underlying regularization (why regularize with the Laplacian?), and Definition 2.1 bypasses this philosophical question. Let us finally note that the notion of a viscosity solution has turned out to be especially relevant to the second order elliptic and parabolic equations – especially those fully nonlinear with respect to the Hessian of the solution. There have been spectacular developments, which are out of the scope of this chapter.

Warning. For the rest of this chapter, a solution of (1.1) or (1.3) will always be meant in the viscosity sense.

One of the main issues of viscosity solutions theory is uniqueness. So, let us give the simplest uniqueness result, and prove it by the method of doubling of variables. This argument appears in almost all uniqueness proofs, in more or less elaborate forms.

Proposition 2.7 *Assume that the Hamiltonian* H *is continuous in all its variables, and satisfies the coercivity assumption* (2.9)*. Consider the equation*

$$H(x,\nabla u) + u = 0, \quad x \in \mathbb{T}^n. \tag{2.11}$$

Let $\underline{u}$ *and* $\overline{u}$ *be, respectively, a viscosity sub- and a super-solution* (1.1)*, then* $\underline{u} \le \overline{u}$*.*

Proof. Assume for a moment that both $\underline{u}$ and $\overline{u}$ are C^1. If x_0 is a minimum of $\overline{u}-\underline{u}$ we have,

$$H(x_0, \nabla \underline{u}(x_0)) + \overline{u}(x_0) \geq 0, \tag{2.12}$$

as $\overline{u}$ is a super-solution, and $\underline{u}$ can be considered a test function. On the other hand, $\underline{u} - \bar{u}$ attains its maximum at x_0, and, as $\underline{u}$ is a sub-solution, and $\overline{u}$ can serve as a test function, we have

$$H(x_0, \nabla \overline{u}(x_0)) + \underline{u}(x_0) \leq 0. \tag{2.13}$$

As x_0 is a minimum of $\overline{u} - \underline{u}$, we have $\nabla \overline{u}(x_0) = \nabla \underline{u}(x_0)$, whence (2.12) and (2.13) imply

$$\underline{u}(x_0) \leq \overline{u}(x_0).$$

As x_0 is a minimum of $\overline{u} - \underline{u}$, we conclude that $\overline{u}(x) \geq \underline{u}(x)$ for all $x \in \mathbb{R}^n$ if both of these functions are $C^1(\mathbb{R}^n)$. Unfortunately, $\underline{u}$ are $\overline{u}$ only continuous, so we can not always use the argument above. Let us define, for all $\varepsilon > 0$, the penalization

$$u_\varepsilon(x, y) = \overline{u}(x) - \underline{u}(y) + \frac{|x-y|^2}{2\varepsilon^2}$$

and let $(x_\varepsilon, y_\varepsilon)$ be a minimum for u_ε.

Exercise 2.8 *Show that*

$$\lim_{\varepsilon \to 0} |x_\varepsilon - y_\varepsilon| = 0.$$

and that the family $(x_\varepsilon, y_\varepsilon)$ *converges, as* $\varepsilon \to 0$*, up to a subsequence, to a point* (x_0, x_0)*, where* x_0 *is a minimum to* $\overline{u} - \underline{u}$*.*

Consider the function

$$\phi(x) = \underline{u}(y_\varepsilon) - \frac{|x - y_\varepsilon|^2}{2\varepsilon^2},$$

as a (smooth) function of the variable x. The difference

$$\overline{u}(x) - \phi(x; y_\varepsilon) = u_\varepsilon(x, y_\varepsilon)$$

attains its minimum at the point $x = x_\varepsilon$, where, as $\overline{u}(x)$ is a super-solution, we have

$$H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + \overline{u}(x_\varepsilon) \geq 0. \tag{2.14}$$

Next, we apply the sub-solution part of Definition 2.7 to the test function

$$\psi(y) = \overline{u}(x_\varepsilon) + \frac{|x_\varepsilon - y|^2}{2\varepsilon^2},$$

considered as a function of y. The difference

$$\underline{u}(y) - \psi(y) = \underline{u}(y) - \overline{u}(x_\varepsilon) - \frac{|x_\varepsilon - y|^2}{2\varepsilon^2} = -u_\varepsilon(x, y_\varepsilon)$$

attains its maximum at $y = y_\varepsilon$, hence

$$H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + \underline{u}(y_\varepsilon) \leq 0; \tag{2.15}$$

The coercivity of the Hamiltonian and (2.15) imply that $|x_\varepsilon - y_\varepsilon|/\varepsilon^2$ is bounded. Hence, as $|x_\varepsilon - y_\varepsilon| \to 0$, it follows that

$$H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) = H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) + o(1).$$

Subtracting (2.15) from (2.14), we obtain

$$\overline{u}(x_\varepsilon) - \underline{u}(y_\varepsilon) \geq o(1).$$

Sending $\varepsilon \to 0$ implies

$$\overline{u}(x_0) - \underline{u}(x_0) \geq 0,$$

and, as x_0 is the minimum of $\overline{u} - \underline{u}$, the proof is complete. □

An immediate consequence is that (2.11) has at most one solution. It has the following easy extension:

Proposition 2.9 *Let $\underline{u}$ (respectively, $\overline{u}$) be an upper semicontinuous (u.s.c.) subsolution (respectively, a lower semicontinuous supersolution) to* (2.11)*, then $\underline{u} \leq \overline{u}$.*

2.2 Steady solutions

If one were to look for solutions of (1.1) which do not depend on t, that is, solutions to

$$H(x, \nabla u) = 0,$$

one would quickly realize that there is no reason why such an equation should have solutions: think, for instance, of $H(x,p)$ such that $H \geq 1$ on $\mathbb{T}^n \times \mathbb{R}^n$ – this is something that we have by no means excluded yet. It is, therefore, desirable to extend a bit the class of solutions and a natural attempt is to look for solutions of (1.1) of the form $-ct + u(x)$, with a constant $c \in \mathbb{R}$. Such function u solves

$$H(x, \nabla u) = c, \quad x \in \mathbb{T}^n. \tag{2.16}$$

Let us point out that (2.16) may have solutions for at most one c. Indeed, assume there exist $c_1 \neq c_2$, such that (2.16) has a solution u_1 for $c = c_1$ and another solution u_2 for $c = c_2$. Let $K > 0$ be such that

$$u_1 - K \leq u_2 \leq u_1 + K.$$

The functions $-c_1 t + u_1 \pm K$ and $-c_2 t + u_2$ solve the Cauchy problem (1.1) with the respective initial data $u_1 \pm K$ and u_2. By the maximum principle (Exercise 2.5), we have

$$-c_1 t + u_1(x) - K \leq -c_2 t + u_2(x) \leq -c_1 t + u_1 + K.$$

This is a contradiction since $c_1 \neq c_2$, and the functions u_1 and u_2 are bounded.

The main result of this section is the following theorem (due to Lions, Papanicolaou, Varadhan [86]) that asserts the existence of a constant c for which (2.16) has a solution.

Theorem 2.10 *Assume $H(x,p)$ is continuous, and the coercivity condition* (2.9) *holds. There is a unique $c \in \mathbb{R}$ for which* (2.16) *has solutions.*

It is important to point out that the periodicity assumption on the Hamiltonian is indispensable – for instance, when $H(x,p)$ is a random function (in x) on $\mathbb{R}^n \times \mathbb{R}^n$, the situation is totally different – an interested reader should consult the literature on stochastic homogenization of the Hamilton-Jacobi equations, a research area that is active and evolving at the moment of this writing.

The homogenization connection

Before proceeding with the proof of the Lions-Papanicolaou-Varadhan theorem, let us explain how the steady equation (2.16) appears in the homogenization context, which was probably the main motivation behind this theorem. Consider the Cauchy problem

$$u_t^\varepsilon + H(x, \nabla u^\varepsilon) = 0, \tag{2.17}$$

in the whole space $x \in \mathbb{R}^n$ (and not on the torus). We assume that the initial datum is slowly varying and large: $u^\varepsilon(0, x) = \varepsilon^{-1} u_0(\varepsilon x)$, and introduce the "slow" variables $y = \varepsilon x$ and $s = \varepsilon t$, as well as the function

$$v^\varepsilon(s, y) = \varepsilon u^\varepsilon(\frac{s}{\varepsilon}, \frac{y}{\varepsilon}).$$

In the new variables the problem takes the form

$$v_s^\varepsilon + H(\frac{y}{\varepsilon}, \nabla v^\varepsilon) = 0, \quad y \in \mathbb{R}^n, \quad t > 0, \tag{2.18}$$

with the initial data $v^\varepsilon(0, y) = u_0(y)$. Let us seek the solution in the form of an asymptotic expansion

$$v^\varepsilon(s, y) = \bar{v}(s, y) + \varepsilon v_1(s, y, \frac{y}{\varepsilon}) + \varepsilon^2 v_2(s, y, \frac{y}{\varepsilon}) + \dots$$

The functions $v_j(s, y, z)$ are assumed to be periodic in the "fast" variable z. Inserting this expansion into (2.18), we obtain in the leading order

$$\bar{v}_s(s, y) + H(\frac{y}{\varepsilon}, \nabla_y \bar{v}(s, y) + \nabla_z v_1(s, y, \frac{y}{\varepsilon})) = 0. \tag{2.19}$$

As is standard in such multiple scale expansions, we consider (2.19) as

$$\bar{v}_s(s, y) + H(z, \nabla_y \bar{v}(s, y) + \nabla_z v_1(s, y, z)) = 0, \tag{2.20}$$

an equation for v_1 as a function of the fast variable $z \in \mathbb{T}^n$, for each $s > 0$ and $y \in \mathbb{R}^n$ fixed. The function $\bar{v}(s, y)$ will then be found from the solvability condition for (2.19). The latter is obtained as follows: for each fixed $p \in \mathbb{R}^n$ consider the problem

$$H(z, p + \nabla_z w) = \bar{H}(p), \tag{2.21}$$

posed on the torus $z \in \mathbb{T}^n$, for an unknown function w. Here, $\bar{H}(p)$ is the unique constant (that depends on p), whose existence is guaranteed by the Lions-Papanicolaou-Varadhan theorem, for which the equation

$$H(z, p + \nabla_z w) = c, \tag{2.22}$$

has a solution. Then, the function $\bar{v}(s, y)$ satisfies the homogenized (or effective) equation

$$\bar{v}_s + \bar{H}(\nabla_y \bar{v}) = 0, \quad \bar{v}(0, y) = u_0(y), \quad s > 0, \ y \in \mathbb{R}^n, \tag{2.23}$$

and the function $\bar{H}(p)$ is called the homogenized Hamiltonian. Note that the effective Hamiltonian does not depend on the spatial variable – the "small scale" variations are averaged out via the above homogenization procedure. Thus, the existence and uniqueness of the constant c for which solution of the steady equation (2.22) exists, is directly related to the homogenization (long time behavior) of solutions of the Cauchy problem (2.17) with slowly varying initial data.

The proof of the Lions-Papanicolaou-Varadhan theorem

As we have already proved uniqueness of the constant c, it only remains to prove its existence. We start with a double approximation problem

$$-\delta\Delta u^{\varepsilon,\delta} + H(x, \nabla u^{\varepsilon,\delta}) + \varepsilon u^{\varepsilon,\delta} = 0, \quad x \in \mathbb{T}^n, \tag{2.24}$$

with $\delta > 0$ and $\varepsilon > 0$. The idea is to send first $\delta \downarrow 0$, and then $\varepsilon \downarrow 0$. Sending $\delta \downarrow 0$ will be harmless, while the limit of εu will produce the correct constant c. The first step is to show that solution of (2.24) exists.

Exercise 2.11 *Show that for all $\delta > 0$ and for all $\varepsilon > 0$, (2.24) has a solution $u^{\varepsilon,\delta}$, which satisfies*

$$-\frac{\|H(\cdot,0)\|_\infty}{\varepsilon} \le u^{\varepsilon,\delta}(x) \le \frac{\|H(\cdot,0)\|_\infty}{\varepsilon}, \tag{2.25}$$

for all $x \in \mathbb{T}^n$. Hint: this can be seen by the sub and supersolution method for the elliptic equations, as

$$-\frac{\|H(\cdot,0)\|_\infty}{\varepsilon} \text{ and } \frac{\|H(\cdot,0)\|_\infty}{\varepsilon}$$

are a pair of ordered sub and super-solutions.

The limit $\delta \downarrow 0$ is taken care of by the Barles-Perthame lemma [7]:

Lemma 2.12 *Fix $\varepsilon > 0$ and consider*

$$\underline{u}^\varepsilon(x) = \limsup_{y\to x,\delta\to 0} u^{\varepsilon,\delta}(y), \qquad \overline{u}^\varepsilon(x) = \liminf_{y\to x,\delta\to 0} u^{\varepsilon,\delta}(y).$$

Then, $\overline{u}^\varepsilon$ (respectively, $\underline{u}^\varepsilon$) is a l.s.c super-solution (respectively, u.s.c sub-solution) to

$$H(x, \nabla u) + \varepsilon u = 0 \qquad (x \in \mathbb{T}^n) \tag{2.26}$$

We leave the proof of this lemma as an exercise. As a hint, we note that if x_ε is a minimum of $\overline{u}^\varepsilon - \phi$, one may use the definition of $\overline{u}^\varepsilon$ to construct a sequence of minima to $u^{\varepsilon,\delta} - \phi$.

We have $\underline{u}^\varepsilon \ge \overline{u}^\varepsilon$, from the definition of $\underline{u}^\varepsilon$ and $\overline{u}^\varepsilon$, thus $\overline{u}^\varepsilon = \underline{u}^\varepsilon$ by Proposition 2.9. Let u^ε be this common value – it is a solution to (2.26). As u^ε is both upper- and lower-semicontinuous, it is continuos, and satisfies

$$\varepsilon|u^\varepsilon(x)| \le \|H(\cdot,0)\|_\infty,$$

as seen from (2.25). In order to pass to the limit $\varepsilon \to 0$ in (2.26), we need a modulus of continuity estimate.

Lemma 2.13 *There is $C > 0$ independent of ε such that $|\mathrm{Lip}\, u^\varepsilon| \le C$.*

Proof. Fix $x \in \mathbb{T}^n$ and, for $K > 0$, consider the function

$$\zeta(y) = u^\varepsilon(y) - u^\varepsilon(x) - K|y - x|.$$

Let $\hat{x}$ be a maximum of $\zeta(y)$ (the point $\hat{x}$ depends on x). If $\hat{x} = x$ for all $x \in \mathbb{T}^n$, we have

$$u^\varepsilon(y) - u^\varepsilon(x) \le K|x - y|,$$

for all $x, y \in \mathbb{T}^n$, which implies that u^ε is Lipschitz. If there exists some x such that $\hat{x} \neq x$, then the function

$$\psi(y) = u^\varepsilon(x) + K|y - x|$$

is, in a vicinity of the point $\hat{x}$, an admissible test function. Moreover, the difference

$$u^\varepsilon(y) - \psi(y) = \zeta(y)$$

attains its maximum at $y = \hat{x}$. The sub-solution condition (2.7) at this point gives:

$$H(\hat{x}, K\frac{\hat{x} - x}{|\hat{x} - x|}) + \varepsilon u^\varepsilon(\hat{x}) \le 0.$$

As $\varepsilon u^\varepsilon(x)$ is bounded by $\|H(\cdot, 0)\|_\infty$, the coercivity condition (2.9) implies the existence of a constant $C > 0$ independent of ε such that $K \le C$. Therefore, if we take $K = 2C$, we must have $\hat{x} = x$ for all $x \in \mathbb{T}^n$, which implies

$$u(y) - u(x) - 2C|y - x| \le 0.$$

The points x and y being arbitrary, this finishes the proof. □

In order to finish the proof of Theorem 2.10, denote by $\langle u^\varepsilon \rangle$ the mean of u^ε over $\mathbb{T}^n$, and set

$$v^\varepsilon = u^\varepsilon - \langle u^\varepsilon \rangle.$$

This function satisfies

$$H(x, \nabla v^\varepsilon) + \varepsilon \langle u^\varepsilon \rangle + \varepsilon v^\varepsilon = 0.$$

Because of Lemma 2.13, the family v^ε converges uniformly, up to a subsequence, to a function $v \in C(\mathbb{T}^n)$, and $\varepsilon v^\varepsilon \to 0$. The bound (2.25) implies that the family $\varepsilon \langle u^\varepsilon \rangle$ is bounded. We may, therefore, assume it convergence (along a subsequence) to a constant denoted by $-c$. By the stability result in Exercise 2.3, v is a viscosity solution of

$$H(x, \nabla v) = c. \tag{2.27}$$

This finishes the proof of Theorem 2.10. □

Non-uniqueness of steady solutions

Once the correct c has been identified, one may wonder about uniqueness of the solution for equation (2.16). Clearly, if u is a solution, $u + q$ is also a solution for every constant q. However, uniqueness modulo constants is also not true. Consider the very simple example

$$|u'| = f(x), \quad x \in \mathbb{T}^1. \tag{2.28}$$

Assume that $f \in C^1(\mathbb{T}^1)$ is 1/2-periodic, satisfies

$$f(x) > 0 \text{ on } (0, 1/2) \cup (1/2, 1), \text{ and } f(0) = f(1/2) = f(1) = 0.$$

and is symmetric with respect to $x = 1/4$ (and thus $x = 3/4$). Let u_1 and u_2 be 1-periodic and be defined, over a period, as follows:

$$u_1(x) = \begin{cases} \displaystyle\int_0^x f(y)\, dy & 0 \le x \le \dfrac{1}{2} \\ \displaystyle\int_x^1 f(y)\, dy & \dfrac{1}{2} \le x \le 1 \end{cases} \qquad u_2(x) = \begin{cases} \displaystyle\int_0^x f(y)\, dy & 0 \le x \le \dfrac{1}{4} \\ \displaystyle\int_x^{1/2} f(y)\, dy & \dfrac{1}{4} \le x \le \dfrac{1}{2} \\ u_2 \text{ is } \dfrac{1}{2}\text{-periodic.} \end{cases}$$

Both are viscosity solutions of (2.28), and u_2 cannot be obtained from u_1 by the addition a constant. A more subtle mechanism is at work. A very remarkable study of this fact may be found in Lions [85] for a multi-dimensional generalization of (2.28), that is,

$$|\nabla u| = f(x), \quad x \in \Omega$$

where Ω is a bounded open subset of $\mathbb{R}^N$ and f is nonnegtive and vanishes only at a finite number of points. The zero set of f is shown to play an important role: essentially, imposing u at those points ensures uniqueness – but not always existence. Fathi's contribution (which we will describe below in detail) is to understand, for a general H, what is going on.

3 The Lagrangian representation of solutions

The Legendre transform and extremal paths

We already know that there is at most one solution to the Cauchy problem (1.1)-(1.2). In this section, we will recall an "explicit" formula for this solution. In the preceding section, we assumed very little about the Hamiltonian H: only a bit of coercivity and continuity. From now on, we will make the smoothness and strict convexity assumptions given in the introduction: $H(x, p)$ is $C^\infty(\mathbb{T}^n \times \mathbb{R}^n)$, uniformly strictly convex in its second variable, and the existence of $\alpha > 0$ so that

$$D_p^2 H(x, p) \ge \alpha I, \qquad [D_p^2 H(x, p)]_{ij} = \frac{\partial^2 H}{\partial p_i \partial p_j}, \tag{3.1}$$

in the sense of quadratic forms, for all $x \in \mathbb{T}^n$ and $p \in \mathbb{R}^n$.

The Legendre transform of H in p (also called the Lagrangian) is

$$L(x, v) = \sup_{p \in \mathbb{R}^n} (p \cdot v - H(x, p)), \quad x \in \mathbb{T}^n, \quad v \in \mathbb{R}^n. \tag{3.2}$$

Under our assumptions on $H(x, p)$, the Lagrangian $L(x, v)$ is C^∞ in x and v, and is uniformly strictly convex in its second variable. Moreover, we have the duality

$$H(x, p) = \sup_{v \in \mathbb{R}^n} (p \cdot v - L(x, v)),$$

and for all $x \in \mathbb{T}^n$, the mapping $v \mapsto \nabla_v L(x, v)$ is a C^∞-diffeomorphism, with

$$(\nabla_v L)^{-1} = \nabla_p H.$$

One of the standard references for the basic properties of the Legendre transform is [106]. However, given our smoothness assumptions on H, they may be deduced in an elementary fashion: strict convexity ensures the existence of a unique minimum, as well as the equivalence between minima and zeros of the gradient of H in p. The implicit function theorem and a few computations will then do the job.

For every time $t > 0$, and two points $x \in \mathbb{T}^n$ and $y \in \mathbb{T}^n$, we define the function

$$h_t(x, y) = \inf_{\gamma(0)=x, \gamma(t)=y} \int_0^t L(\gamma, \dot{\gamma}) \, ds. \tag{3.3}$$

Here, the infimum is taken over all paths γ on $\mathbb{T}^n$, that are piecewise C^1. The quantity

$$\int_0^t L(\gamma, \dot{\gamma}) \, ds$$

is usually referred to as the *Lagrangian action*, or simply the action. This is a classical minimisation problem, which admits the following result (Tonelli's theorem).

Proposition 3.1 *Given any $(t, x, y) \in \mathbb{R}_+^* \times \mathbb{T}^n \times \mathbb{T}^n$, there exists at least one minimizing path $\gamma(s) \in C^2([0, t]; \mathbb{T}^n)$, such that*

$$h_t(x, y) = \int_0^t L(\gamma, \dot{\gamma}) \, ds.$$

Moreover there is $C(t, |x - y|) > 0$ such that

$$\|\dot{\gamma}\|_{L^\infty([0,t])} + \|\ddot{\gamma}\|_{L^\infty([0,t])} \le C(t, |x - y|).$$

The function C tends to $+\infty$ as $t \to 0$ – keeping $|x - y|$ fixed. The function $\gamma(t)$ solves the Euler-Lagrange equation

$$\frac{d}{ds} \nabla_v L(\gamma, \dot{\gamma}) - \nabla_x L(\gamma, \dot{\gamma}) = 0. \tag{3.4}$$

Once again, we leave the proof as an exercise but give a hint for the proof: consider a minimizing sequence γ_n. First, use the strict convexity of L to obtain the H^1-estimates for γ_n, thus ensuring compactness in the space of continuous paths and weak convergence to $\gamma \in H^1([0, t])$ with fixed ends. Next, show that the convexity of L implies that γ is, indeed, a minimizer. Finally, derive the Euler-Lagrange equation and show that γ is actually C^∞. Such a curve γ is called an *extremal*.

Back to the Hamilton-Jacobi equations

We now relate the solutions of the Cauchy problem for the Hamilton-Jacobi equations to the Lagrangian. Given the initial data $u_0 \in C(\mathbb{T}^n)$, define the function

$$u(t, x) = \mathcal{T}(t) u_0(x) = \inf_{y \in \mathbb{T}^n} (u_0(y) + h_t(y, x)). \tag{3.5}$$

Exercise 3.2 *Show that the infimum in (3.5) is attained. Also show that* $(\mathcal{T}(t))_{t>0}$ *is a semi-group: one has*

$$\mathcal{T}(t+s)u_0 = \mathcal{T}(t)\mathcal{T}(s),$$

that is,

$$u(t,x) = \inf_{y\in\mathbb{T}^n}(u(s,y) + h_{t-s}(y,x)), \tag{3.6}$$

for all $0 \le s \le t$*, and* $\mathcal{T}(0) = I$*.*

This semigroup is referred to as the *Lax-Oleinik semigroup.* Here is its link to the Hamilton-Jacobi equations.

Theorem 3.3 *The function* $u(t,x) := \mathcal{T}(t)u_0(x)$ *is the unique solution to the Cauchy problem*

$$\begin{aligned} &u_t + H(x,\nabla u) = 0, \\ &u(0,x) = u_0(x). \end{aligned} \tag{3.7}$$

Proof. Let us first show the super-solution property: take $t_0 > 0$ and $x_0 \in \mathbb{T}^n$ and let ϕ be a test function such that (t_0,x_0) is a minimum for $u-\phi$. Without loss of generality, we may assume that $u(t_0,x_0) = \phi(t_0,x_0)$. Consider y_0 such that

$$u(t_0,x_0) = u_0(y_0) + h_{t_0}(y_0,x_0).$$

Let also γ be an extremal of the action between the times $t=0$ and $t=t_0$, going from y_0 to x_0: $\gamma(0)=y_0$, $\gamma(t_0)=x_0$. We have, for all $0\le t\le t_0$:

$$\phi(t,\gamma(t)) \le u(t,\gamma(t)) \le u_0(y_0) + \int_0^t L(\gamma,\dot\gamma)\,ds,$$

both $\le$ being an equality for $t=t_0$, which implies

$$\frac{d}{dt}\left(u_0(y_0) + \int_0^t L(\gamma,\dot\gamma)\,ds - \phi(t,\gamma(t))\right)\Big|_{t=t_0} \le 0,$$

or, in other words

$$\phi_t(t_0,x_0) + \dot\gamma(t_0)\cdot\nabla\phi(t_0,x_0) - L(\gamma(t_0),\dot\gamma(t_0)) \ge 0.$$

Which implies, via (3.2):

$$\phi_t(t_0,x_0) + H(x_0,\nabla\phi(t_0,x_0)) \ge 0.$$

To show the sub-solution property, consider a test function $\phi(t,x)$, as well as $t_0>0$ and $x_0\in\mathbb{T}^n$, where the difference $u-\phi$ attains its maximum, and assume, once again, that $u(t_0,x_0)=\phi(t_0,x_0)$. Given $v\in\mathbb{R}^n$, define the curve

$$\gamma(s) = x_0 + (t_0 - s)v.$$

Using the semigroup property (3.6), we obtain, for all $t\le t_0$, since $u(t,x)\le\phi(t,x)$ for all $0\le t\le t_0$ and $x\in\mathbb{T}^n$:

$$\begin{aligned} u(t_0,x_0) &\le u(t,x_0-(t_0-t)v) + h_{t_0-t}(x-(t_0-t)v,x_0) \\ &\le \phi(t,x_0-(t_0-t)v) + h_{t_0-t}(x-(t_0-t)v,x_0), \end{aligned}$$

both inequalities becoming an equality at $t = t_0$. Just as above, differentiating in t at $t = t_0$ gives

$$\phi_t(t_0, x_0) + v \cdot \nabla\phi(t_0, x_0) - L(x_0, v) \le 0,$$

that is

$$\phi_t(t_0, x_0) + H(x_0, \nabla\phi(t_0, x_0)) \le 0,$$

by (3.2), beacuse v is arbitrary. □

Instant regularization to Lipschitz

We conclude this section with a remarkable result on instant smoothing. We will show that if the initial datum u_0 is continuous on $\mathbb{T}^n$, the solution of the Cauchy problem $u(t, x)$ becomes instantaneously Lipschitz. The improved regularity comes from the strict convexity of the Hamiltonian: indeed, nothing of that sort is true for the eikonal equation, as can be seen from the following example. Consider the initial value problem

$$u_t + |u_x| = 0, \quad u(0, x) = u_0(x), \quad x \in \mathbb{T}^1$$

whose solution is

$$u(t, x) = \inf_{|x-y|\le t} u_0(y).$$

If $u_0(x)$ is not Lipschitz in x, neither is $u(t, x)$. On the other hand, if the Hamiltonian is strictly convex we have the following result.

Theorem 3.4 *Let $u(t, x)$ be the unique solution to the Cauchy problem*

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \\ &u(0, x) = u_0(x), \end{aligned} \tag{3.8}$$

with $u_0 \in C(\mathbb{T}^n)$. For all $t > 0$, there is $C_t > 0$ such that $|\mathrm{Lip}\ u(t, x)| \le C_t$ on $[t, +\infty) \times \mathbb{T}^n$. The constant C_t does not depend on the initial condition and blows up as $t \downarrow 0$.

Proof. It is sufficient to consider time intervals of length one, and repeat the argument on the subsequent intervals. Given $0 < t \le 1$, and $x \in \mathbb{T}^n$, consider the extremal γ such that

$$u(t, x) = u_0(\gamma(0)) + \int_0^t L(\gamma, \dot\gamma)\ ds.$$

Take $h \in \mathbb{R}^n$ (we may always assume $x + h \in \mathbb{T}^n$), and define the curve

$$\widetilde\gamma(s) = \gamma(s) + \frac{s}{t}h, \quad 0 \le s \le t,$$

so that $\widetilde\gamma(0) = \gamma(0)$ and $\widetilde\gamma(t) = x + h$. We have

$$\begin{aligned} \int_0^t (L(\widetilde\gamma, \dot{\widetilde\gamma}) - L(\gamma, \dot\gamma))\ ds = & \int_0^t [L(\gamma(s) + \frac{s}{t}h, \dot\gamma(s) + \frac{1}{t}h) - L(\gamma(s), \dot\gamma(s))]\ ds \\ \le & \int_0^t \frac{1}{t}\Big(sh \cdot \nabla_x L(\gamma, \dot\gamma) + h \cdot \nabla_v L(\gamma, \dot\gamma)\Big)\ ds + C_t|h|^2. \end{aligned}$$

We may now use the Euler-Lagrange equation

$$\frac{d}{ds}\nabla_v L(\gamma(s), \dot{\gamma}(s)) - \nabla_x L(\gamma(s), \dot{\gamma}(s)) = 0$$

to rewrite the last line above as

$$\int_0^t (L(\widetilde{\gamma}, \dot{\widetilde{\gamma}}) - L(\gamma, \dot{\gamma}))\ ds \le h \cdot \nabla_v L(\gamma(t), \dot{\gamma}(t)) + C_t |h|^2.$$

It is now sufficient to write, in view of formula (3.5) for $u(t, x)$:

$$u(t, x+h) = u(t, \widetilde{\gamma}(t)) \le u(\gamma(0)) + \int_0^t L(\widetilde{\gamma}, \dot{\widetilde{\gamma}})\ ds = u(t, x) + \int_0^t (L(\widetilde{\gamma}, \dot{\widetilde{\gamma}}) - L(\gamma, \dot{\gamma}))\ ds.$$

We obtain

$$u(t, x+h) - u(t, x) \le h \cdot \nabla_v L(\gamma(t), \dot{\gamma}(t)) + K|h|^2, \tag{3.9}$$

which proves the Lipschitz regularity in space, because both γ and $\dot{\gamma}$ are bounded, as long as $t \ge t_0 > 0$.

In order to prove the Lipschitz regularity in time, let us examine a small variation of t, denoted by $t + \tau$ with $t + \tau > 0$. Perturbing the extremal γ into

$$\widetilde{\gamma}(s) = \gamma(\frac{t}{t+\tau} s),$$

we still have $\widetilde{\gamma}(0) = \gamma(0)$, $\widetilde{\gamma}(t+\tau) = \gamma(t) = x$. The same computation as above gives $u(t+\tau, x) - u(t, x) \le C_t |\tau|$, hence the result. □

Remark 3.5 (i) It follows that $u(t, x)$ is almost everywhere differentiable.
(ii) Take $t > 0$ and $\gamma(s)$ an extremal such that u is differentiable at $x := \gamma(t)$. We have then

$$\nabla u(t, x) = \nabla u(t, \gamma(t)) = \nabla_v L(x, \dot{\gamma}(t)).$$

The Hamilton-Jacobi equation yields

$$u_t(t, \gamma(t)) = -H(x, \nabla u(t, x)) = -H(\gamma(t), \nabla u(t, \gamma(t))).$$

4 Semi-concavity and $C^{1,1}$ regularity

As we have mentioned, equation (1.1) may have several Lipschitz solutions, so it is worth asking whether the unique viscosity solution has additional regularity features. A relevant notion is that of semi-concavity.

Semi-concavity

Definition 4.1 *If B is an open ball in $\mathbb{R}^n$, F a closed subset of B and K a positive constant, we say that $u \in C(B)$ is K-semi-concave on F if: for all $x \in F$, there is $l_x \in \mathbb{R}^n$ such that for all $h \in \mathbb{R}^n$ satisfying $x + h \in B$, we have:*

$$u(x+h) \leq u(x) + l_x \cdot h + K|h|^2. \tag{4.1}$$

The function u is said to be K-semi convex on F if $-u$ is K-semi-concave on F.

The next theorem is crucial for the sequel. If u is continuous in an open ball B in $\mathbb{R}^n$, and F is a closed set of B, we say that $u \in C^{1,1}(F)$ if u is differentiable in F and ∇u is Lipschitz over F.

Theorem 4.2 *Let B be an open ball of $\mathbb{R}^n$ and F closed in B. Assume $u \in C(B)$ is K-semi-concave and K-semi-convex in F, then $u \in C^{1,1}(F)$.*

Proof. For all $x \in F$, there are two vectors l_x and m_x such that:

$$\forall h \in \mathbb{R}^N, \qquad \begin{array}{rl} u(x+h) \leq & u(x) + l_x \cdot h + K|h|^2 \\ u(x+h) \geq & u(x) + m_x \cdot h - K|h|^2 \end{array}$$

which yields, after subtracting:

$$(l_x - m_x) \cdot h \leq 2K|h|^2.$$

As this is true for all h, we conclude that $l_x = m_x$ and, therefore, u is differentiable at x.

Consider then $(x, y, h) \in F \times F \times \mathbb{R}^n$. The semi-convexity and semi-concavity inequalities, written, respectively, between $x + h$ and x, x and y, $x + h$ and y, give:

$$\begin{array}{l} |u(x+h) - u(x) - \nabla u(x) \cdot h| \leq K|h|^2 \\ |u(x) - u(y) - \nabla u(y) \cdot (x-y)| \leq K|x-y|^2 \\ |u(y) - u(x+h) + \nabla u(y) \cdot (x+h-y)| \leq K|x+h-y|^2. \end{array}$$

Adding the three inequalities, we obtain:

$$|(\nabla u(x) - \nabla u(y)) \cdot h| \leq 3K(|h|^2 + |x-y|^2). \tag{4.2}$$

Taking

$$h = |x-y| \frac{\nabla u(x) - \nabla u(y)}{|\nabla u(x) - \nabla u(y)|},$$

inequality (4.2) becomes

$$|(\nabla u(x) - \nabla u(y))| \leq 3K|x-y|,$$

which is the Lipschitz property of ∇u that we sought. $\square$

Improved regularity of the viscosity solutions

Let us come back to the solution $u(t,x)$ of the Cauchy problem

$$\begin{aligned} &u_t + H(x, \nabla u) = 0, \\ &u(0,x) = u_0(x). \end{aligned} \tag{4.3}$$

Choose $t > 0$ and $x \in \mathbb{T}^n$, and an extremal curve γ such that

$$u(t, x = \gamma(t)) = u_0(\gamma(0)) + \int_0^t L(\gamma, \dot\gamma)\ ds.$$

We also have, for all $s \in (0,t)$:

$$u(s, \gamma(s)) \le u_0(\gamma(0)) + \int_0^s L(\gamma, \dot\gamma)\ d\sigma.$$

Imagine that for some $0 \le s \le t$, we have

$$u(s, \gamma(s)) < u_0(\gamma(0)) + \int_0^s L(\gamma, \dot\gamma)\ d\sigma.$$

That is, there exists a curve $\gamma_1(s')$, $0 \le s' \le s$, such that $\gamma_1(s) = \gamma(s)$, and

$$u_0(\gamma_1(0)) + \int_0^s L(\gamma, \dot\gamma)\ d\sigma < u_0(\gamma(0)) + \int_0^s L(\gamma, \dot\gamma)\ d\sigma.$$

Then, we would consider the concatenated curve $\widetilde\gamma(s)$ so that $\widetilde\gamma(s') = \gamma_1(s')$ for $0 \le s' \le s$, and $\widetilde\gamma(s') = \gamma(s')$ for $s \le s' \le t$. We would have

$$u(t, \gamma(t)) = u_0(\gamma(0)) + \int_0^s L(\gamma, \dot\gamma)\ d\sigma + \int_s^t L(\gamma, \dot\gamma) d\sigma > u_0(\widetilde\gamma(0)) + \int_0^t L(\widetilde\gamma, \dot{\widetilde\gamma})\ ds,$$

which would contradict the extremal property of the curve γ between the times 0 and t. So, for all $0 \le s \le s' \le t$ we have:

$$u(s', \gamma(s')) = u_0(\gamma(0)) + \int_0^{s'} L(\gamma, \dot\gamma)\ d\sigma = u(s, \gamma(s)) + \int_s^{s'} L(\gamma, \dot\gamma)\ d\sigma. \tag{4.4}$$

Definition 4.3 *We say that $\gamma : [0,t] \to \mathbb{T}^n$ is calibrated by u.*

Let us define the conjugate semigroup of the Lax-Oleinik semigroup by:

$$\forall u_0 \in C(\mathbb{T}^n), \quad \widetilde{\mathcal{T}}(t) u_0(x) = \sup_{y \in \mathbb{T}^n} (u_0(y) - h_t(x,y)). \tag{4.5}$$

We will often denote $\widetilde u(t,x) = \widetilde{\mathcal{T}}(t) u_0(x)$. The following lemma is proved exactly as Theorem 3.4:

Lemma 4.4 *Let $u_0 \in C(\mathbb{T}^n)$ and $\sigma > 0$. There is $K(\sigma) > 0$ such that $\widetilde{\mathcal{T}}(\sigma) u_0$ is $K(\sigma)$-semiconvex. The constant $K(\sigma)$ blows up as $\sigma \to 0$.*

Given $0 < s < s'$, we define the set $\Gamma_{s,s'}[u_0]$ as follows: it is the union of all points $x \in \mathbb{T}^n$ so that the extremal calibrated by u, which passes through the point x at the time s can be continued until the time s'.

Corollary 4.5 *Let $u(t,x) = \mathcal{T}(t)u_0(x)$, and $0 < s_1 < s_2$, then for any $0 < \varepsilon < s_1$, the function $u \in C^{1,1}([s_1, s_2] \times \Gamma_{s_1,s_2+\varepsilon})$.*

Proof. We are under the assumptions of Theorem 3.4, so there is $K > 0$ depending on s_1 such that inequality (3.9) is true:

$$u(s, x+h) - u(s, x) \leq h \cdot \nabla_v L(\gamma(s), \dot{\gamma}(s)) + K|h|^2. \tag{4.6}$$

The function $u(s, \cdot)$ is thus K-semi-concave on Γ_s for all $s_1 \leq s \leq s_2$. Furthermore, note that for all $y \in \mathbb{R}^n$ we have

$$u(s_2 + \varepsilon, y) \leq u(s, \gamma(s)) + h_{s_2+\varepsilon-s}(\gamma(s), y),$$

and the calibration relation (4.4) implies that equality is attained when $y = \gamma(s_2+\varepsilon)$. Here, we use the assumption that the extremal γ can be continued past the time s, until the time $s_2+\varepsilon$. We conclude that, for all $s_1 \leq s \leq s_2$:

$$u(s, \gamma(s)) = \sup_{y \in \mathbb{T}^n} (u(s_2 + \varepsilon, y) - h_{s_2+\varepsilon-s}(\gamma(s), y)) = \widetilde{\mathcal{T}}(s_2 + \varepsilon - s)u(s_2 + \varepsilon, \cdot)(\gamma(s)).$$

It follows now from Lemma 4.4 that there is a constant $\widetilde{K}$ depending on ε, such that $u(s, \cdot)$ is $\widetilde{K}$-semi-convex in x on $[s_1, s_2] \times \Gamma_{s_1,s_2+\varepsilon}$. We conclude from Theorem 4.2 that the function $u(s, \cdot)$ is $C^{1,1}$ in x on Γ_s. To end the proof, one just has to invoke Remark 3.5 to obtain the corresponding regularity in the time variable. □

5 Critical sub-solutions and the Fathi-Siconolfi theorem

5.1 Sub-solutions of the steady problem

Basic properties of the sub-solutions

We now shift from the Cauchy problem to the main subject of this chapter, the steady problem (1.3):

$$H(x, \nabla u) = c. \tag{5.1}$$

Note that, without loss of generality, we may assume $c = 0$: we just have to redefine H as $H - c$. Hence, we study "the true steady equation"

$$H(x, \nabla u) = 0, \quad x \in \mathbb{T}^n. \tag{5.2}$$

Warning. We will from now on always assume that the constant c is chosen so that (5.2) has solutions. In this section we do not worry about the existence issue and assume that the Hamiltonian is such that the solutions exist.

A solution u of (5.2) is a solution of the Cauchy problem for

$$u_t + H(x, \nabla u) = 0,$$

with itself as the initial data, in other words:

$$u(x) = \mathcal{T}(t)u(x), \tag{5.3}$$

for all $t \geq 0$. The proof of Corollary 4.5 (see (4.6) above) implies the semi-concavity of u. Moreover, if $x \in \mathbb{T}^n$ lies on an extremal calibrated by u, so that $x = \gamma(s)$ with $s > 0$, and this extremal can be extended until a time $s' > s$, then u is $C^{1,1}$ at x, with the corresponding regularity constants depending only on s and $s' - s$.

Note that (5.3) means that

$$u(x) = \inf_{y \in \mathbb{T}^n} [u(y) + h_t(y, x)], \tag{5.4}$$

for all $t \geq 0$. It follows that

$$u(x) - u(y) \leq h_t(y, x), \tag{5.5}$$

for all $x, y \in \mathbb{T}^n$ and $t \geq 0$. The following criterion, in the spirit of (5.5), will be useful to decide whether or not a function is a subsolution to (5.2).

Proposition 5.1 *A function $u \in C(\mathbb{T}^n)$ is a subsolution to* (5.2) *if and only if*

$$u(x) - u(y) \leq h_t(y, x), \tag{5.6}$$

for all $t > 0$ and $x, y \in \mathbb{T}^n$. Here, $h_t(y, x)$ is the minimum of the action between y and x, and is given by (3.3)*.*

Proof. First, assume that $u(x)$ is a sub-solution, and consider the solution of the Cauchy problem

$$\begin{aligned} &v_t + H(x, \nabla v) = 0 \\ &v(0, x) = u(x). \end{aligned} \tag{5.7}$$

As $u(x)$ is a sub-solution also to the time-dependent problem, we have $v(t, x) \geq u(x)$ for all $t \geq 0$ and $x \in \mathbb{R}^n$, by the maximum principle. However, $v(t, x)$ is given by

$$v(t, x) = \inf_{y \in \mathbb{T}^n} [u(y) + h_t(y, x)].$$

It follows that

$$u(x) \leq v(t, x) \leq u(y) + h_t(y, x),$$

for all $x, y \in \mathbb{T}^n$ and $t \geq 0$.

On the other hand, if $u(y)$ satisfies (5.6), we may simply follow the corresponding part of the proof of Theorem 3.3. Consider a function $\phi(x)$ such that the difference $u - \phi$ attains its maximum at a point x_0. Again, without loss of generality we may assume that $u(x_0) = \phi(x_0)$. Given $v \in \mathbb{R}^n$ and $t > 0$, define the curve

$$\gamma(s) = x_0 - (t - s)v, \quad 0 \leq s \leq t,$$

so that $\gamma(0) = x_0 - tv$, and $\gamma(t) = x_0$, and write

$$\begin{aligned} u(x_0) &\leq u(x_0 - tv) + h_t(x_0 - tv, x_0) \leq \phi(x_0 - tv) + h_t(x_0 - tv, x_0) \\ &\leq \phi(x_0 - tv) + \int_0^t L(x_0 - (t - s)v, v) ds. \end{aligned}$$

All inequalities above are equalities at $t = 0$. Differentiating in t at $t = 0$ gives

$$-v \cdot \nabla\phi(x_0) + L(x_0, v) \geq 0.$$

As this inequality holds for all $v \in \mathbb{R}^n$, we conclude that

$$H(x_0, \nabla\phi) = \sup_{v\in\mathbb{R}^n} [v \cdot \nabla\phi - L(x_0, v)] \leq 0,$$

whence $u(x)$ is, indeed, a sub-solution. □

Next, we give a criterion (in terms of the extremal curves) for a sub-solution to be a solution.

Proposition 5.2 *A subsolution to* (5.2) *is a solution of this equation in an open subset U of $\mathbb{T}^n$ if and only if: for all $x \in U$, there is $t_x > 0$ and a curve $\gamma : [0, t_x] \to \mathbb{T}^n$, such that $\gamma(t_x) = x$, and*

$$u(x) = u(\gamma(0)) + \int_0^{t_x} L(\gamma, \dot\gamma)ds. \tag{5.8}$$

Proof. The proof follows that of Theorem 3.3 – we will show that if (5.8) holds then $u(x)$ is not only a sub-solution but also a super-solution. Let ϕ be a function such that $u - \phi$ attains its minimum at a point $x_0 \in \mathbb{T}^n$, and $\phi(x_0) = u(x_0)$, and assume that $x_0 \in U$, so that there exist a curve γ, and a time t_0 so that $\gamma(t_0) = x_0$

$$u(x_0) = u(\gamma(0)) + \int_0^{t_0} L(\gamma, \dot\gamma)ds. \tag{5.9}$$

Then we have, as u is a sub-solution:

$$\phi(\gamma(t)) \leq u(\gamma(t)) \leq u(\gamma(0)) + \int_0^t L(\gamma, \dot\gamma)ds,$$

with the equality at $t = t_0$. Differentiating at $t = t_0$ we get

$$\dot\gamma(t_0) \cdot \nabla\phi(x_0) - L(x_0, \dot\gamma(t_0)) \geq 0,$$

whence

$$H(x_0, \nabla\phi_0) \geq 0.$$

Therefore, $u(x)$ is a both a sub- and super-solution, hence a solution on the open set U. □

Proposition 5.3 *For every solution u of* (5.2)*, and for all $x \in \mathbb{T}^n$, there is an "eternal" extremal $\gamma : (-\infty, 0] \to \mathbb{T}^n$, such that $\gamma(0) = x$ and*

$$\forall t > 0, \quad u(x) - u(\gamma(-t)) = \int_{-t}^0 L(\gamma, \dot\gamma)\ ds.$$

Proof. Consider, for all $T > 0$, an extremal $\gamma_T : [-T, 0] \to \mathbb{T}^n$ calibrating u and such that $\gamma_T(0) = x$. The family γ_T is relatively compact in the C^1_{loc} topology – this follows from Proposition 3.1. Extraction of a subsequence converging in $C^1_{loc}(\mathbb{R}_-)$ does the job. □

The critical sub-solutions

If solutions of (5.2) exist, then

$$H(x, \nabla u) \leq c$$

admits sub-solutions (but not solutions) for all $c > 0$, but not for $c < 0$. This motivates the following.

Definition 5.4 *A subsolution of* (5.2) *is called a critical subsolution. A critical subsolution u is said to be strict at a point $x \in \mathbb{T}^n$ if there exists U, open neighbourhood of x and $c_x < 0$ such that u solves $H(x, \nabla u) \leq c_x$ in U, in the viscosity sense. A critical subsolution is strict in an open subset U of $\mathbb{T}^n$ if it is strict at each point of U.*

5.2 The Mañé potential and the Peierls barrier

We introduce now the basic players in the study of the critical sub-solutions: the Mañé potential and the Peierls barrier, and describe some of their basic properties. They are defined as follows:

$$\phi(x, y) = \inf_{t>0} h_t(x, y), \qquad h(x, y) = \liminf_{t\to+\infty} h_t(x, y). \tag{5.10}$$

The function ϕ is known as the Mañé potential, and the function h as the Peierls barrier. Note that both are finite: to see that, we first observe that the function $h_t(x, y)$ is uniformly bounded from below. Indeed, as throughout this section we assume that solutions of

$$H(x, \nabla u) = 0, \tag{5.11}$$

exist, we automatically have

$$u(x) - u(y) \leq h_t(x, y),$$

thus $h_t(x, y) \geq -2\|u\|_{L^\infty}$. On the other hand, it is easy to see that there exists C_0 so that $|h_{t=1}(x, y)| \leq C_0$ for all $x, y \in \mathbb{T}^n$. This immediately implies that $\phi(x, y)$ is finite. To see that $h(x, y)$ is finite, note that for each $x \in \mathbb{T}^n$ and $t > 0$ there exists $\bar{y}_t(x)$ so that

$$u(x) = u(\bar{y}_t(x)) + h_t(x, \bar{y}_t(x)),$$

and thus

$$h_t(x, \bar{y}_t(x)) \leq 2\|u\|_{L^\infty}.$$

Hence, we can write

$$h_t(x, y) \leq h_{t-1}(x, \bar{y}_{t-1}(x)) + h_{t=1}(\bar{y}_{t-1}(x), y) \leq 2\|u\|_{L^\infty} + C_0,$$

and we conclude that $h(x, y)$ is also finite if a solution of (5.11) exists.

The definition of $\phi(x, y)$ and $h(x, y)$ as "the smallest cost of transport disregarding the time" from x to y shows that

$$\begin{aligned} \phi(x, z) &\leq \phi(x, y) + h_t(y, z), \\ h(x, z) &\leq h(x, y) + h_t(y, z), \end{aligned} \tag{5.12}$$

for all $t \in \mathbb{R}$ and $x, y, z \in \mathbb{T}^n$. It follows that $\phi_x(y) = \phi(x, y)$ and $h_x(y) = h(x, y)$ are (as functions of $y \in \mathbb{T}^n$ with x fixed) Lipschitz sub-solutions of (5.2). In fact, more is true: h_x is, actually, a solution (hence, for it to be finite the Hamiltonian has to be such that at least one solution of (5.11) exists).

Proposition 5.5 *For all $x \in \mathbb{T}^n$, the function h_x is a solution of* (5.2) *over $\mathbb{T}^n$. The function ϕ_x is a solution of* (5.2) *over $\mathbb{T}^n \backslash \{x\}$.*

Proof. To show that h_x is a solution, consider a point $y \in \mathbb{T}^n$, and a sequence $t_n \to +\infty$ such that $h_{t_n}(x, y)$ tends to $h(x, y) = h_x(y)$. Consider also a sequence of extremals γ_n achieving the minimum of the action between x and y in the time t_n, such that $\gamma_n(t_n) = y$, as well as the time shifts $\widetilde{\gamma}_n(s) = \gamma_n(t_n + s)$, $-t_n \leq s \leq$. Then, for any $0 \leq t \leq t_n$ we have

$$h_{t_n}(x, y) = \int_{-t_n}^{0} L(\widetilde{\gamma}_n, \dot{\widetilde{\gamma}}_n)\ ds, \quad h_{t_n}(x, \widetilde{\gamma}_n(-t)) \leq \int_{-t_n}^{-t} L(\widetilde{\gamma}_n, \dot{\widetilde{\gamma}}_n)\ ds,$$

so that

$$h_{t_n}(x, \widetilde{\gamma}_n(-t)) + \int_{-t}^{0} L(\widetilde{\gamma}_n, \dot{\widetilde{\gamma}}_n)\ ds \leq h_{t_n}(x, y).$$

Using the relative compactness of the sequence $\widetilde{\gamma}_n$ in the C^1_{loc} topology, and the fact that h is a lim inf, we may pass to the limit $n \to +\infty$ above:

$$h(x, \gamma(-t)) + \int_{-t}^{0} L(\gamma, \dot{\gamma})\ ds \leq h(x, y),$$

where $\gamma : (-\infty, 0]$ is an eternal extremal which is the local uniform limit of $\widetilde{\gamma}_n$. It follows that

$$h(x, y) \geq h(x, \gamma(-t)) + h_t(\gamma(-t), y),$$

Because h_x is a subsolution, we obtain

$$h_x(y) = h_x(\gamma(-t)) + h_t(\gamma(-t), y),$$

and

$$h(x, \gamma(-t)) + \int_{-t}^{0} L(\gamma, \dot{\gamma})\ ds = h(x, y).$$

Invoking Proposition 5.2 shows that $h_x(y)$ is a solution.

To show that ϕ_x is a solution on $\mathbb{T}^n \setminus \{x\}$, consider once again $y \in \mathbb{T}^n$, $y \neq x$. It is enough to assume that the inf in formula (5.10) is attained at some finite time $t_0 > 0$: if not, we have $\phi_x(y) = h_x(y)$, and the previous result applies. Thus, $\phi_x(y) < h_x(y)$, hence there exists $t_0 > 0$ so that $\phi_x(y) = h_{t_0}(x, y)$, and, in particular, there exists an extremal γ so that $\gamma(0) = x$, $\gamma(t_0) = y$, and

$$\phi_x(y) = \int_{0}^{t_1} L(\gamma, \dot{\gamma}) ds.$$

Moreover, the same is true for z close enough to y. As $t_0 > 0$, we may take $z = \gamma(t_1)$ for some time $0 < t_1 < t_0$, sufficiently close to t_0. It is easy to see that the curve $\gamma(s)$, $0 \leq s \leq t_1$

realizes the infimum for the point z as well (otherwise it would not realize the minimum for the point y), so that

$$\phi_x(z) = \int_0^{t_1} L(\gamma, \dot{\gamma}) ds.$$

Consider the piece of γ that connects z and y: $\gamma_1(s) = \gamma(t+s)$, $0 \le s \le t_0 - t_1$, then

$$\phi_x(y) = \phi_x(\gamma_1(0)) + \int_0^{t_0 - t_1} L(\gamma_1, \dot{\gamma}_1) ds,$$

thus Proposition 5.2, once again, implies that $\phi_x(y)$ is a solution, not just a sub-solution. □

Let us close this initial presentation of the Mañé potential and the Peierls barrier by the following formulae, which are once again simple consequences of (5.10), (5.12), and of relation (5.6) for any solution of equation (5.2): first, for all $x \in \mathbb{T}^n$ we have

$$h(x,x) \ge 0, \;\; \phi(x,x) = 0, \tag{5.13}$$

and, second:

$$\forall x, y \in \mathbb{T}^n, \qquad h(x,y) + h(y,x) \ge 0. \tag{5.14}$$

5.3 The Fathi-Siconolfi theorem and uniqueness sets

We define the *Aubry set* as

$$\mathcal{A} = \{x \in \mathbb{T}^n : \; h(x,x) = 0\}. \tag{5.15}$$

The Aubry set is closed (recall that h is Lipschitz). The PDE interest in the Aubry set comes from its connection to the uniqueness of the solutions to

$$H(x, \nabla u) = 0. \tag{5.16}$$

We say that a closed set $B \in \mathbb{T}^n$ is a uniqueness set for a Hamiltonian $H(x,p)$ if the Dirichlet problem

$$H(x, \nabla u) = 0, \quad x \in \mathbb{T}^n \backslash B, \tag{5.17}$$

with u imposed on B, admits *at most* one solution. Furthermore, B is a strong uniqueness set for (5.16) if given an u.s.c. subsolution $\underline{u}$, and an l.s.c supersolution $\overline{u}$ to (5.16), with $\underline{u} \le \overline{u}$ on B, then $\underline{u} \le \overline{u}$ on $\mathbb{T}^n$. The PDE interest of the Aubry set comes from the following.

Theorem 5.6 *The Aubry set $\mathcal{A}$ is a strong uniqueness set for* (5.2).

As an example, let us see what happens for the classical mechanics Hamiltonian

$$H(x,p) = |p|^2 - f(x).$$

Here, f is a positive potential with a nontrivial zero set. Equation

$$|\nabla u(x)|^2 - f(x) = 0 \tag{5.18}$$

admits solutions – and thus $c = 0$. In order to see that, consider the Cauchy problem

$$\begin{aligned} &v_t + |\nabla v(x)|^2 - f(x) = 0, \\ &v(0,x) = 0. \end{aligned} \tag{5.19}$$

Exercise 5.7 *Prove that $v(t,x) \geq 0$, and the function $v(t,x)$ is monotonically increasing in time.*

In addition to the monotonicity of $v(t,x)$ in time, we also know that $v(t,x) = 0$ on the set $\{f = 0\}$.

Exercise 5.8 *Prove that (5.18) has a solution, which is the limit of $v(t,x)$ as $t \to +\infty$.*

The Lagrangian is

$$L(x,v) = \frac{|v|^2}{4} + f(x),$$

thus

$$h(x,y) = \liminf_{t\to+\infty} \inf_{\gamma} \int_0^t \left(\frac{|\dot\gamma|^2}{4} + f(\gamma) \right) ds \geq 0.$$

We have $h(x,x) = 0$ if and only if $f(x) = 0$. Indeed, if $f(x) = 0$, we trivially have $h(x,x) = 0$. On the other hand, if $f(x) > 0$, it follows from Proposition 3.1 that the velocity of an extremal is bounded, hence $h(x,x) > 0$ as soon as $f(x) > 0$. On the other hand, the constants provide a family of smooth critical subsolutions, which are strict outside of the Aubry set $\{h(x,x) = 0\} = \{f(x) = 0\}$.

Exercise 5.9 *What happens if the Hamiltonian is even in p: $H(x,p) = H(x,-p)$?.*

Exercise 5.10 *(Lions [85]). If f is smooth over $\mathbb{T}^n$ and has d zeroes, find all the viscosity solutions of*

$$|\nabla u(x)| = f(x), \quad x \in \mathbb{T}^n.$$

In general, the situation is less obvious. There is, however, an easily obtained family of C^∞ sub-solutions of the super-critical equation

$$H(x, \nabla v) = c \quad \text{with} \quad c > 0,$$

starting with a solution of

$$H(x, \nabla u) = 0.$$

Indeed, if ρ_ε is an approximation of the identity, we have

$$\begin{aligned} H(x, \rho_\varepsilon * u) &= H\Big(x, \int_{\mathbb{T}^n} \rho_\varepsilon(y) \nabla u(x-y)\, dy\Big) \leq \int_{\mathbb{T}^n} \rho_\varepsilon(y) H(x, \nabla u(x-y))\, dy \quad \text{(by Jensen)} \\ &\leq O(\varepsilon) + \int_{\mathbb{T}^n} \rho_\varepsilon(x-y) H(x-y, Du(x-y))\, dy = O(\varepsilon). \end{aligned} \tag{5.20}$$

It is easy to see that the critical value $c = 0$ cannot be attained with this argument. However, it is not useless to mention it here, because it will reappear later.

Strong uniqueness and strict sub-solutions

The following lemma (due to Barles [6]) is a criterion for a set to be a strong uniqueness set, and will be the basis for the proof of Theorem 5.6.

Lemma 5.11 *Assume there exists a C^1 subsolution of the steady equation* (5.16)*, which is strict outside a set $B \subset \mathbb{T}^n$. Then, B is a strong uniqueness set for* (5.16)*, that is: if $\underline{u}$ (resp. $\overline{u}$) is an u.s.c. subsolution (resp. a l.s.c supersolution) to* (5.16)*, with $\underline{u} \le \overline{u}$ on B, then $\underline{u} \le \overline{u}$.*

Proof. Assume that there exists $\bar{x}$ such that $\overline{u}(\bar{x}) < \underline{u}(\bar{x})$ – we must have $\bar{x} \notin B$. Let ψ be the critical C^1 sub-solution, strict outside B.

Exercise 5.12 *Let $u(x)$ be a sub-solution, and $\psi(x)$ a C^1 strict sub-solution to (5.16). Then, if the Hamiltonian $H(x,p)$ is convex, the function*

$$u(x;t) = (1-t)\underline{u}(x) + t\psi(x)$$

is also a strict subsolution outside B, for all $t \in (0,1]$.

In addition, we have

$$\overline{u}(x) < u(x;t), \tag{5.21}$$

in a neighborhood of $\bar{x}$, when $t > 0$ is sufficiently small. We will now apply the argument of Proposition 2.7. Define the function

$$u_\varepsilon(x,y) = \overline{u}(x) - u(y;t) + \frac{|x-y|^2}{2\varepsilon^2},$$

and let $(x_\varepsilon, y_\varepsilon)$ be the minimum for u_ε over $\mathbb{T}^n \times \mathbb{T}^n$. One immediately sees that

$$|x_\varepsilon - y_\varepsilon| \to 0, \quad \text{as } \varepsilon \to 0, \tag{5.22}$$

and, of course, the family $(x_\varepsilon, y_\varepsilon)$ converges, along a subsequence to a limit (x_0, x_0) (which depends on $t \in (0,1)$.

Consider the function

$$\eta(x) = u(y_\varepsilon; t) - \frac{|x-y_\varepsilon|^2}{2\varepsilon^2},$$

then the difference

$$\overline{u}(x) - \eta(x) = u_\varepsilon(x, y_\varepsilon)$$

attains its minimum at the point $x = x_\varepsilon$. Since $\overline{u}$ is a super-solution, we have

$$H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) \ge 0. \tag{5.23}$$

On the other hand, for the function

$$\psi(y) = \overline{u}(x_\varepsilon) + \frac{|x-x_\varepsilon|^2}{2\varepsilon^2},$$

the difference

$$u(y;t) - \psi(y) = u_t(y) - \overline{u}(x_\varepsilon) - \frac{|x - x_\varepsilon|^2}{2\varepsilon^2} = -u_\varepsilon(x_\varepsilon, y),$$

attains its maximum at $y = y_\varepsilon$, hence

$$H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) \leq 0. \tag{5.24}$$

We deduce from (5.24) that

$$|y_\varepsilon - x_\varepsilon| \leq C\varepsilon^2.$$

Moreover, if $x_0 \notin B$, then, there exists $\alpha_0 > 0$ so that for $\varepsilon > 0$ sufficiently small, we have

$$\text{dist}(x_\varepsilon, B), \text{dist}(y_\varepsilon, B) \geq \alpha_0 > 0,$$

and thus

$$H(y_\varepsilon, \nabla u(y_\varepsilon; t)) \leq -c_t < 0, \quad H(x_\varepsilon, \nabla u(x_\varepsilon; t)) \leq -c_t < 0,$$

with some $c_t > 0$. It follows that

$$o(1) = H(y_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) - H(x_\varepsilon, \frac{y_\varepsilon - x_\varepsilon}{\varepsilon^2}) \geq c_t > 0 \text{ independent of } \varepsilon,$$

which is a contradiction. We conclude that $x_0 \in B$ for all $t \in (0,1)$. It is easy to see, however, that, as $\overline{u}(x) \geq \underline{u}(x)$ for $x \in B$, the points $x_\varepsilon, y_\varepsilon$ can not be too close to B,when $t > 0$ is small, because of (5.21), which means that $x_0 \in B$ is impossible. □

A proof of Theorem 5.6

We will prove Theorem 5.6 using Lemma 5.11. The existence result for critical subsolutions [62] is as follows.

Theorem 5.13 *There is a critical subsolution ψ of* (5.2)*, of class C^1 over $\mathbb{T}^n$, which is strict over $\mathbb{T}^n \backslash \mathcal{A}$. Moreover, ψ is of class $C^{1,1}$ over $\mathcal{A}$.*

Theorem 5.6 is an immediate consequence of Theorem 5.13 and Lemma 5.11.

We now turn to the proof of Theorem 5.13. The idea is to choose, as a critical sub-solution, a convex combination of subsolutions of the form

$$\psi(x) = \sum_{n \in I\!N} \frac{\phi_{x_n}(x)}{2^{n+1}},$$

the base points x_n being chosen outside of the Aubry set $\mathcal{A}$. This will define a strict subsolution, which is $C^{1,1}$ over $\mathcal{A}$. The problem with this direct approach is that $\phi_x(y)$ does not have to be C^1 at $y = x$ – to overcome this, we will regularize each ϕ_{x_n} outside of $\mathcal{A}$. This will ensure that (a properly redefined) ψ is globally C^1 on $\mathbb{T}^n$ while keeping the sub-solution property. Many extensions are available in [62].

Step 1. Behaviour of ϕ_x on $\mathcal{A}$. If u is a critical subsolution, define $\mathcal{I}(u)$ as the set of points in $\mathbb{T}^n$ which are passed by some extremal $\gamma : \mathbb{R} \to \mathbb{T}^n$ calibrated by u. A priori, if u is a sub-solution (and not a solution) we do not know if there are any extremals calibrated by u. The main result here is

Lemma 5.14 *For every point $x \in \mathcal{A}$ there exists an extremal $\gamma : \mathbb{R} \to \mathbb{T}^n$ such that $\gamma(0) = x$, which is calibrated by every subsolution.*

Proof. Let us construct the extremal, with the help of the function h. Given a point $x \in \mathcal{A}$, we are going to construct an extremal $\gamma : \mathbb{R} \to \mathbb{T}^n$, such that $\gamma(0) = x$, and

$$h(\gamma(t), x) = -\int_0^t L(\gamma, \dot{\gamma})\ ds, \quad h(x, \gamma(-t)) = -\int_{-t}^0 L(\gamma, \dot{\gamma})\ ds, \tag{5.25}$$

for any $t > 0$. As $h(x, x) = 0$, we may find a sequence of extremals $\gamma_m : [0, t_m] \to \mathbb{T}^n$, with

$$\gamma_m(0) = \gamma_m(t_m) = x,$$

such that

$$\lim_{m \to +\infty} \int_0^{t_m} L(\gamma_m, \dot{\gamma}_m)\ ds = 0. \tag{5.26}$$

We may assume that γ_m converges, as well at its derivatives, locally uniformly. Let γ be this local uniform limit, fix $t > 0$, and set

$$d_m = |\gamma(t) - \gamma_m(t)|.$$

Let $\mu_m : [0, d_m] \to \mathbb{T}^n$ be an extremal of the action between the points $\gamma(t)$ and $\gamma_m(t)$. Gluing together μ_m and γ_m, suitably reparametrized, yields a Lipschitz curve

$$\widetilde{\gamma} : [t - d_m, t_m] \to \mathbb{T}^n,$$

which connects the point $\gamma(t)$ to x. We have

$$\int_{t-d_m}^{t_m} L(\widetilde{\gamma}, \dot{\widetilde{\gamma}})\ ds \le C d_m + \int_t^{t_m} L(\gamma_m, \dot{\gamma}_m)\ ds = C d_m + \int_0^{t_m} L(\gamma_m, \dot{\gamma}_m) ds - \int_0^t L(\gamma_m, \dot{\gamma}_m)\ ds.$$

Minimizing the left side over all paths going from $\gamma(t)$ to x gives

$$h_{t_m - t + d_m}(\gamma(t), x) \le C d_m + \int_0^{t_m} L(\gamma_m, \dot{\gamma}_m) - \int_0^t L(\gamma_m, \dot{\gamma}_m)\ ds.$$

This, thanks to (5.26) and the definition of h as a lower limit, implies

$$h(\gamma(t), x) \le -\int_0^t L(\gamma, \dot{\gamma})\ ds.$$

On the other hand, (5.14) implies that

$$h(\gamma(t), x) \ge -h(x, \gamma(t)) \ge -\int_0^t L(\gamma, \dot{\gamma})\ ds.$$

This implies the first part of (5.25). The second part is obtained in an analogous fashion.

With (5.25) in hand, let u be a critical sub-solution. We have, for the constructed extremal γ and $x \in \mathcal{A}$, and $t > 0$:

$$u(\gamma(t)) - u(x) \leq \int_0^t L(\gamma, \dot{\gamma})\ ds = -h(\gamma(t), x), \tag{5.27}$$

by the definition of a sub-solution. But we also have, for all s:

$$u(x) - u(\gamma(t)) \leq h_s(\gamma(t), x).$$

Passing to the lower limit leads to:

$$u(x) - u(\gamma(t)) \leq h(\gamma(t), x). \tag{5.28}$$

Putting (5.27) and (5.28) together gives

$$u(\gamma(t)) - u(x) = \int_0^t L(\gamma, \dot{\gamma})\ ds. \tag{5.29}$$

A similar argument applies to $t < 0$. We see that γ is calibrated by u on $[0, +\infty)$. □

An important consequence is that, for all $x \in \mathbb{T}^n \setminus \mathcal{A}$, the function ϕ_x is $C^{1,1}$ on $\mathcal{A}$. Moreover, the Lipschitz constant of $\nabla\phi_x$ on $\mathcal{A}$ only depends on the Hamiltonian, any point of $\mathcal{A}$ belonging to an extremal calibrated by ϕ_x.

Step 2: Construction of a critical subsolution outside of the set $\mathcal{A}$. The key ingredient this time is the

Lemma 5.15 *If $x \in \mathbb{T}^n \backslash \mathcal{A}$, then ϕ_x is not a viscosity solution on $\mathbb{T}^n$.*

Proof. Assume that ϕ_x is a solution of (5.2) over all of $\mathbb{T}^n$. According to Proposition 5.3, we may then find an extremal $\gamma : \mathbb{R}_- \to \mathbb{T}^n$ such that $\gamma(0) = x$ and such that, for all $t > 0$:

$$-\phi_x(\gamma(-t)) = \phi_x(x) - \phi_x(\gamma(-t)) = \int_{-t}^0 L(\gamma, \dot{\gamma})\ ds. \tag{5.30}$$

Fix now $t > 0$ and $\varepsilon > 0$. The definition of ϕ_x implies the existence of a time $t_\varepsilon > 0$ and of an extremal $\gamma_\varepsilon : [0, t_\varepsilon] \to \mathbb{T}^n$ such that $\gamma_\varepsilon(0) = x$, $\gamma_\varepsilon(t_\varepsilon) = \gamma(-t)$, and such that

$$\int_0^{t_\varepsilon} L(\gamma_\varepsilon, \dot{\gamma}_\varepsilon)\ ds \leq \phi_x(\gamma_\varepsilon(t_\varepsilon)) + \varepsilon = \phi_x(\gamma(-t)) + \varepsilon. \tag{5.31}$$

Next, define the (Lipschitz) glued curve

$$\widetilde{\gamma}_\varepsilon(s) = \gamma_\varepsilon(s) \text{ for } 0 \leq s \leq t_\varepsilon, \text{ and } \widetilde{\gamma}_\varepsilon(s) = \gamma(s - t_\varepsilon - t) \text{ for } t_\varepsilon \leq s \leq t + t_\varepsilon.$$

Combining (5.30) and (5.31), we obtain:

$$\int_0^{t+t_\varepsilon} L(\widetilde{\gamma}, \dot{\widetilde{\gamma}}_\varepsilon)\ ds \leq \varepsilon.$$

We may thus go to the lower limit $t \to +\infty$ to obtain $h(x, x) \leq \varepsilon$, hence, $x \in \mathcal{A}$. □

Lemmas 5.15 and 5.14 imply

Corollary 5.16 *The Mañé potential ϕ_x is a viscosity solution on the whole $\mathbb{T}^n$ if and only if $x \in \mathcal{A}$.*

The construction of a critical subsolution goes as follows. For all $x \in U := \mathbb{T}^n \backslash \mathcal{A}$, the function ϕ_x is not a solution at x. As ϕ_x is a sub-solution on $\mathbb{T}^n$ but not a solution at x, it is not a super-solution at x. Thus, there exist an open neighbourhood U_x of x, a function $\theta_x \in C^1(U_x)$ and a real number $c_x < 0$ such that $\phi_x \geq \theta_x$ on U_x, $\phi_x(x) = \theta_x(x)$ and

$$H(y, \nabla\theta_x) \leq c_x,$$

over U_x. Subtracting a quadratic function from θ_x we may ensure that $\phi_x > \theta_x$ on $U_x \backslash \{x\}$. Then, there is $\varepsilon_x > 0$ and Ω_x, an open neighbourhood of x inside U_x, such that $\phi_x = \theta_x + \varepsilon_x$ on $\partial\Omega_x$. So, denote

$$u_x(y) = \begin{cases} \theta_x(y) + \varepsilon_x & \text{for } y \in \Omega_x \\ \phi_x(y) & \text{for } y \notin \Omega_x. \end{cases}$$

We will choose U_x so that, in addition, they do not intersect the Aubry set $\mathcal{A}$, which leads to $u_x(y) = \phi_x(y)$ for all $y \in \mathcal{A}$. The function u_x is still a critical subsolution over $\mathbb{T}^n$, as the maximum of subsolutions, but it is now strict over Ω_x. It is not yet C^1 on $\mathbb{T}^n$ since it is only Lipschitz on $\partial\Omega_x$. Let us extract from the cover

$$U = \bigcup_{x \in U} \Omega_x$$

a countable sub-cover

$$U = \bigcup_{n \in \mathbb{N}} \Omega_{x_n}.$$

By the convexity of H, the function

$$\underline{\phi} = \sum_{n \in \mathbb{N}} \frac{u_{x_n}}{2^{n+1}} \tag{5.32}$$

is still a critical subsolution, which coincides with

$$\underline{\phi} = \sum_{n \in \mathbb{N}} \frac{\phi_{x_n}}{2^{n+1}} \tag{5.33}$$

on $\mathcal{A}$. It is of the class $C^{1,1}$ on $\mathcal{A}$, as seen from the convergence of the series

$$\sum \frac{\nabla\phi_{x_n}}{2^{n+1}}$$

over $\mathcal{A}$. As a simple exercise, the reader may check that $\underline{\phi}$ is a strict subsolution over $\mathbb{T}^n \backslash \mathcal{A}$.

Step 3: Regularization. The next task is to smooth the subsolution $\underline{\phi}$ constructed in Step 2 over $U := \mathbb{T}^n \backslash \mathcal{A}$ (for the moment it is only Lipschitz outside $\mathcal{A}$ because each u_{x_n} is only Lipschitz on $\partial\Omega_{x_n}$), keeping it a strict subsolution over U, and keeping its values over $\mathcal{A}$ unchanged. What is going to help here is that $\underline{\phi}$ is Lipschitz, as well as the mollification computation (5.20). First, we need a special partition of unity.

Lemma 5.17 *There is a covering B_n of U by open balls of radii r_n, and a sequence ψ_n of C^∞ functions over U such that*

- *The number $N(n)$ of indices k such that $\mathrm{supp}\psi_n \cap \mathrm{supp}\psi_k$ is not empty, is finite for each n.*
- *For all $n \in \mathbb{N}$, we have $\|\nabla \psi_n\|_\infty \leq \frac{1}{r_n}$;*
- *for all $x \in U$ we have $\sum_{n\in\mathbb{N}} \psi_n(x) = 1$.*

Proof. Let, for all $n \geq 1$:

$$F_n = \{x \in U,\ \frac{1}{n+2} \leq d(x, \mathcal{A}) \leq \frac{1}{n+1}\}, \qquad U_n = \{x \in U,\ \frac{1}{n+3} < d(x, \mathcal{A}) < \frac{1}{n}\}.$$

It suffices to cover F_n with $\widetilde{N}(n)$ closed balls, each of them in U_n, and to consider a classical partition of unity: $(\psi_{k,n})_{1\leq k\leq \widetilde{N}(n)}$ relative to these balls. □

Define the function ϕ as

$$\forall x \notin \mathcal{A},\quad \phi(x) = \sum_{n\in I\!N} \psi_n \rho_{\varepsilon_n} * \underline{\phi}, \qquad \forall x \in \mathcal{A},\quad \phi(x) = \underline{\phi}(x).$$

The function $\rho_\varepsilon = \varepsilon^{-n}\rho(x/\varepsilon)$ is a classical approximation of identity. Recall the existence of $c_n < 0$ such that $H(x, \nabla\underline{\phi}) \leq c_n$ over B_n; we choose

$$\varepsilon_n \ll \min(c_n, \min_{k\in N(n)} r_k, \frac{1}{|N(n)|} \min_{k\in N(n)} d(\partial U_k, \mathcal{A})^2). \tag{5.34}$$

Now, arguing as in computation (5.20), using the estimates

$$\|\nabla \psi_n\|_\infty \leq \frac{1}{r_n}$$

and finally the definition (5.34) of ε_n, we obtain

$$H(x, \nabla \phi) \leq c_n + O(\frac{\varepsilon_n}{r_n})$$

over B_n. This guarantees that ϕ is a strict C^1 critical subsolution over U.

To show that ϕ is C^1 over $\mathbb{T}^n$, it suffices to show that this function and its gradient match at the interface $\partial\mathcal{A}$ – everywhere else it is smooth by its definition. To this end, choose $x \in U$ and $n \in \mathbb{N}$ such that $x \in B_n$. Because $\underline{\phi}$ is Lipschitz, there is $C > 0$ such that

$$\|\rho_{\varepsilon_n} * \underline{\phi} - \underline{\phi}\|_\infty \leq C\varepsilon_n$$

for all n, and we deduce:

$$|\phi(x) - \underline{\phi}(x)| \leq C \sum_{k\in N(n)} \varepsilon_k \leq C \min_{k\in N(n)} d(\partial U_k, \mathcal{A})^2 \leq C d(x, \mathcal{A})^2$$

This implies the matching of ϕ and its gradient at the interface.

5.4 Invariant regions

We finish this chapter with some implications to the invariant regions of the corresponding Hamiltonian system. The following result comes almost immediately.

Theorem 5.18 *Choose $x \in \mathcal{A}$ and let $\gamma : \mathbb{R} \to \mathbb{T}^n$ be calibrated by the critical subsolution ϕ of Theorem 5.13, such that $\gamma(0) = x$. Then, for all $t \in \mathbb{R}$, $\gamma \in \mathcal{A}$.*

Proof. Take $t \in \mathbb{R}$ and apply Lemma 5.14: the extremal $\gamma_t(s) = \gamma(t+s)$ is calibrated by *every* critical subsolution, so, in particular, by the C^1 critical subsolution ϕ of Theorem 5.13. If $\gamma(t)$ were outside the Aubry set $\mathcal{A}$, this would be a contradiction to ϕ being a strict subsolution at that point. □

The value of the result is in its interpretation. Here is a corollary of Theorem 5.13:

Corollary 5.19 *Let ϕ the critical subsolution of Theorem 5.13. Then $\nabla u = \nabla \phi$ on $\mathcal{A}$, for all critical subsolutions u of* (5.2).

Proof. If $x \in \mathcal{A}$ and u is a critical subsolution, let $\gamma : \mathbb{R} \to \mathbb{T}^n$ be an extremal calibrated by all critical solutions. From remark 3.5, we have

$$\nabla \phi(\gamma(t)) = \nabla_v L(\gamma(t), \dot{\gamma}(t)) = \nabla u(\gamma(t)).$$

This is the sought for result. □

Another consequence of the invariance of the set $\mathcal{A}$ and the fact that ϕ is a solution on $\mathcal{A}$, is

Theorem 5.20 *The set $\{(x, \nabla\phi(x)), \quad x \in \mathcal{A}\}$ is an invariant region for the Hamiltonian system*

$$\left\{ \begin{array}{rl} \dot{X} = & \nabla H(X,P) \\ \dot{P} = & -\nabla H(X,P) \end{array} \right. \tag{5.35}$$

The (widespread) appellation "weak KAM theory" originates from this type of property. Notice that we are here outside every perturbative framework.

Chapter 6

The two dimensional Euler equations

1 Introduction: the Euler equations and the vorticity formulation

In this chapter, we will study the 2D incompressible Euler equations of the fluid mechanics. These equations describe the flow of an incompressible, inviscid fluid, and were first derived by Leonhard Euler in 1755 [51]. Let D be a compact smooth domain in $\mathbb{R}^d$, $d = 2$ or 3. The equation is given by

$$\begin{aligned} &\partial_t u + (u \cdot \nabla)u + \nabla p = 0, \\ &\nabla \cdot u = 0, \\ &u(x, 0) = u_0(x), \end{aligned} \tag{1.1}$$

along with the no flow on the boundary condition

$$u \cdot \nu|_{\partial D} = 0. \tag{1.2}$$

Here, $u(x)$ is the vector field describing the fluid velocity, and p is the pressure which mathematically can be thought of as a Lagrange multiplier needed to accommodate the incompressibility constraint. The Euler equations are also often considered in the whole space case with the decay conditions at infinity, or on the torus – which is equivalent to taking periodic initial data in $\mathbb{R}^d$. There are many great texts outlining the derivation and the basic properties of the equation, see for example [31], [87], [48] and [89].

The Euler equations are maybe the most fundamental and widely used partial differential equations. They are nonlinear and nonlocal, the latter property a consequence of the nonlocal dependence of pressure on the fluid velocity. On the physical level this simply because pushing the fluid in one region produces an instantaneous pressure in a different region, because of the fluid incompressibility. Mathematically, taking the divergence of (1.1) and using the incompressibility of u, we obtain the Poisson equation for there pressure:

$$-\Delta p = \nabla \cdot (u \cdot \nabla u),$$

which shows the non locality of the pressure-velocity relation. This will be even more clear from the nonlocal Biot-Savart law for the vorticity form of the equation presented below.

This explains, from the mathematical point of view, why the analysis of the Euler equation is challenging. From the intuitive point of view, anyone who observed the flow of a river, or the intricate structures of the fluid motion in a rising smoke, or a tornado, can understand that only a very rich and complex equation has a chance of modeling these exquisite phenomena. The solutions of the Euler equations are often very unstable, and prone to creation of small scale structures. Due to the central role of these equations in mathematical physics, a lot of studies have focused on these problems over the 250 years that have passed since its discovery. We have no hope of covering much of this research here, so after a brief overview we will focus on a few specific questions, including some very recent developments.

The theory of the existence, uniqueness and regularity of solutions to the Euler equations is quite different in two and three spatial dimensions. In the two dimensional case, for smooth initial data there exists a unique global in time smooth solution, while for the three dimensional case an analogous result is only known locally in time. The question of global existence of smooth solutions to the Euler equations in three dimensions is a major open problem. This difference can be illustrated on a basic level by rewriting the Euler equations in a different form. An important quantity in the fluid mechanics is the vorticity ω, which describes the rotational motion of the fluid and is given by $\omega = \mathrm{curl} u$. In three dimensions, if we apply curl to equation (1.1), we obtain the Euler equation in the vorticity form:

$$\omega_t + (u \cdot \nabla)\omega = (\omega \cdot \nabla)u, \tag{1.3}$$

with the initial condition $\omega(0, x) = \omega_0(x)$.

Exercise 1.1 Use vector algebra to derive the vorticity equation (1.3) in three dimensions.

The vector field u can be recovered from ω via the Biot-Savart law. Consider the (vector-valued) stream function ψ defined by the Dirichlet problem

$$-\Delta\psi = \omega, \quad \psi|_{\partial D} = 0. \tag{1.4}$$

This problem has a unique solution by the classical results of the elliptic theory (see e.g. [69]). We denote this solution by $\psi = (-\Delta_D)^{-1}\omega$. One can also write

$$\psi(x) = \int_D G_D(x, y)\omega(y)\, dy \tag{1.5}$$

where G_D is the Dirichlet Green's function of the Laplacian. Then one can show that u is given by (see, for example, [48, 87])

$$u = \mathrm{curl}\psi. \tag{1.6}$$

In particular, the velocity u defined by the Biot-Savart law (1.5)-(1.6), satisfies the no flow boundary condition (1.2).

Exercise 1.2 Verify that u given by (1.5)-(1.6) satisfies $\mathrm{curl} u = \omega$ in D, and $u \cdot \nu = 0$ on ∂D. You have to use the divergence free property of u and some vector identities (or brute force computations).

On the other hand, in the two dimensional case the term in the right side of (1.3) vanishes. Indeed, the solutions of the two-dimensional Euler equations can be thought of as solutions of the three-dimensional equations of the special form $(u_1(x_1,x_2), u_2(x_1,x_2), 0)$. In that case, the vorticity vector has only one non-zero component:

$$\omega = (0, 0, \partial_1 u_2 - \partial_2 u_1),$$

and can be regarded as a scalar. Then, the term in the right side of (1.3) is simply

$$(\omega \cdot \nabla) u = \omega_3 \partial_3 u,$$

but the two dimensional u does not depend on x_3. Thus, in two dimensions, the vorticity equation simplifies. We will use the notation

$$\omega = \partial_1 u_2 - \partial_2 u_1, \tag{1.7}$$

instead of ω_3. This scalar vorticity satisfies

$$\partial_t \omega + (u \cdot \nabla)\omega = 0, \tag{1.8}$$
$$u = \nabla^\perp (-\Delta_D)^{-1} \omega, \tag{1.9}$$
$$\omega(0,x) = \omega_0(x),$$

where $\nabla^\perp = (\partial_2, -\partial_1)$. Note that the flow u defined by (1.9) automatically satisfies the boundary condition

$$u \cdot \nu = 0 \text{ on } \partial D.$$

This is because the gradient of the stream function

$$\psi = (-\Delta_D)^{-1}\omega, \quad u = \nabla^\perp \psi,$$

is normal to ∂D at the boundary.

Exercise 1.3 Verify that if $u(t,x)$ satisfies the Euler equations in two dimensions, then the vorticity $\omega(t,x)$ given by (1.7) satisfies (1.8), and $u(t,x)$ and $\omega(t,x)$ are related via (1.9).

This simpler form of the Euler equations in two dimensions has significant consequences. As we will see, any L^p norm of the vorticity is conserved for smooth solutions of (1.8). In particular, $\|\omega\|_{L^\infty}$ does not change. In contrast, in three dimensions, the amplitude of vorticity can and usually does grow due to the non-zero term on the right hand side of (1.3). This term is often called the vortex stretching term in the literature.

Our focus in the present chapter will be on the basic questions of existence, uniqueness, and regularity properties of the solutions to the two dimensional Euler equations. First, we will present the existence and uniqueness theory of solutions due to Yudovich [121] which works for a very natural class of initial data. We will then study the small scale formation in the smooth solutions of the 2D Euler equations, proving an upper bound for the growth of the derivatives of the solution as well as constructing examples that show that in general this upper bound is sharp. The set of techniques we will need in this chapter is a beautiful mix of Fourier analysis, ODE methods, comparison principles, and all sorts of other PDE estimates.

2 The Yudovich theory

The Yudovich theory proves existence and uniqueness of solutions to the 2D Euler equations with a bounded initial vorticity. Thus, the regularity assumptions on the initial data are fairly mild. The L^∞ class for the vorticity is very natural since it is preserved by the evolution. In addition, many phenomena in nature, such as hurricanes or tornados, feature vorticities with a sharp variation. As we will see, if the initial condition is more regular, this regularity is reflected in the additional regularity of the solution, even though the quantitative estimates can deteriorate very quickly. Our exposition in this section roughly follows [89].

It is not immediately clear how one can define the low regularity solutions (such as L^∞) of the vorticity equation (1.8) since we need to take derivatives. A "canonical" way around that is to define a weak solution of a nonlinear equation via the multiplication of the equation by a test function and integration by parts, and then try to obtain some a priori bounds and use some compactness arguments to show that such weak solution exists. However, there is a more elegant (and efficient) approach for the two-dimensional Euler equations, via a reformulation of the problem that allows us to define a weak solution in an appropriate sense. Given a divergence-free flow $u(t,x)$, we may define the particle trajectories $\Phi_t(x)$ by

$$\frac{d\Phi_t(x)}{dt} = u(t,\Phi_t(x)), \;\; \Phi_0(x) = x. \tag{2.1}$$

If u is sufficiently regular and incompressible, (2.1) defines a volume preserving map for each t.

Exercise 2.1 Verify the claim that the map $x \to \Phi_t(x)$ is measure-preserving for each t fixed if $\nabla \cdot u = 0$. You can find the proof for example in [31] or [87].

A direct calculation using the method of characteristics shows that if $\omega(t,x)$ is a smooth solution of (1.8), then

$$\omega(t,\Phi_t(x)) = \omega_0(x), \;\; \text{thus } \omega(t,x) = \omega_0(\Phi_t^{-1}(x)). \tag{2.2}$$

In addition, if we denote, as before, by $G_D(x,y)$ the Green's function for the Dirichlet Laplacian in a domain D (in the sense that the solution of (1.4) is given by (1.5)), and set $K_D(x,y) = \nabla_x^\perp G_D(x,y)$, then the Biot-Savart law in two dimensions can be written as

$$u(t,x) = \int_D K_D(x,y)\omega(t,y)\,dy. \tag{2.3}$$

A classical C^1 solution of the two-dimensional Euler equations (1.8) satisfies the system (2.1), (2.2) and (2.3). On the other hand, a direct computation shows that a smooth solution of (2.1), (2.2) and (2.3) gives rise to the classical solution of (1.8). Thus, for smooth solutions the two problems are equivalent. We will generalize the notion of the solution to the 2D Euler equations by saying that a triple $(\omega, u, \Phi_t(x))$ solves the 2D Euler equations if it satisfies (2.1), (2.2) and (2.3). The obvious next task is to make sense of the solutions of the latter system with the only requirement that $\omega_0 \in L^\infty$. Classically, for the trajectories of (2.1) to be well-defined, the flow $u(t,x)$ needs to be Lipschitz in x. Thus, if it were true that if $\omega(t,x)$ is in L^∞, the Biot-Savart law would give a Lipschitz function $u(t,x)$, then it would be very reasonable to expect (2.1), (2.2) and (2.3) to be a well-posed system. This, however, is not quite true – the regularity for $u(t,x)$ when $\omega \in L^\infty$ is slightly lower than Lipschitz. Nevertheless, we will see that this lower regularity is sufficient to define the trajectories of the ODE (2.1), making the system well-posed.

The regularity of the flow

To construct the solutions of the 2D Euler equations in the trajectory formulation (2.1)-(2.3) with the vorticity $\omega_0 \in L^\infty$, we need to establish the regularity of the fluid flow given by (2.3) for a vorticity in L^∞. The following proposition summarizes some well known properties of the Dirichlet Green's function (see, e.g. [69]).

Proposition 2.2 *If $D \subset \mathbb{R}^2$ is a domain with a smooth boundary, the Dirichlet Green's function $G_D(x, y)$ has the form*

$$G_D(x,y) = \frac{1}{2\pi}\log|x-y| + h(x,y).$$

Here, for each $y \in D$, $h(x, y)$ is a harmonic function solving

$$\Delta_x h = 0, \quad h|_{x\in\partial D} = -\frac{1}{2\pi}\log|x-y|. \tag{2.4}$$

We have $G_D(x, y) = G_D(y, x)$ for all $(x, y) \in D$, and $G_D(x, y) = 0$ if either x or y belongs to ∂D. In addition, we have the estimates

$$|G_D(x,y)| \le C(D)(\log|x-y|+1) \tag{2.5}$$
$$|\nabla G_D(x,y)| \le C(D)|x-y|^{-1}, \tag{2.6}$$
$$|\nabla^2 G_D(x,y)| \le C(D)|x-y|^{-2}. \tag{2.7}$$

Sometimes, G_D can be computed explicitly in a closed form (for example for a plane, a half-plane, a disk, a corner, see e.g. [52]), or as an infinite series (for example for a square or a rectangle, or a torus). For most domains only estimates are available. The following lemma outlines a key property which allows to construct unique solutions for bounded vorticity.

Lemma 2.3 *The kernel $K_D(x, y) = \nabla^\perp G_D(x, y)$ satisfies*

$$\int_D |K_D(x,y) - K_D(x',y)|\,dy \le C(D)\phi(|x-x'|), \tag{2.8}$$

where

$$\phi(r) = \begin{cases} r(1-\log r) & r<1 \\ 1 & r\ge 1, \end{cases} \tag{2.9}$$

with a constant $C(D)$ which depends only o the domain D.

Proof. In what follows, $C(D)$ denotes constants that may depend only on the domain D, and may change from line to line. To show (2.8), we may assume that $r = |x - x'| < 1$, otherwise (2.8) follows from the simple observation that

$$|K_D(x,y)| \le C(D)|x-y|^{-1},$$

so that

$$\int_D |K_D(x,y)|dy \le C(D),$$

which implies (2.8) for $x, x' \in D$ such that $|x - x'| > 1$. Assume now that $r < 1$ and suppose first that the interval connecting the points x and x' lies entirely inside D. Let us set

$$A = \{y \in D : \ |y - x| \leq 2r\}.$$

The estimate (2.6) implies

$$\int_{D \cap A} |K_D(x, y) - K_D(x', y)|\, dy \leq C(D) \int_{B_{2r}(x)} \frac{1}{|x - y|}\, dy \leq C(D) r.$$

To bound the remainder of the integral, observe that for every y,

$$|K(x, y) - K(x', y)| \leq r |\nabla K(x''(y), y)|, \tag{2.10}$$

where $x''(y)$ lies on the interval connecting x and x'. This follows from the mean value theorem and the assumption that the interval connecting x and x' lies in D. Then, by (2.7) and the choice of the set A, so that the distances $|x - y|$, $|x' - y|$ and $|x'' - y|$ are all comparable if $y \in A^c$, we have

$$\int_{D \cap A^c} |K_D(x, y) - K_D(x', y)|\, dy \leq C(D) r \int_{D \cap A^c} \frac{dy}{|x''(y) - y|^2}$$
$$\leq C(D) r \int_r^{C(D)} s^{-1}\, ds \leq C(D) r (1 - \log r).$$

The case where the interval connecting x and x' does not lie entirely in D is similar, one just needs to replace this interval by a curve connecting x and x' with the length of the order r. We briefly sketch the argument. The following lemma can be proved by standard methods using the compactness of the domain and the regularity of the boundary, so we do not present its proof.

Lemma 2.4 *Fix $\varepsilon > 0$ and let $D \subset \mathbb{R}^2$ be bounded domain with a smooth boundary. Then there exists $r_0 = r_0(D, \varepsilon) > 0$ such that if $x_0 \in \partial D$, and $r \leq r_0$, then $B_r(x_0) \cap \partial D$ is a curve that, by a rotation and a translation of the coordinate system, can be represented as a graph $x_2 = f(x_1)$, with $x_0 = (0, 0)$. The function f is C^∞, and $f'(x_{0,1}) = 0$. Moreover, the part of the boundary ∂D within $B_r(x_0)$ lies in the narrow angle between the the lines $x_2 = \pm\varepsilon x_1$.*

With this lemma, suppose we have x and x' such that the interval connecting these points does not lie in D. It is enough to consider the case where $|x - x'| = r < r_0/2$, where r_0 is as in Lemma 2.4 corresponding to a sufficiently small ε. Indeed, the larger values of $|x - x'|$ can be handled by adjusting $C(D)$ in (2.8). Find a point $x_0 \in \partial D$ closest to x (it does not have to be unique). Note that by the assumption that the interval (x, x') crosses the boundary, we must have $|x - x_0| \leq r_0/2$ and $|x' - x_0| < r_0$. Thus,. both x and x' lie in the disk $B(x_0, r)$ where ∂D lies between the lines $x_2 = \pm\varepsilon x_1$. It is also not hard to see that x must lie on the vertical x_2-axis of a system of coordinates centered at x_0, with the horizontal x_1-axis tangent to ∂D at x_0. Since by assumption the interval between x and x' does not lie in D, we know that x' must lie in the narrow angle between the lines $x_2 = \pm\varepsilon x_1$. Otherwise, the interval (x, x') could not have crossed the boundary. Now take a curve connecting x and x' consisting of a straight vertical interval from x' to a point on one of the lines $x_2 = \pm\varepsilon x_1$ which is closest to x, and then an interval connecting this point to x. We can smooth out this curve without changing its length by much. It is easy to see that the length of this curve does not exceed $2r$ if ε is small enough. The rest of the proof goes through as before. □

Now we can state the regularity result for the fluid velocity.

Corollary 2.5 *The fluid velocity u satisfies*

$$\|u\|_{L^\infty} \le C(D)\|\omega\|_{L^\infty}, \tag{2.11}$$

and

$$|u(x) - u(x')| \le C\|\omega\|_{L^\infty}\phi(|x - x'|), \tag{2.12}$$

with the function $\phi(r)$ defined in (2.9).

Proof. By (2.6), we have, for any $x, y \in D$,

$$|K_D(x, y)| \le C(D)|x - y|^{-1},$$

so that

$$\left|\int_D K_D(x,y)\omega(y)\,dy\right| \le C(D)\|\omega\|_{L^\infty}\int_D \frac{1}{|x-y|}\,dy \le C(D)\|\omega\|_{L^\infty},$$

which is (2.11). The proof of (2.12) is immediate from Lemma 2.3, as

$$u(t, x) = \int_D K_D(x, y)\omega(t, y)dy,$$

and we are done. □

We say that u is log-Lipschitz if it satisfies (2.12). We will see that this bound is in fact sharp; there are velocities that correspond to bounded vorticities which are just log-Lipschitz and in particular fail to be Lipschitz.

Trajectories for log-Lipschitz velocities

As the fluid velocity with an L^∞-vorticity is not necessarily Lipschitz but only log-Lipschitz, we may not use the classical results on the existence and uniqueness of the solutions of ODEs with Lipschitz velocities. Nevertheless, as we show next, the log-Lipschitz regularity is sufficient for determining fluid trajectories uniquely.

Lemma 2.6 *Assume that the velocity field $b(t, x)$ satisfies*

$$b \in C([0,\infty) \times \mathbb{R}^d) \;\; |b(t, x) - b(t, y)| \le C\phi(|x - y|), \;\; \forall t > 0, \tag{2.13}$$

with the function $\phi(r)$ given by (2.9). Then the Cauchy problem in $\mathbb{R}^d$

$$\frac{dx}{dt} = b(t, x), \;\; x(0) = x_0, \tag{2.14}$$

has a unique solution.

Note that the log-Lipschitz regularity is border-line: the familiar example of the ODE

$$\dot{x} = x^\beta, \quad x(0) = 0,$$

with $\beta \in (0,1)$ has two solutions: $x(t) \equiv 0$, and

$$x(t) = \frac{t^p}{p^p}, \quad p = \frac{1}{1-\beta},$$

so that ODE's with Hölder (with an exponent smaller than one) velocities may have more than one solution. Existence of solutions, on the other hand, does not really require the log-Lipschitz condition: uniform continuity of $b(t,x)$ and at most linear growth as $|x| \to +\infty$ would be sufficient.

Proof. Let us first show existence of a solution using the standard Picard iteration: set

$$x_n(t) = x_0 + \int_0^t b(s, x_{n-1}(s))\,ds, \quad x_0(t) \equiv x_0.$$

Then, as usual, we have

$$|x_n(t) - x_{n-1}(t)| \le \int_0^t |b(s, x_{n-1}(s)) - b(s, x_{n-2}(s))|\,ds \le C \int_0^t \phi(|x_{n-1}(s) - x_{n-2}(s)|)\,ds. \tag{2.15}$$

As the function $\phi(r)$ is concave, we have

$$\phi(r) \le \varepsilon(1 + \log \varepsilon^{-1}) + (r - \varepsilon)\log \varepsilon^{-1} = \varepsilon + r \log \varepsilon^{-1},$$

for every $\varepsilon < 1$. Using this in (2.15) gives

$$|x_n(t) - x_{n-1}(t)| \le C \log(\varepsilon^{-1}) \int_0^t |x_{n-1}(s) - x_{n-2}(s)|\,ds + Ct\varepsilon.$$

Iterating, we get for any $0 \le t \le T$ (check this using the induction on n)

$$|x_n(t) - x_{n-1}(t)| \le CT\varepsilon \sum_{k=0}^{n-2} \frac{C^k (\log \varepsilon^{-1})^k t^k}{k!} + \frac{C^{n-1} t^{n-1} (\log \varepsilon^{-1})^{n-1}}{(n-1)!} \sup_{0 \le t \le T} |x_1(t) - x_0|.$$

As

$$|x_1(t) - x_0| \le Ct,$$

we have

$$|x_n(t) - x_{n-1}(t)| \le CT\varepsilon \exp(CT \log \varepsilon^{-1}) + \frac{C^n T^n (\log \varepsilon^{-1})^{n-1}}{(n-1)!}.$$

Choose now $\varepsilon = \exp(-n)$ and T sufficiently small so that $1 - CT > 1/2$, so that

$$|x_n(t) - x_{n-1}(t)| \le CT \exp(-n/2) + \frac{C^n T^n n^{n-1}}{(n-1)!}.$$

The Stirling formula

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

implies that if T is sufficiently small (independently of n and x_0), then

$$|x_n(t) - x_{n-1}(t)| \leq \alpha^n(T),$$

with $\alpha(T) < 1$. Thus, $x_n(t)$ converges uniformly to a limit $x(t)$. The uniformity of the convergence implies that the limit satisfies the integral equation

$$x(t) = x_0 + \int_0^t b(x(s), s)\, ds.$$

As b is locally bounded, we may differentiate this formula and obtain the desired ODE

$$\frac{dx(t)}{dt} = b(t, x(t)), \quad x(0) = x_0.$$

Since the existence time T is independent of the starting point x_0, the construction can be iterated in time, leading to a global in time solution.

Next, we show the uniqueness of the solution – here, the log-Lipchitz property will play a crucial role. Suppose that $x(t)$ and $y(t)$ are two different solutions of (2.14) with the same initial data, and set $z(t) = x(t) - y(t)$. Then by the log-Lipschitz assumption on b, we have

$$|\dot{z}(t)| \leq C\phi(z(t)), \;\; z(0) = 0.$$

In order to show that $z(t) \equiv 0$, define $f_\delta(t)$ as the solution of

$$\dot{f}_\delta = 2C\phi(f_\delta(t)), \;\; f_\delta(0) = \delta > 0.$$

We claim that $|z(t)| \leq f_\delta(t)$ for all t. Indeed, this is true for some initial time interval, since $\delta > 0$ and both $z(t)$ and $f_\delta(t)$ are continuous. Let $t_1 > 0$ be the smallest time such that $z(t_1) = f_\delta(t_1)$ (the case $z(t_1) = -f_\delta(t_1)$ is similar). But then

$$\dot{z}(t_1) - \dot{f}_\delta(t_1) \leq -C\phi(z(t_1)) < 0,$$

contradicting the definition of t_1. Thus, no such t_1 exists and

$$|z(t)| \leq f_\delta(t) \text{ for all } t \geq 0, \text{ and all } \delta > 0. \tag{2.16}$$

Next, we show that for any $t > 0$ fixed we have

$$\lim_{\delta \to 0^+} f_\delta(t) = 0. \tag{2.17}$$

It suffices to consider the case where δ is small and times are small enough so that $f_\delta(t) < 1$. Then we have

$$\frac{d}{dt} \log f_\delta(t) = 2C(1 - \log f_\delta(t)).$$

Solving this differential equation leads to

$$1 - \log f_\delta(t) = (1 - \log \delta)e^{-2Ct},$$

whence (2.17) follows. Together, (2.16) and (2.17) imply that $z(t) \equiv 0$, that is, the solution $x(t)$ of (2.14) is unique. □

Exercise 2.7 Identify the place the uniqueness proof above where we have used the log-Lipschitz condition on the function $b(t, x)$, that is, where the proof would have failed, for example, for $\phi(r) = r^\beta$, with $\beta \in (0, 1)$.

The approximation scheme

Now let us return to our plan to construct a triple $(\omega, u, \Phi_t(x))$ solving (2.1), (2.2) and (2.3), with the initial vorticity $\omega_0 \in L^\infty$. Let us define a sequence of approximations

$$\begin{aligned}
&\frac{d}{dt}\Phi_t^n(x) = u^n(\Phi_t^n(x), t), \\
&u^n(x,t) = \int_D K_D(x,y)\omega^{n-1}(y,t)\,dy, \\
&\omega^n(x,t) = \omega_0((\Phi_t^n)^{-1}(x)),
\end{aligned} \tag{2.18}$$

with $\omega^0(t,x) \equiv \omega_0(x) \in L^\infty$ for all $t \geq 0$. Each successive approximation involves solving a linear problem

$$\omega_t^n + (u^n \cdot \nabla)\omega^n = 0,$$

with the flow

$$u^n(t,x) = \int_D K_D(x,y)\omega^{n-1}(t,y)\,dy,$$

computed from the previous iteration. By Corollary 2.5 and Lemma 2.6, the solutions are well-defined and unique. Note that each $\omega_n \in L^\infty$, with

$$\|\omega_n(t)\|_{L^\infty} \leq \|\omega_0\|_{L^\infty}.$$

Therefore, Corollary 2.5 implies that all $u^n(x,t)$ are uniformly bounded and log-Lipschitz:

$$|u^n(t,x) - u^n(t,x')| \leq C(D)\phi(|x-x'|). \tag{2.19}$$

The Hölder regularity of the approximate trajectories

We have a uniform continuity bound on $\Phi_t^n(x)$.

Lemma 2.8 *For every n and any x, y with $|x-y| \leq 1$, we have*

$$|x-y|^{e^{Ct}} \leq |\Phi_t^n(x) - \Phi_t^n(y)| \leq C|x-y|^{e^{-Ct}}. \tag{2.20}$$

This is a rather remarkable estimate: we can show that $\Phi_t^n(x)$ is Hölder continuous in space for any $t \geq 0$, but the Hölder exponent deteriorates in time. This is a reflection of the complexity of the dynamics: the exponent in the upper bound in (2.20) tends to zero as $t \to +\infty$ because two trajectories that start very close at $t = 0$ may diverge very far at large times. On the other hand, the exponent in the lower bound in (2.20) grows as $t \to +\infty$ because even if at the time $t = 0$ the starting points x and y are "relatively far apart" (but with $|x-y| \leq 1$), they can extremely close at large times. This deterioration of estimates is not an artefact of the proof – we will later see that the trajectories of the Euler equations can get extremely close at large times.

Proof. Let us fix x and y, and set $F(t) = |\Phi_t(x) - \Phi_t(y)|$. We compute

$$\left|\frac{d}{dt}F^2(t)\right| = 2\left|(\Phi_t^n(x) - \Phi_t^n(y)) \cdot (u(\Phi_t^n(x)) - u(\Phi_t^n(y)))\right| \leq 2CF(t)\phi(F(t)),$$

so that

$$|F'(t)| \le CF(t)\max(1, 1 - \log F(t)).$$

Clearly $|F(t)| \le C(D)$ for all t, so it suffices to consider the case when $F(t) \le 1/2$. Then, we have

$$|F'(t)| \le CF(t) \log F(t)^{-1},$$

which leads to

$$[\log F(0)]e^{Ct} \le \log F(t) \le [\log F(0)]e^{-Ct}.$$

The estimate (2.20) follows immediately from integrating this inequality and taking into account that $F(0) = |x - y|$. □

Convergence of the approximation scheme

Let us now investigate the convergence of the sequence $(\omega^n, u^n, \Phi_t^n)$. The estimate (2.20) implies that for every $T > 0$, we have

$$\Phi_t^n(x) \in C^{\alpha(T)}([0,T] \times D),$$

for some $\alpha(T) > 0$. The Arzela-Ascoli theorem implies that we can find a subsequence n_j such that $\Phi_t^{n_j}(x)$ converges uniformly to $\Phi_t(x) \in C([0,T] \times D)$. Moreover, since (2.20) is uniform in n, the limit $\Phi_t(x)$ also satisfies (2.20) and

$$\Phi_t(x) \in C^{\alpha(T)}([0,T] \times D).$$

In addition, as all Φ_t^n are measure-preserving, so is $\Phi_t(x)$. The lower bound in (2.20) implies that $\Phi_t(x)$ is invertible. As Φ_t^{-1} satisfies the same estimate (2.20), it also belongs to $C^{\alpha(T)}([0,T] \times D)$. We may then define the vorticity

$$\omega(t,x) = \omega_0(\Phi_t^{-1}(x)),$$

and the fluid velocity

$$u(t,x) = \int_D K_D(x,y)\omega(t,y)\,dy.$$

For simplicity of notation, we relabel the subsequence n_j by n.

Lemma 2.9 *We have $|u(t,x) - u_n(t,x)| \to 0$, as $n \to \infty$, uniformly in D for every $t \in [0,T]$.*

Proof. Note that

$$|u(t,x) - u_n(t,x)| = \left| \int_D (K_D(x, \Phi_t(z)) - K_D(x, \Phi_t^n(z)))\, \omega_0(z)\,dz \right|.$$

Given $\varepsilon > 0$, choose n so that $|\Phi_t(x) - \Phi_t^n(x)| < \delta$, for every $x \in D$, $t \in [0,T]$, with $\delta > 0$ to be determined later. Then we have

$$|u(t,x) - u_n(t,x)| \le \|\omega_0\|_{L^\infty} \int_D |K_D(x,z) - K_D(x, y(z))|\,dz. \tag{2.21}$$

Here, the map $y(z) = \Phi_t^n \circ \Phi_t^{-1}(z)$ is measure preserving, and

$$|y(z) - z| = |\Phi_t^n(\Phi_t^{-1}(z)) - \Phi_t(\Phi_t^{-1}(z))| < \delta,$$

for every z. As usual, we split the integral in (2.21) into two regions: in the first one we have

$$\int_{B_{3\delta}(x)\cap D} |K_D(x,z) - K_D(x,y(z))|\, dz \le 2C \int_{B_{3\delta}(x)} \frac{dz}{|x-z|} \le 2C\delta,$$

while in the second

$$\begin{aligned}\int_{B_{3\delta}^c(x)\cap D} |K_D(x,z) - K_D(x,y(z))|\, dz &\le C\delta \int_{B_{3\delta}^c(x)\cap D} |\nabla K_D(x,p(z))|\, dz \\ &\le C\delta \int_{B_\delta^c} \frac{dz}{|x-z|^2} \le C\delta \log \delta^{-1}. \qquad (2.22)\end{aligned}$$

Here, $p(z)$ is a point on a curve of length $\le 2\delta$ that connects z and $y(z)$. If the interval connecting these points lies in D then this interval can be used as this curve. If not, one can use an argument similar to that in the proof of Lemma 2.3. Thus choosing δ sufficiently small we can make sure that the difference of the velocities does not exceed ε. □

Exercise 2.10 Fill in all the details in the last step in the proof of the Lemma.

We are now ready to show that

$$\frac{d}{dt}\Phi_t(x) = u(t, \Phi_t(x)).$$

Indeed, we have

$$\Phi_t^n(x) = x + \int_0^t u^n(\Phi_s^n(x), s)\, ds,$$

and, taking $n \to \infty$, using Lemma 2.9 and the definition of $\Phi_t(x)$, we obtain

$$\Phi_t(x) = x + \int_0^t u(\Phi_s(x), s)\, ds.$$

Thus, the limit triple $(\omega(t,x), u(t,x), \Phi_t(x))$ satisfies the Euler equations, completing the proof of the existence of solutions.

Uniqueness of the solutions

Let us now, finally, state the main result on the existence and uniqueness of the solutions with $\omega_0 \in L^\infty$. The existence part of this theorem summarizes what has been proved above using the approximation scheme.

Theorem 2.11 *Given $T > 0$, there exists $\alpha(T) > 0$ so that for any $\omega_0 \in L^\infty(D)$ there is a unique triple $(\omega(t,x), u(t,x), \Phi_t(x))$, with the vorticity $\omega \in L^\infty([0,T], L^\infty(D))$, the fluid* *velocity $u(t,x)$ uniformly bounded and log-Lipschitz in x, and $\Phi_t \in C^{\alpha(T)}([0,T]\times D)$ a measure preserving, invertible mapping of D, which satisfy*

$$\begin{aligned}
&\frac{d\Phi_t(x)}{dt} = u(\Phi_t(x)), \ \ \Phi_0(x) = x, \\
&\omega(t,x) = \omega_0(\Phi_t^{-1}(x)), \\
&u(t,x) = \int_D K_D(x,y)\omega(y,t)\,dy.
\end{aligned} \tag{2.23}$$

It is clear from the statement of the theorem that $\omega(t,x)$ converges to $\omega_0(x)$ as $t \to 0$ in the weak-$*$ sense in L^∞: for any test function $\eta \in L^1(D)$ we have

$$\int_D \omega(t,x)\eta(x)dx = \int_D \omega_0(\Phi_t^{-1}(x))\eta(x)dx = \int_D \omega_0(x)\eta(\Phi_t(x))dx \to \int_D \omega_0(x)\eta(x)dx, \tag{2.24}$$

as $t \to 0$. Indeed, as ω is uniformly bounded in $L^\infty(D)$, it suffices to check (2.24) for smooth functions η, for which we have

$$\int_D |\eta(\Phi_t(x)) - \eta(x)|dx \le \|\nabla\eta\|_{L^\infty} \int_D |\Phi_t(x) - x|dx \le C(D)\|\nabla\eta\|_{L^\infty}\|u\|_{L^\infty} t.$$

Proof of Theorem 2.11. As we have already established the existence and regularity of the solutions, it remains only to prove the uniqueness. Suppose that there are two solution triples $(\omega^1, u^1, \Phi_t^1)$ and $(\omega^2, u^2, \Phi_t^2)$ and set

$$\eta(t) = \frac{1}{|D|}\int_D |\Phi_t^1(x) - \Phi_t^2(x)|\,dx.$$

Let us write

$$|\Phi_t^1(x) - \Phi_t^2(x)| \le \int_0^t |u^1(s,\Phi_s^1(x)) - u^1(s,\Phi_s^2(x))|\,ds + \int_0^t |u^1(s,\Phi_s^2(x)) - u^2(s,\Phi_s^2(x))|\,ds. \tag{2.25}$$

By Corollary 2.5, the first integral in the right side of (2.25) can be bounded by

$$C\|\omega_0\|_{L^\infty}\int_0^t \phi(|\Phi_s^1(x) - \Phi_s^2(x)|)\,ds.$$

For the second integral in (2.25), consider the difference

$$\begin{aligned}
u^1(s,\Phi_s^2(x)) - u^2(s,\Phi_s^2(x)) &= \int_D K_D(\Phi_s^2(x), y)\omega^1(s,y)\,dy - \int_D K_D(\Phi_s^2(x), y)\omega^2(s,y)\,dy \\
&= \int_D \big(K_D(\Phi_s^2(x), \Phi_s^1(y)) - K_D(\Phi_s^2(x), \Phi_s^2(y))\big)\,\omega_0(y)\,dy.
\end{aligned}$$

Averaging (2.25) in x, we now obtain

$$\begin{aligned}
\eta(t) &\le \frac{C\|\omega_0\|_{L^\infty}}{|D|}\int_0^t ds \int_D \phi(|\Phi_s^1(x) - \Phi_s^2(x)|)\,dx \\
&+ \frac{C}{|D|}\int_0^t ds \int_D |\omega_0(y)| \int_D |K_D(x,\Phi_s^1(y)) - K_D(x,\Phi_s^2(y))|\,dxdy \\
&\le C(D)\|\omega_0\|_{L^\infty}\int_0^t ds \int_D \phi(|\Phi_s^1(x) - \Phi_s^2(x)|)\frac{dx}{|D|}.
\end{aligned} \tag{2.26}$$

We used Lemma 2.3 in the last step. As the function ϕ is concave, we may use Jensen's inequality to exchange ϕ and averaging in the last expression in (2.26):

$$\eta(t) \leq C(D)\|\omega_0\|_{L^\infty} \int_0^t \phi(\eta(s))\, ds.$$

In addition, we have $\eta(0) = 0$. An argument very similar to the proof of uniqueness in Lemma 2.6 (based on the log-Lipschitz property of the function ϕ) can be now used to prove that $\eta(t) = 0$ for all $t \geq 0$.

Exercise 2.12 Work out the details of this argument.

This completes the proof of the theorem. □

Regularity of the solutions for regular initial data

If ω_0 possesses additional regularity, then so does the solution $\omega(t, x)$. This is expressed by the following theorem.

Theorem 2.13 *Suppose that $\omega_0 \in C^k(D)$, $k \geq 1$. Then the solution described in Theorem 2.11, satisfies, in addition, the following properties, for each $t > 0$ fixed:*

$$\omega(t) \in C^k(D), \quad \Phi_t(x) \in C^{k,\alpha(t)}, \text{ and } u \in C^{k,\beta}(D),$$

for all $\beta < 1$. In addition, the kth order derivatives of u are log-Lipschitz.

The first proof of a related result goes back to works of Wolibner and of Hölder in the early 1930s. We will provide a detailed argument for the case of $k = 1$, larger values of k will be left as an exercise. The following result is classical.

Theorem 2.14 *Suppose that D is a domain in $\mathbb{R}^d$ with a smooth boundary, and let ψ be the solution of the Dirichlet problem*

$$-\Delta\psi = \omega,$$
$$\psi|_{\partial D} = 0.$$

If $\omega \in C^\alpha(D)$, $\alpha > 0$, then $\psi \in C^{2,\alpha}(D)$ and $\|\partial_{ij}\psi\|_{C^\alpha} \leq C(\alpha, D)\|\omega\|_{C^\alpha}$.

This result was originally proved by Kellogg in 1931. Schauder later established a similar bound for more general elliptic operators. Such estimates are commonly called the Schauder estimates, see [69].

We have already proved that if $\omega_0 \in L^\infty$, then $\Phi_t^{-1}(x) \in C^{\alpha(t)}(D)$. Since

$$\omega(t, x) = \omega_0(\Phi_t^{-1}(x)),$$

if, in addition, we know that $\omega_0 \in C^1(D)$, we would automatically have $\omega(t, x) \in C^{\alpha(t)}(D)$. By Theorem 2.14, we then have $u(x, t) \in C^{1,\alpha(t)}(D)$. A simple calculation using the trajectories equation leads to

$$\frac{d}{dt}|\Phi_t(x) - \Phi_t(y)|^2 \leq C\|\nabla u(\cdot, t)\|_{L^\infty}|\Phi_t(x) - \Phi_t(y)|^2, \tag{2.27}$$

where we now know that the derivatives of u are bounded for all t, even though their size may grow with time. Integrating (2.27) in time and using the initial condition

$$|\Phi_0(x) - \Phi_0(y)| = |x-y|,$$

we obtain

$$\exp\Big\{ -\int_0^t \|\nabla u(\cdot,s)\|_{L^\infty}\, ds\Big\} \le \frac{|\Phi_t(x)-\Phi_t(y)|}{|x-y|} \le \exp\Big\{ \int_0^t \|\nabla u(\cdot,s)\|_{L^\infty}\, ds\Big\}. \qquad (2.28)$$

This inequality will be useful for us later. For now, we observe that it implies that $\Phi_t(x)$ is Lipschitz for every $t \ge 0$. By the Rademacher theorem (see e.g. [53]), it follows that $\Phi_t(x)$ is differentiable almost everywhere. We would like to show that in fact $\Phi_t(x) \in C^{1,\alpha(t)}(D)$ for every $t \ge 0$. For this purpose we need a couple of technical lemmas. In what follows we adopt the summation convention: we sum over repeated indexes.

Lemma 2.15 *For every $t \ge 0$, for a.e. x, we have*

$$\partial_j \Phi_t^k(x) = \delta_{jk} + \int_0^t \partial_l u^k(\Phi_s(x),s)\partial_j\Phi_s^l(x)\, ds. \qquad (2.29)$$

Proof. Note that at this point we can talk only about almost everywhere representation for the derivatives of $\Phi_t(x)$ since this is what Rademacher theorem gives us. Let $y = x + e_j\Delta x$, where e_j is a unit vector in jth direction. Consider

$$\frac{\Phi_t^k(y)-\Phi_t^k(x)}{\Delta x} = \delta_{jk} + \int_0^t \frac{u^k(\Phi_s(y),s)-u^k(\Phi_s(x),s)}{\Delta x}\, ds. \qquad (2.30)$$

Now

$$\begin{aligned}\frac{u^k(\Phi_s(y),s)-u^k(\Phi_s(x),s)}{\Delta x} &= \frac{u^k(\Phi_s^1(y),\Phi_s^2(y)) - u^k(\Phi_s^1(x),\Phi_s^2(y))}{\Phi_s^1(y)-\Phi_s^1(x)}\,\frac{\Phi_s^1(y)-\Phi_s^1(x)}{\Delta x}\\ &+\frac{u^k(\Phi_s^1(x),\Phi_s^2(y)) - u^k(\Phi_s^1(x),\Phi_s^2(x))}{\Phi_s^2(y)-\Phi_s^2(x)}\,\frac{\Phi_s^2(y)-\Phi_s^2(x)}{\Delta x}.\end{aligned}$$

Since $u \in C^{1,\alpha}(D)$ it is not difficult to show, using mean value theorem, that the first factors in both products on the right hand side converge uniformly in x to $\partial_l u^k(\Phi_s(x),s)$, $l = 1,2$ respectively. On the other hand, the ratios $(\Delta x)^{-1}(\Phi_s^l(y) - \Phi_s^l(x))$ we control in L^∞ by the Lipschitz estimate (2.28). Moreover, by the Fubini theorem, for a.e. x, for a.e. s, the ratio converges to $\partial_j\Phi_s^l(x)$. By the dominated convergence theorem, we have the convergence of the integral in (2.30) to the integral in (2.29). □

Observe that (2.29) allows us to define $\nabla\Phi_t(x)$ for almost every x for all $t \ge 0$.

Lemma 2.16 *For every $t \ge 0$, the function $\partial_j\Phi_t^k(x)$ satisfies Hölder estimate in x for a.e. x. This allows us to extend the function $\partial_j\Phi_t^k(x)$ for all x so that it belongs to $C^{\alpha(t)}(D)$ and (2.29) holds for all x, t.*

Proof. From (2.29) we find that

$$\partial_t \partial_j \Phi_t^k(x) = \partial_l u^k(\Phi_t(x), t) \partial_j \Phi_t^l(x)$$

for a.e. x for all t. Consider

$$\partial_t(\partial_j \Phi_t^k(x) - \partial_j \Phi_t^k(y)) = (\partial_l u^k(\Phi_t(x), t) - \partial_l u^k(\Phi_t(y), t))\partial_j \Phi_t^l(x) + \partial_l u^k(\Phi_t(y), t)(\partial_j \Phi_t^l(x) - \partial_j \Phi_t^l(y)).$$

It follows that

$$\partial_t |\partial_j \Phi_t^k(x) - \partial_j \Phi_t^k(y)| \leq \|\Phi_t\|_{Lip} \|\nabla u\|_{C^{\alpha(t)}} |\Phi_t(x) - \Phi_t(y)|^{\alpha(t)} + \|\nabla u\|_{L^\infty} |\partial_j \Phi_t^l(x) - \partial_j \Phi_t^l(y)|,$$

where we denote by $\|\Phi_t\|_{Lip}$ the Lipschitz bound we have on $\Phi_t(x)$ in x for a given t. Let us denote

$$F(t) = \sum_{k,j} |\partial_j \Phi_t^k(x) - \partial_j \Phi_t^k(y)|.$$

Then we get

$$F'(t) \leq \|\nabla u\|_{L^\infty} F(t) + |x-y|^{\alpha(t)} \|\Phi_t\|_{Lip}^2 \|\nabla u\|_{C^{\alpha(t)}}.$$

By applying Gronwall Lemma we obtain that on any time interval $[0, T]$, the first order derivatives of $\Phi_t(x)$ are Hölder in x with exponent $\alpha(T) > 0$. □

Exercise. Fill in all the details in the proofs of the above two lemmas.

Now the proof of Theorem 2.13 in the case $k = 1$ is immediate.

Proof. [Proof of Theorem 2.13] Since $\Phi_t(x)$ is measure preserving, we have that $\det \nabla \Phi_t = 1$ and the derivatives of the inverse map $\Phi_t^{-1}(x)$ satisfy the same bounds as those of Φ_t. This and Lemma 2.16 imply immediately that $\omega(x,t) = \omega_0(\Phi_t^{-1}(x))$ is $C^1(D)$ for all times. □

Exercise. Carry out the analogous computations for the case of $k > 1$, proving Theorem 2.13 in this case.

2.1 The Kato estimate and upper bound on growth of the gradient of vorticity

In the case of a regular initial vorticity ω_0, an interesting question is how fast the higher derivatives of the solution may grow. This question is linked with small scale creation in fluids, a phenomenon that is ubiquitous in nature and engineering. We witness this process in observing thin filaments in turbulent flows, in the structure of hurricanes and in boiling water in our kitchen. Our main result in this section addresses such upper bound on the growth of small scales in solutions.

Theorem 2.17 *Assume that $\omega_0 \in C^1(D)$. Then the gradient of the solution $\omega(x,t)$ satisfies the following bound*

$$\|\omega(\cdot,t)\|_{C^k} \leq (\|\omega_0\|_{C^k} + 1)^{C \exp \|\omega_0\|_{L^\infty} t} \tag{2.31}$$

for all $t \geq 0$.

Remark. This bound is implicit already in the work of Wolibner; it has been stated explicitly by Yudovich.

A key step in the proof is the following inequality due to Kato.

Proposition 2.18 (Kato) *Let D be a smooth compact domain, $\omega \in C^\alpha(D)$, $\alpha > 0$, $u = \nabla^\perp(-\Delta_D)^{-1}\omega$. Then*

$$\|\nabla u\|_{L^\infty} \leq C(\alpha, D)\|\omega\|_{L^\infty}\left(1+\log\left(1+\frac{\|\omega\|_{C^\alpha}}{\|\omega\|_{L^\infty}}\right)\right). \tag{2.32}$$

Remarks. 1. The operators $\partial_{jk}(-\Delta)^{-1}$ are called (iterated) Riesz transforms. Calderon-Zygmund theory proves that Riesz transforms are bounded on all L^p, $1 < p < \infty$ (see e.g. [109]). The derivatives of the fluid velocity u are exactly Riesz transforms of vorticity. However we need L^∞ bound since this is what appears in (2.28). The L^∞ bound on Riesz transform is not true, and we need a little extra - a logarithm - of a higher order norm of ω to control the L^∞ norm of ∇u.
2. The proposition also has applications to three dimensional case, where it leads to a well known conditional regularity statement for the solutions of 3D Euler equation called Beale-Kato-Majda criterion [11]. In three dimensions, there is no control on $\|\omega\|_{L^\infty}$ anymore. However, using the bound (2.32), one can show that finiteness of the integral $\int_0^T \|\omega\|_{L^\infty}\,dt$ implies regularity of the solution on $[0,T]$. Thus $\|\omega\|_{L^\infty}$ "controls" the possible blow up in 3D case: solutions cannot develop a singularity without $\int_0^T \|\omega\|_{L^\infty}\,dt$ also becoming infinite.

Before proving Proposition 2.18, we need the following lemma.

Lemma 2.19 *Let D be a smooth compact domain, and $u(x)$ be given by the Biot-Savart law*

$$u(x) = \int_D K_D(x,y)\omega(y)\,dy.$$

Then

$$\nabla u(x) = P.V.\int_D \nabla K_D(x,y)\omega(y)\,dy + \frac{(-1)^i}{2}\omega(x)(1-\delta_{ij}). \tag{2.33}$$

Proof. Notice that the derivative of u needs to be defined and computed in the distributional sense due to the singularity of the kernel. Such computation is fairly standard and can be found for example in [87].

Exercise. Carry out the computation to verify (2.33).

□

Now we are ready to prove Proposition 2.18.

Proof. [Proof of Proposition 2.18] Let

$$\delta = \min\left(c, \left(\frac{\|\omega_0\|_{L^\infty}}{\|\omega(x,t)\|_{C^\alpha}}\right)^{1/\alpha}\right),$$

where $c > 0$ is some fixed constant that depends on D, chosen so that the set of points $x \in D$ with $\mathrm{dist}(x, \partial D) \geq 2\delta$ is not empty. Consider first any interior point x such that $\mathrm{dist}(x,\partial D) \geq 2\delta$. Let us look at the representation (2.33). The part of the integral over the complement of the ball centered at x with radius δ can be estimated as

$$\left|\int_{B^c_\delta(x)} \nabla K_D(x,y)\omega(y)\,dy\right| \leq C\|\omega_0\|_{L^\infty}\int_{B^c_\delta(x)} |x-y|^{-2}\,dy \leq C\|\omega_0\|_{L^\infty}(1+\log\delta^{-1}), \tag{2.34}$$

where we used a bound (2.7) from the Proposition 2.2.

Now recall that the Dirichlet Green's function is given by

$$G_D(z, y) = \frac{1}{2\pi} \log|z - y| + h(z, y), \tag{2.35}$$

where h is harmonic in D in z for each fixed y and has boundary values $-\frac{1}{2\pi}\log|z-y|$. Any second order partial derivative at $z = x$ of the first summand on the right hand side of (2.35) is of the form $r^{-2}\Omega(\phi)$ where r, ϕ are radial variables centered at x, and $\Omega(\phi)$ is mean zero. For this part, we can write

$$\begin{aligned} \left| P.V. \int_{B_\delta(x)} \partial^2_{x_i x_j} \log|x-y|\omega(y)\,dy \right| &= \left| \int_{B_\delta(x)} \partial^2_{x_i x_j} \log|x-y|(\omega(y)-\omega(x))\,dy \right| \\ \le C\|\omega(x,t)\|_{C^\alpha} \int_0^\delta r^{-1+\alpha}\,dr &\le C(\alpha)\delta^\alpha \|\omega(x,t)\|_{C^\alpha} \le C(\alpha)\|\omega_0\|_{L^\infty} \end{aligned} \tag{2.36}$$

by our choice of δ. Finally, notice that our assumptions on x, the boundary values for h, and the maximum principle together guarantee that we have $|h(z,y)| \le C\log\delta^{-1}$ for all $y \in B_\delta(x)$, $z \in D$. Standard estimates for harmonic functions (see e.g. [52]) give, for each fixed $y \in B_\delta(x)$,

$$|\partial^2_{x_i x_j} h(x,y)| \le C\delta^{-4}\|h(z,y)\|_{L^1(B_\delta(x),dz)} \le C\delta^{-2}\log\delta^{-1}.$$

This gives

$$\left| \int_{B_\delta(x)} \partial^2_{x_i x_j} h(x,y)\omega(y,t)\,dy \right| \le C\|\omega_0\|_{L^\infty}\log\delta^{-1}. \tag{2.37}$$

Together, (2.37), (2.36) and (2.34) prove the Proposition at interior points.

Now if x' is such that $\mathrm{dist}(x', \partial D) < 2\delta$, find a point x such that $\mathrm{dist}(x, \partial D) \ge 2\delta$ and $|x' - x| \le C(D)\delta$. By Schauder estimate (see Theorem 2.14) we have

$$|\nabla u(x') - \nabla u(x)| \le C(\alpha, D)\delta^\alpha \|\omega\|_{C^\alpha}. \tag{2.38}$$

At x, interior bounds apply, which together with Theorem 2.14 gives desired bound at any $x' \in D$. □

Given Proposition 2.18, the proof of the estimate (2.31) and so of Theorem 2.17 follows.

Proof. [Proof of Theorem 2.17] Let us come back to the two sided bound (2.28) and use the estimate (2.32). We obtain

$$f(t)^{-1} \le \frac{|\Phi_t(x) - \Phi_t(y)|}{|x-y|} \le f(t), \tag{2.39}$$

where

$$f(t) = \exp\left(C\|\omega_0\|_{L^\infty} \int_0^t \left(1 + \log\left(1 + \frac{\|\nabla\omega(x,s)\|_{L^\infty}}{\|\omega_0\|_{L^\infty}} \right) \right) ds \right).$$

Of course, the bound (2.39) also holds for Φ_t^{-1}. On the other hand,

$$\|\nabla\omega(x,t)\|_{L^\infty} = \mathrm{sup}_{x,y} \frac{|\omega_0(\Phi_t^{-1}(x)) - \omega_0(\Phi_t^{-1}(y))|}{|x-y|} \le \|\nabla\omega_0\|_{L^\infty} \mathrm{sup}_{x,y} \frac{|\Phi_t^{-1}(x) - \Phi_t^{-1}(y)|}{|x-y|}. \tag{2.40}$$

Combining (2.40) and (2.39), we obtain

$$\|\nabla\omega(x,t)\|_{L^\infty} \le \|\nabla\omega_0\|_{L^\infty} \exp\left(C\|\omega_0\|_{L^\infty} \int_0^t \left(1+\log\left(1+\frac{\|\nabla\omega(x,s)\|_{L^\infty}}{\|\omega_0\|_{L^\infty}}\right)\right) ds \right),$$

or

$$\log\|\nabla\omega(x,t)\|_{L^\infty} \le \log\|\nabla\omega_0\|_{L^\infty} + C\|\omega_0\|_{L^\infty} \int_0^t \left(1+\log\left(1+\frac{\|\nabla\omega(x,s)\|_{L^\infty}}{\|\omega_0\|_{L^\infty}}\right)\right) ds.$$

Let $A = ||\omega_0||_{L^\infty}$, $B = ||\nabla\omega_0||_{L^\infty}$ and consider the solution $y = y(t)$ of

$$\frac{y'}{y} = CA\,(1+\log(1+y))\,, \qquad y(0) = \frac{B}{A} = y_0\,. \tag{2.41}$$

By Gronwall's lemma it is enough to bound $y(t)$. The solution of (2.41) is given by

$$\int_{y_0}^{y(t)} \frac{dy}{y\,(1+\log(1+y))} = CAt\,. \tag{2.42}$$

Hence

$$\log\left(1+\log(1+y(t))\right) - \log\left(1+\log(1+y_0)\right) + \int_{y_0}^{y(t)} dy\,\left[\frac{1}{y(1+\log(1+y))} - \frac{1}{(1+y)(1+\log(1+y))}\right] = CAt\,.$$

The integrand in the last expression is positive and hence

$$1+\log(1+y(t)) \le (1+\log(1+y_0))\exp(CAt)\,. \tag{2.43}$$

This implies the double exponential upper bound we seek. □

The question of how sharp the double exponential bound is has been open for a long time. This is what we will discuss in the next three sections.

2.2 The Denisov example

The first works constructing examples with growth in derivatives of vorticity are due to Yudovich [119, 120]. He considered growth on the boundary of the domain, and his construction required that the boundary has a flat piece. The bounds on the growth are not explicit, but it is shown that $\limsup_{t\to\infty} \|\nabla\omega(\cdot,t)\|_{L^\infty} = \infty$. Generally, small scale generation at the boundary fits well with physical intuition. It is known that boundaries are important for fluid motion, and in particular influence generation of turbulence (see e.g. [68]). In later works [81, 92], it was shown that the small scale generation at the boundary is in some sense generic.

Nadirashvili [96] has constructed examples with linear growth in the vorticity gradient in the bulk of the domain in the case of an annulus. He called such solutions "wandering", since, at least in a relatively strong norm, they travel to infinity as time passes. The argument is based on constructing a stable background flow that can stretch a small perturbation, creating small scales. We will outline a similar philosophy in more detail below when discussing

Denisov's example. This example provides the best known rate of growth for the gradient of vorticity in the bulk of the fluid, away from the boundary, when starting with smooth initial data. The example is set on the torus $\mathbb{T}^2$. The existence and global regularity of the solutions to 2D Euler equation on the torus can be proved similarly to the bounded domain case considered in the previous section. We will spell out some of the differences (such as the form of the Biot-Savart law) below. Amazingly, the rate of growth that the example provides is just superlinear, leaving a huge gap with the double exponential upper bound. In Section 2.4 below, we will see an example showing that growth on the boundary (as opposed to in the bulk) can indeed happen at a double exponential rate.

Let us now build an explicit example of solution to 2D Euler equation with infinite growth of the gradient. This example is due to Denisov [45]. Such specific examples are very useful for developing intuition. The basic idea behind building this example will be simple: try finding a stable stationary flow, and perturb it a little. The background stable flow should be chosen so that it will then stretch the perturbation, creating gradients. Since it is stable, the process will continue indefinitely, resulting in unbounded gradient growth. The plan is not easy to implement, however, since the equation is strongly nonlinear and nonlocal. No matter how small the perturbation is, it will interact with the background flow, and this interaction is difficult to control for large times. Let us start with some examples of stationary 2D Euler flows and see what sort of small scale creation they can be expected to provide.

In the periodic case, the Biot-Savart law is given by $u = \nabla^\perp(-\Delta)^{-1}\omega$, where $-\Delta$ is the Lapalacian on $\mathbb{T}^2$. Suppose for simplicity that our torus $\mathbb{T}^2$ has size 2π in both directions. This is of course not crucial but will make the computations simpler by eliminating some constant factors. The inverse of the Laplacian is easiest to define through Fourier transform:

$$(-\Delta)^{-1}f(x) = \sum_{k\in\mathbb{Z}^2} e^{ikx}|k|^{-2}\hat{f}(k),$$

where

$$\hat{f}(k) = \int_{\mathbb{T}^2} e^{-ikx} f(x)\,dx.$$

Note that the inverse Laplacian is only defined on functions which have mean zero. We will assume in this section that ω_0 satisfies this requirement. It is not hard to see that the solution $\omega(x,t)$ in this case satisfies the same requirement for all times.

A stationary flow satisfies $(u\cdot\nabla)\omega = 0$, or $\nabla^\perp(-\Delta)^{-1}\omega\cdot\nabla\omega = 0$ at every x. Denote $\psi = (-\Delta)^{-1}\omega$ the stream function of the flow. The flow is clearly stationary if $-\Delta\psi = f(\psi)$ for some smooth function f. The simplest examples of stream functions of stationary flows are just eigenfunctions of periodic Laplacian.

Example 1. Shear flow: $\psi(x_1,x_2) = \omega(x_1,x_2) = \cos x_2$; $u(x_1,x_2) = (\sin x_2, 0)$. This is a flow with straight line trajectories. Indeed, the characteristics are given by $X_1 = x_1 - t\sin x_2$, $X_2 = x_2$. Consider the passive scalar equation

$$\partial_t\varphi + (u\cdot\nabla)\varphi = 0, \;\; \varphi(x,0) = \varphi_0(x). \tag{2.44}$$

The equation has the same form as 2D Euler, except u and φ are independent; u is given and the initial data φ_0 is arbitrary, and doesn't have to be the vorticity corresponding to u. Taking a shear flow in (2.44), we can solve for $\varphi(x,t)$ explicitly. We have $\varphi(X_1,X_2,t) = \varphi_0(x_1,x_2)$, and taking into account formulas for X_1, X_2 we get $\varphi(x_1, x_2, t) = \varphi_0(x_1 + t\sin x_2, x_2)$. On the torus, all coordinates are taken modulo 2π. This suggests that we can only expect linear growth in the gradient and higher order Sobolev norms of φ if we adopt shear flow as our background flow. This is closely related to the setting of Nadirashvili example where he proves linear growth.

Example 2. Cellular flow. A simple cellular flow is given by $\psi(x_1, x_2) = \frac{1}{2}\omega(x_1, x_2) = \sin x_1 \sin x_2$; $u(x_1, x_2) = (\cos x_2 \sin x_1, -\cos x_1 \sin x_2)$. This flow has four vortices in the four quadrants of the plane, and, in particular, a hyperbolic point near the origin. The x_2 direction is contracting. A trajectory starting at $(0, x_2)$ is a straight line with $X_2(t)$ a solution of $X_2' = -\sin X_2$. Thus if x_2 is small, then $X_2(t) \sim x_2 e^{-t}$. Then for a solution $\varphi(x_1, x_2, t)$ of the passive scalar equation (2.44) with such u, we have $\varphi(0, x_2, t) \sim \varphi_0(0, x_2 e^t)$. Therefore exponential growth in gradient can be expected here. The example that we will discuss is a version of a cellular flow; for technical reasons it will be defined a little differently. Also, due to nonlinearity and nonlocality of 2D Euler equation, we will not be able to prove exponential growth but will settle for a weaker result.

Before we start the actual construction, it is worthwhile to note that even for the simpler passive scalar, we only presented a scenario with exponential growth, but not double exponential. For smooth solutions, the double exponential growth is strictly nonlinear phenomena and cannot be captured by passive scalar. However there is a stationary flow generated by singular stationary vorticity that can lead to double exponential growth in passive scalar. Such example is due to Bahouri and Chemin, and we will discuss it in the next section.

Exercise. To explain the statement that double exponential growth cannot happen in a smooth passive flow, prove that if u is smooth in (2.44), then $\|\varphi(x,t)\|_s \leq Ce^{Ct}$ for all times, where C depends only on u, s and φ_0.

The following more explicit form of the periodic Biot-Savart law will be useful for us in the construction.

Proposition 2.20 *Let $\omega \in L^\infty(\mathbb{T}^2)$ be a mean zero function. Then the vector field $u = \nabla^\perp(-\Delta)^{-1}\omega$ is given by*

$$u(x) = -\frac{1}{2\pi}\lim_{\gamma\to 0}\int_{\mathbb{R}^2} \frac{(x-y)^\perp}{|x-y|^2}\omega(y)e^{-\gamma|y|^2}\,dy, \tag{2.45}$$

where ω has been extended periodically to all $\mathbb{R}^2$.

Proof. We have by definition $u = \nabla^\perp(-\Delta)^{-1}\omega$, that is

$$u(x) = \sum_{k\in\mathbb{Z}^2} e^{ikx}\frac{ik^\perp}{|k|^2}\hat{\omega}(k).$$

To link this expression with (2.45), observe first that for a smooth ω,

$$\sum_{k\in\mathbb{Z}^2} e^{ikx}\frac{ik^\perp}{|k|^2}\hat{\omega}(k) = \lim_{\gamma\to 0}\int_{\mathbb{R}^2} e^{ipx}\frac{ip^\perp}{|p|^2}\int_{\mathbb{R}^2} e^{-ipy-\gamma|y|^2}\omega(y)\,dydp, \tag{2.46}$$

where the function $\omega(y)$ is extended periodically to the whole plane.

Exercise. Check the above identity by substituting Fourier series for $\omega(y)$ on the right hand side and integrating in y to obtain Gaussian approximation of identity.

On the other hand, recall that the inverse Laplacian on the whole plane is given by

$$(-\Delta)^{-1}f(x) = \int_{\mathbb{R}^2} e^{ikx}\frac{1}{|k|^2}\int_{\mathbb{R}^2} e^{-iky}f(y)\,dy = -\frac{1}{2\pi}\int_{\mathbb{R}^2}\log|x-y|f(y)\,dy$$

if the function f is sufficiently regular and quickly decaying.

Therefore the expression on the right hand side of (2.46) is equal to

$$\int_{\mathbb{R}^2} e^{ipx}\frac{1}{|p|^2}\int_{\mathbb{R}^2} e^{-ipy}\nabla^{\perp}\left(\omega(y)e^{-\gamma|y|^2}\right)\,dydp = -\frac{1}{2\pi}\int_{\mathbb{R}^2}\log|x-y|\nabla^{\perp}\left(\omega(y)e^{-\gamma|y|^2}\right).$$

Integrating by parts we obtain (2.45). $\Box$

Exercise. Observe that the decay of the kernel in (2.45) is not sufficient to guarantee the convergence of the integral when $\gamma = 0$. Prove that nevertheless the limit on (2.45) exists and is finite if ω is periodic, mean zero and bounded.

The formula (2.45) shows that 2D Euler evolution on torus can be equivalently viewed as evolution on the plane with periodic initial data if we understand Biot-Savart law in the sense of the principal value integral (2.45).

We now start the construction of an example where gradient of the solution of 2D Euler equation grows faster than linearly. More precisely, we will prove the following theorem.

Theorem 2.21 *There exists $\omega_0 \in C^\infty(\mathbb{T}^2)$ such that for the corresponding solution of 2D Euler equation $\omega(x,t)$ we have*

$$\frac{1}{T^2}\int_0^T \|\nabla\omega(\cdot,t)\|_{L^\infty}\,dt \overset{T\to\infty}{\longrightarrow} +\infty. \tag{2.47}$$

This shows faster than linear growth on average, or on a subsequence of times tending to infinity.

Our basic background flow will be really similar to that in the Example 2 above, in fact it will be the same flow, but arranged slightly differently. We will set $\omega^*(x_1,x_2) = \cos x_1 + \cos x_2$. Then the stream function $\psi^*(x_1,x_2) = \omega^*(x_1,x_2)$, and $u^*(x_1,x_2) = (-\sin x_2, \sin x_1)$. The torus $(-\pi,\pi]^2$ contains two stagnation points of the flow, $(0,0)$ and (π,π). The four lines $x_2 = \pm x_1 \pm \pi$ are separatrices of the flow, and the points $(\pi,0)$ and $(0,\pi)$ where the separatrices intersect are hyperbolic points. Consider the $D \equiv (\pi,0)$ point. The change of coordinates $\xi = (x_1+x_2-\pi)/2,\ \eta = (x_2-x_1+\pi)/2$ transforms characteristic equations near the point D to

$$\xi' = \sin\xi\cos\eta,\ \ \eta' = -\sin\eta\cos\xi.$$

Let us now consider the 2D Euler equation

$$\partial_t\omega + (u\cdot\nabla)\omega = 0,\ \ u = (\partial_2(-\Delta)^{-1}\omega, -\partial_1(-\Delta)^{-1}\omega),\ \ \omega(x,0) = \omega_0(x),\ \ x\in\mathbb{T}^2. \tag{2.48}$$

We set $\omega(x,t) = \omega^*(x) + \varphi(x,t)$ and $u(x,t) = u^*(x) + v(x,t)$. In the construction, we will take $\varphi(x,0)$ as a small perturbation of $\omega^*(x)$. We will need several auxiliary lemmas in the proof, starting with stability and symmetry lemmas proving that solution will remain close to $\omega^*(x)$ in L^2 sense. Let P_1 be the orthogonal projector on the unit sphere in $\mathbb{Z}^2$ on Fourier side, and P_2 be the projection on the orthogonal complement of functions supported on the unit sphere on Fourier side. Namely, if $f(x) = \sum_{k\in\mathbb{Z}^2} \hat{f}(k)e^{ikx}$, then $P_1 f(x) = \sum_{|k|=1} \hat{f}(k)e^{ikx}$, $P_2 f(x) = \sum_{|k|\neq 1} \hat{f}(k)e^{ikx}$.

Lemma 2.22 *[Stability Lemma] Suppose that the initial data $\omega_0(x)$ is mean zero. Suppose that $\|P_2\omega_0(x)\|_{L^2} \leq \varepsilon$ for some $\varepsilon > 0$. Then $\|P_2\omega(\cdot,t)\|_{L^2} \leq \sqrt{2}\varepsilon$ for all $t > 0$.*

Proof. Observe that the mean zero property is conserved by 2D Euler evolution, as can be checked by integrating (2.48) over $\mathbb{T}^2$ and integrating by parts in nonlinear term. Next, observe that the following two quantities are conserved by Euler evolution:

$$\int_{\mathbb{T}^2} \omega(x,t)^2\,dx = C_1, \quad \int_{\mathbb{T}^2} \omega(x,t)\psi(x,t)\,dx = C_2. \tag{2.49}$$

Note that the second quantity in (2.49) is just the energy of the flow $\int_{\mathbb{T}^2} |u|^2\,dx$.

Exercise. Prove (2.49) directly from (2.48).

Observe now that

$$\sum_{|k|>1}\left(1-\frac{1}{|k|^2}\right)|\hat{\omega}(k,t)|^2 = C_1 - C_2$$

does not depend on time. At time $t=0$ by assumption this expression does not exceed ε^2. The same is then true for all times. But since $1-|k|^{-2} \geq 1/2$ if $|k|>1$, it follows that $\|P_2\omega(\cdot,t)\|_{L^2}^2 \leq 2\varepsilon^2$. $\Box$

Note that the Fourier transform of ω^* is supported on the unit sphere in $\mathbb{Z}^2$, with $\hat{\omega^*}(1,0) = \hat{\omega^*}(-1,0) = \hat{\omega^*}(0,1) = \hat{\omega^*}(0,-1) = 1/2$. We also work with real valued solutions $\omega(x,t)$, so $\hat{\omega}(k,t) = \overline{\hat{\omega}}(-k,t)$. Yet the Stability Lemma alone is not enough to conclude L^2 stability of ω^* to small perturbations, as energy might shift between different modes with $|k|=1$.

Lemma 2.23 *[Symmetries Lemma] Consider 2D Euler equation on $\mathbb{T}^2$. Let $\omega_0(x)$ be smooth initial data. Assume that ω_0 is even: $\omega_0(x) = \omega_0(-x)$. Then the solution $\omega(x,t)$ remains even for all $t>0$. Assume that ω_0 is invariant under rotation by $\pi/2 : \omega_0(x_1,x_2) = \omega_0(-x_2,x_1)$. Then the solution $\omega(x,t)$ remains invariant under rotation by $\pi/2$.*

Proof. The proof uses uniqueness of smooth solutions to (2.48). We show that if $\omega(x_1,x_2,t)$ is a solution, then so is $\omega(-x_1,-x_2,t)$ and $\omega(-x_2,x_1,t)$. Given the assumption on symmetry of initial data, this would imply, by uniqueness, that the solution must possess the same symmetry.

To prove that $\omega(-x_1,-x_2,t)$ and $\omega(-x_2,x_1,t)$ are also solutions, consider the Fourier transform of the 2D Euler equation:

$$\partial_t\hat{\omega}(k) = \sum_{l+m=k}\frac{\langle m^\perp, l\rangle}{|m|^2}\hat{\omega}(m)\hat{\omega}(l), \quad m^\perp = (-m_2,m_1). \tag{2.50}$$

Consider $\omega_1(x,t) = \omega(-x,t)$. Observe that $\hat{\omega}_1(k) = \overline{\hat{\omega}}(k)$. Applying complex conjugation to (2.50), we see that ω_1 solves 2D Euler equation, too.

Similarly, consider $\omega_2(x_1, x_2, t) = \omega(-x_2, x_1, t)$. A simple computation shows that in this case $\hat{\omega}_2(k_1, k_2, t) = \hat{\omega}(-k_2, k_1, t)$. Then

$$\partial_t \hat{\omega}(k_1, k_2, t) = \partial_t \hat{\omega}(-k_2, k_1, t) = \sum_{l+m=k^{\perp}} \frac{\langle m^{\perp}, l\rangle}{|m|^2} \hat{\omega}(m_1, m_2, t)\hat{\omega}(l_1, l_2, t).$$

Relabeling indices, we get that ω_2 solves 2D Euler equation. ☐

Exercise. Does 2D Euler equation preserve the $\omega(x_1, x_2) = \omega(-x_1, x_2)$ symmetry? How about $\omega(x_1, x_2) = -\omega(-x_1, x_2)$? $\omega(x_1, x_2) = \omega(x_2, x_1)$? $\omega(x_1, x_2) = -\omega(-x_1, -x_2)$?

Now suppose that our perturbation $\varphi(x, 0)$ of $\omega^*(x)$ is such that $\omega_0(x)$ is even and symmetric under rotation by $\pi/2$. Then the following Corollary follows from Lemma 2.23 and Lemma 2.22.

Corollary 2.24 *Suppose that $\|\omega_0(x) - \omega^*(x)\|_{L^2} \leq \varepsilon$. Then $\|\omega(x, t) - \omega^*(x)\|_{L^2} \leq C_0\varepsilon$ for all times $t > 0$.*

Proof. Lemma 2.22 implies that $\|P_2\omega(\cdot, t)\|_{L^2} \leq \sqrt{2}\varepsilon$. Furthermore, Lemma 2.23 implies that $\hat{\omega}(1, 0, t) = \hat{\omega}(0, 1, t) = \hat{\omega}(-1, 0, t) = \hat{\omega}(0, -1, t)$ and are real valued. Therefore, $\omega(x, t) = c(t)\omega^*(x) + P_2\omega(x, t)$. A calculation leads to the estimates $|c(t) - 1| \leq C_1\varepsilon$, from which the Corollary follows. ☐

Before we start the construction let us state one more general elementary lemma we will need.

Lemma 2.25 *Let $a_j > 0$ be such that $\sum_{j=1}^{\infty} a_j < \infty$. Then*

$$\frac{1}{N^2} \sum_{j=1}^{N} a_j^{-1} \overset{N\to\infty}{\longrightarrow} \infty.$$

Proof. Observe that

$$\min_{x_i > 0, x_1 + \dots + x_n = \sigma} \sum_{i=1}^{n} x_i^{-1} = n^2 \sigma^{-1},$$

and the minimum is achieved when $x_i = \sigma/n$ for each i.

Exercise Prove this claim.

Now set

$$\tau_N = \sum_{j=N/2}^{N} a_j \overset{N\to\infty}{\longrightarrow} 0.$$

Finally, note that

$$\frac{1}{N^2} \sum_{j=1}^{\infty} a_j^{-1} \geq \frac{1}{N^2} \sum_{j=N/2}^{N} a_j^{-1} \geq N^{-2} \frac{N^2}{4} \frac{1}{\tau_N} \to \infty$$

as $N \to \infty$. ☐

Now let us designate the initial data that we will use in our construction. While the construction is carried out on the torus, it is sometimes more convenient to think of functions defined on $\mathbb{R}^2$ which are $2\pi-$periodic in both x_1 and x_2. Let U_δ be a disc of radius $\sqrt{\delta}$ centered at the origin $(0,0)$. Recall the coordinates $\xi = (x_1+x_2-\pi)/2,\ \eta = (x_2-x_1+\pi)/2$ in which the saddle point D corresponds to $\xi=\eta=0$. The direction ξ is expanding at D and the direction η is contracting. Define $P_\delta = \{|\xi|<0.1,\ |\eta|<\delta\}$. Rotate P_δ by $\pi/2$ around the origin in the original (x_1,x_2) coordinate system and denote its image by P'_δ. Consider ω_0 defined on $\mathbb{T}^2$ as follows.

- $\omega_0(x) = \omega^*(x)$ outside P_δ, P'_δ, and U_δ.
- In (ξ,η) coordinate system, $\omega_0 = f(\xi,\eta)$ in P_δ. The function $f \in C_0^\infty(P_\delta)$ is even, and satisfies $4 \geq f \geq -1$. The level set $f(\xi,\eta)=4$ is equal to $\{\eta=0,\ |\xi|\leq 0.08\}$, the level set $f(\xi,\eta)=3$ is equal to the ellipse $\{(\xi/0.09)^2+(2\eta/\delta)^2=1\}$. In P'_δ, define $\omega_0(x)$ so that it is invariant under rotation by $\pi/2$ with respect to the origin.
- Inside U_δ, we set $\omega_0 = \omega^* + \phi_\delta$, where $\phi_\delta \in C_0^\infty(U_\delta)$ is designed so that $\int_{\mathbb{T}^2}\omega_0(x)\,dx = 0$ and ω_0 obeys symmetry conditions as described in the next item.
- ω_0 is even and symmetric with respect to rotation by $\pi/2$.

Observe that by this construction, ω_0 is smooth. We note also that since we chose f to be even in (ξ,η) coordinates, and ω^* is even with respect to the hyperbolic point D as well, then ω_0 is even with respect to the point $D: \omega_0(x_1-\pi,x_2) = \omega_0(-x_1-\pi,-x_2)$. By Lemma 2.23, solution $\omega(x,t)$ inherits this property. Moreover, if a function $g(x)$ is even with respect to some point, then $(-\Delta)^{-1}g(x)$ is also even with respect to the same point. This has a useful consequence that $u(D,t) = \nabla^\perp(-\Delta)^{-1}\omega(D,t) = 0$ for all time, so the point D is left fixed by the flow. Notice that D is not hyperbolic anymore, as our definition of $f(\xi,\eta)$ destroys hyperbolicity near D. However, the flow still possesses hyperbolic structure outside the small region near D, and we will use this to prove the growth of $\nabla\omega$.

Let us recall our notation $\omega(x,t) = \omega^*(x)+\varphi(x,t)$, $u(x,t) = u^*(x)+v(x,t)$. By Corollary 2.24 and definition of ω_0, $\|\varphi(\cdot,t)\|_{L^2} \leq C\delta^{1/2}$ for all $t\geq 0$. Due to L^∞ maximum principle for $\omega(x,t)$, we also have $\|\varphi(x,t)\|_{L^\infty}\leq C$. Interpolating, we get $\|\varphi(\cdot,t)\|_{L^p}\leq C\delta^{1/p}$, for every $p\geq 2$. We also have

Lemma 2.26
$$\|v(\cdot,t)\|_{L^\infty} \leq C(p)\delta^{1/p},$$
for every $p>2$.

Proof. Recall that by (2.45),
$$v_{1,2}(x,t) = \frac{1}{2\pi}\lim_{\gamma\to 0}\int_{\mathbb{R}^2}\frac{\mp y_{2,1}}{|y|^2}\varphi(x-y,t)e^{-\gamma|x-y|^2}\,dy,$$
where φ is extended periodically to all $\mathbb{R}^2$. Split integration into two parts, over the unit ball B_1 and its complement. Then by Hölder inequality,
$$\left|\int_{B_1}|y|^{-1}|\varphi(x-y,t)|\,dy\right| \leq C\|\varphi(\cdot,t)\|_{L^p}(2-q)^{-1/q} \leq C(p)\delta^{1/p},$$

where $p^{-1} + q^{-1} = 1$, $p > 2$. For the rest of the estimate, set $\varphi = \Delta\psi$, and integrate by parts. We obtain

$$\left| \int_{B_1^c} \frac{y_{1,2}}{|y|^2} \Delta\psi(x-y) e^{-\gamma|x-y|^2}\, dy \right| \leq C \int_{\partial B_1} \left(\left| \frac{\partial \psi}{\partial n} \right| + |\psi| \right) d\sigma + \int_{B_1^c} \psi(x-y) \Delta \left(\frac{y_{1,2}}{|y|^2} e^{-\gamma|x-y|^2} \right) dy.$$

Using Sobolev imbedding theorem and the bounds we have for φ, as well as some straightforward estimates, we can complete the proof of the lemma.

Exercise Carry out the calculations carefully to complete the proof.
$\square$

Let us now choose δ so that $\|v(\cdot,t)\|_{L^\infty} \leq 0.001$ for all t. Let us zoom into the point D. Characteristic curves near D are given by solutions of $x_1'(t) = \sin x_2 - v_1(x,t)$, $x_2'(t) = -\sin x_1 - v_2(x,t)$. In the ξ, η coordinates this becomes $\xi' = \cos\eta \sin\xi - (v_1 + v_2)/2$, $\eta' = \sin\eta\cos\xi + (v_1 - v_2)/2$. We will write this in a shortcut notation

$$\xi' = \sin\xi\cos\eta + \mu_1, \quad \eta' = -\sin\eta\cos\xi + \mu_2, \tag{2.51}$$

where $\|\mu_{1,2}\|_{L^\infty} \leq 0.001$.

Lemma 2.27 *Consider the Cauchy problem* (2.51) *with initial data* $\xi(t_0) = \xi_0$, $\eta(t_0) = \eta_0$. *If* $|\xi_0| \leq 0.03$ *and* $|\eta_0| \leq 0.1$, *then* $|\eta(t_0+1)| \leq 0.1$. *Also, if* $0.03 > |\xi_0| > 0.02$ *and* $|\eta_0| < 0.1$, *then* $|\xi(t_0+1)| > 0.03$. *More generally, if* $0.03 > |\xi_0| > (3-\tau)/100$, $0 \leq \tau \leq 1$, *and* $|\eta_0| < 0.1$, *then* $|\xi(t_0+\tau)| > 0.03$.

Proof. Observe that $|\xi'| < |\xi| + 0.001$, $|\xi_0| \leq 0.03$ imply that

$$|\xi(t)| \leq |\xi_0| e + 0.001 \int_0^1 e^t\, dt < 0.04e$$

for $t \in (t_0, t_0+1)$. Now at $\eta = 0.1$ we have $\eta' \geq -\sin 0.1 \cos 0.2 + 0.001 < 0$ for all times in (t_0, t_0+1) and so the trajectory cannot pass or arrive at this value of η. The case of $\eta = -0.1$ is similar. Thus, for $t \in (t_0, t_0+1)$ we have $|\eta(t)| < 0.1$. For the second statement of the lemma notice that for $0.04 > \xi_0 > 0.02$, due to bounds we showed, $\xi'(t) \geq 0.9\xi - 0.001$ in the time interval (t_0, t_0+1). Then $\xi(t) \geq 0.02e^{0.9} - 0.001(e-1) > 0.003$.

For the last statement, following the same estimates, we have to check that

$$(3-\tau)e^{0.9\tau} - 0.1(e^\tau - 1) \geq (3-\tau)(1+0.9\tau) - 0.3\tau \geq 3 + 0.5\tau \geq 3,$$

which is correct. $\square$

Now we are ready to prove Theorem 2.21.

Proof. [Proof of Theorem 2.21] Denote R_s the rectangle $|\eta| < 0.1$, $|\xi| < 0.01s$. Denote $E(t_1,t_2)$ the Euler flow map from time t_1 to time t_2. $E(t_1,t_2)$ is a smooth area preserving diffeomorphism given by solutions to characteristic equations (2.51). The map $E(t_1,t_2)$ has a fixed point D and is centrally symmetric with respect to D. Consider $S_0 = R_3 \cap \{x : \omega_0(x) \geq 3\}$. This set is bounded by intervals lying on lines $\xi = \pm 0.03$ and parts of the ellipse where $\omega_0(x) = 3$. Split $S_0 = S_0^1 \cup S_0^2$, where $S_0^1 = S_0 \cap R_2$, and S_0^2 is the rest of S_0. Look at $E(0,1)S_0$.

By Lemma 2.27, this set is contained in $|\eta| < 0.1$. Denote $S_1 = E(0,1)S_0 \cap R_3$, and keep only simply connected component of this set containing the point D. S_1 will be bounded by intervals lying on the lines $\xi = \pm 0.03$ and parts of the level set $\omega(x,1) = 3$. By Lemma 2.27 the set $E(0,1)S_0^2$ gets transported out of R_3, and so $|S_1| \le |S_0| - |S_0^2|$. Notice that S_1 contains a part of the level set $\omega(x,1) = 4$. Moreover, since $E(0,1)S_0$ is contained in $|\eta| < 0.1$ and the ends of $\omega_0 = 4$ curve get transported out of R_3, the part of the level set $\omega = 4$ lying in S_1 contains a curve passing through the point D and connecting two points $P_1^{\pm}$ lying on $\xi = \pm 0.03$. Now let us split $S_1 = S_1^1 \cup S_1^2$, where $S_1^1 = S_1 \cap R_2$, and S_1^2 is the rest of S_1. We now iterate time in unit steps, obtaining a sequence of sets $S_{n+1} = E(n, n+1)S_n$. All properties of the set S_1 described above continue to hold for S_n. In particular, $|S_{n+1}| \le |S_n| - |S_n^2|$, which implies that $\sum_n |S_n^2| < \infty$. On the other hand, for each fixed ξ, $0.02 < |\xi| < 0.03$, a section of the set S_n^2 at level ξ must contain an interval $[\eta_1, \eta_2]$ such that $\omega(\eta_1, \xi) = 3$ and $\omega(\eta_2, \xi) = 4$. This implies that

$$|S_n^2| \ge 0.01 \|\nabla \omega(\cdot, n)\|_{L^\infty}^{-1}.$$

The application of Lemma 2.25 then gives

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{n=0}^{\infty} \|\nabla \omega(\cdot, n)\|_{L^\infty} = +\infty.$$

This is a discrete version of (2.47). One can obtain the continuous version by using the last statement of Lemma 2.27.

Exercise. Prove (2.47), and thus finish the proof of the Theorem. You will need to use the last statement of Lemma 2.27, taking τ small (and passing to the limit $\tau \to 0$). Otherwise the argument above will require only a few adjustments. □

2.3 The Bahouri-Chemin example

The scenario of the previous section gave us an example of superlinear growth in vorticity gradient. One could guess that, since the scenario is based on a hyperbolic point of a cellular flow, it may be possible to obtain exponential growth in a similar scenario with a better technical effort. This is what a computation in the Example 2 in the previous section would suggest. Of course, in the Denisov's construction, the flow is modified near the hyperbolic point so that it is no longer clear if exponential growth should persist. But in principle, the mechanism by which exponential growth of the derivatives can appear in solutions of 2D Euler equation is more or less clear. But how can one get double exponential growth? In this section we discuss an example of a singular stationary solution to 2D Euler equation which provides a hint into how double exponential growth can be achieved. This example is due to Bahouri and Chemin [4]. It also shows that many of the estimates we obtained when developing Yudovich theory of solutions with bounded vorticity are optimal. In this example, the fluid velocity u is just log-Lipschitz, and the flow map $\Phi_t(x)$ is indeed Hölder continuous with the exponent that is exponentially decaying in time.

Consider the singular "cross" solution corresponding to $\omega_0(x_1, x_2) = 1$ for $0 \le x_1, x_2 \le \pi$, $\omega_0(x_1, x_2) = -\omega_0(-x_1, x_2) = -\omega_0(x_1, -x_2)$ on the torus $(x_1, x_2) \in [-\pi, \pi)^2$. The solution corresponding to such ω can be rigorously defined, as it has been proved by Yudovich that any initial data in L^∞ leads to unique solution of 2D Euler equation. The first observation is an important conservation of symmetry.

Lemma 2.28 *The oddness with respect to an axis property is preserved by the 2D Euler evolution. That is, if $\omega_0 \in L^\infty$, $\omega_0(-x_1, x_2) = -\omega_0(x_1, x_2)$, then the corresponding Yudovich solution $\omega(x_1, x_2, t)$ satisfies $\omega(-x_1, x_2, t) = -\omega(x_1, x_2, t)$ for every $t \geq 0$.*

Remark. Clearly it is sufficient to prove conservation of odd symmetry with respect to $x_1 = 0$ axis, since the choice of coordinates does not affect the properties of the solutions.

Proof. The proof is similar to that of Lemma 2.23. Suppose that $(\omega(x_1, x_2, t), u(x_1, x_2, t), \Phi_t(x_1, x_2))$ is a Yudovich solution of the 2D Euler equation. A direct computation verifies that in this case $(-\omega(-x_1, x_2, t), (-u_1(-x_1, x_2, t), u_2(-x_1, x_2, t)), (-\Phi_t^1(-x_1, x_2), \Phi_t^2(-x_1, x_2))$ is also a Yudovich solution of the 2D Euler equation. But if ω_0 is odd with respect to x_1, the initial data for these two solutions coincide. Hence by uniqueness of solutions these two solutions must be equal. □

Now we can verify that the singular cross solution is stationary.

Lemma 2.29 *The solution of the 2D Euler equation corresponding to the initial data ω_0 described above is stationary, that is, $\omega(x, t) \equiv \omega_0(x)$ for all t.*

Proof. Since ω_0 is odd with respect to both x_1 and x_2, then by Lemma 2.28 the solution $\omega(x, t)$ has the same property. The stream function $\psi(x, t)$ is also odd with respect to both variables by a simple computation. Then $u_1 = \partial_2 \psi$ is odd with respect to x_1. Therefore, $u_1(0, x_2, t) = 0$ for all t. Similarly, $u_2(x_1, 0, t) = 0$ for all t. Moreover, by choosing a different system of coordinates and running the same argument, we see that $u_1(\pi, x_2, t) = u_2(x_1, \pi, t) = 0$. This shows that the particle trajectories never cross the lines $x_1 = 0, \pi$ and $x_2 = 0, \pi$. Since $\omega(x, t) = \omega_0(\Phi_t^{-1}(x))$ and given the structure of ω_0, this shows that $\omega(x, t) \equiv \omega_0$ for all times. □

For simplicity of notation and in order to work with positive rather than negative quantities, we will change the sign of the Biot-Savart law in this and next sections. This of course has no influence on the essence of the results and essentially is just a convention.

We now establish a key property of the fluid velocity in the Bahouri-Chemin example.

Proposition 2.30 *Consider the singular cross solution described above. Then for small positive x_1, we have*

$$u_1(x_1, 0) = \frac{4}{\pi} x_1 \log x_1 + O(x_1). \tag{2.52}$$

Proof. Let us use the Biot-Savart law (2.45) (changing the sign as we discussed):

$$u_1(x_1, 0) = -\frac{1}{2\pi} \lim_{\gamma \to 0} \int_{\mathbb{R}^2} \frac{y_2}{(x_1 - y_1)^2 + y_2^2} \omega(y) e^{-\gamma |y|^2} \, dy. \tag{2.53}$$

Denote $S = [-1, 1] \times [-1, 1]$. Let us represent $u_1(x_1, 0)$ as a sum of two components $u_1^S(x_1, 0) + u_1^F(x_1, 0)$. Here $u_1^S(x_1, 0)$ contains a contribution over integration over S in (2.53), while $u_1^F(x_1, 0)$ contains the contribution from integration over the complement of S (the "far field"). We first claim that $|u_1^F(x_1, 0)| \leq C \|\omega\|_{L^\infty} x_1$, and this will be left as an exercise.

Exercise. Verify that

$$\left| \lim_{\gamma \to 0} \int_{\mathbb{R}^2 \setminus S} \frac{y_2}{(x_1-y_1)^2+y_2^2} \omega(y) e^{-\gamma |y|^2} \, dy \right| \le C \|\omega\|_{L^\infty} x_1.$$

One way to perform this computation is to use the oddness of ω which leads to extra cancellation and faster decay in the kernel.

Next, in the $u_1^S(x_1,0)$ part, we can freely pass to the limit $\gamma \to 0$ and use symmetry to simplify the expression:

$$\pi u_1^S(x_1,0) = -\frac{1}{2} \int_S \frac{y_2}{(x_1-y_1)^2+y_2^2} \omega(y) \, dy = -\int_0^1 dy_2 \int_{-1}^1 \frac{y_2}{(x_1-y_1)^2+y_2^2} \omega(y_1,y_2) =$$
$$-2x_1 \int_0^1 \int_0^1 \frac{y_1 y_2}{((x_1-y_1)^2+y_2^2)((x_1+y_1)^2+y_2^2)} \, dy_1 dy_2. \tag{2.54}$$

In the last step we used that $\omega(y_1,y_2)=1$ on $[0,1]\times[0,1]$. Let us consider the contributions from different regions of integration in (2.54).

1. The region $[0,1]\times[0,2x_1]$.

$$\int_0^1 \int_0^{2x_1} \frac{y_1 y_2}{((x_1-y_1)^2+y_2^2)((x_1+y_1)^2+y_2^2)} \, dy_1 dy_2 \le$$
$$C \int_0^1 dz_1 \int_0^{x_1} dz_2 \frac{x_1 z_2}{(z_1^2+z_2^2)(x_1^2+z_2^2)} \le C \int_0^{x_1} \frac{x_1}{x_1^2+z_2^2} \arctan z_2^{-1} \, dz_2 \le C.$$

2. The region $[0,2x_1]\times[2x_1,1]$.

$$\int_0^{2x_1} dy_1 \int_{2x_1}^1 dy_2 \frac{y_1 y_2}{((x_1-y_1)^2+y_2^2)((x_1+y_1)^2+y_2^2)} \, dy_1 dy_2 \le$$
$$C \int_0^{x_1} dz_1 \int_{x_1}^1 dz_2 \frac{x_1 z_2}{(z_1^2+z_2^2)^2} \le C \int_0^{x_1} \frac{dz_1}{z_1^2+x_1^2} \le C.$$

3. The region $[2x_1,1]\times[2x_1,1]$. Here, the first observation is that

$$\left| \int_{2x_1}^1 \int_{2x_1}^1 \frac{y_1 y_2}{((x_1-y_1)^2+y_2^2)((x_1+y_1)^2+y_2^2)} \, dy_1 dy_2 - \int_{x_1}^1 \int_{x_1}^1 \frac{y_1 y_2}{(y_1^2+y_2^2)^2} dy_1 dy_2 \right| \le C.$$

Exercise. Verify this claim by direct computation.

Next, we have

$$\int_{x_1}^1 \int_{x_1}^1 \frac{y_1 y_2}{(y_1^2+y_2^2)^2} dy_1 dy_2 = \int_{x_1}^1 y_1 \left(\frac{1}{y_1^2+x_1^2} - \frac{1}{y_1^2+1} \right) dy_1 =$$
$$\int_{x_1^2}^1 \frac{dz_1}{z_1+x_1^2} + O(1) = -2\log x_1 + O(1).$$

Collecting all the estimates, we arrive at (2.52). $\square$

Observe that the estimate (2.52) corresponds to u_1 being just log-Lipchitz near the origin. Hence the estimates on fluid velocity in Yudovich theory are qualitatively sharp. A characteristic curve starting at a point $(x_1^0, 0)$ will be just a line $\Phi_t((x_1^0, 0)) \equiv (x_1(t), 0)$ moving towards the origin. If x_1^0 is sufficiently small, it will satisfy $x_1'(t) \leq x_1(t) \log x_1(t)$, and so $(\log x_1(t))' \leq \log x_1(t)$, $\log x_1(t) \leq e^t \log x_1^0$, $x_1(t) \leq x_1(0)^{\exp(t)}$. Such an estimate has several consequences. First, since the origin is a stationary point of the flow, the inverse flow map $\Phi_t^{-1}(x)$ can be Hölder continuous only with decaying in time exponent (at most e^{-t}). In fact, the exponent is a little weaker than that since our estimate on the characteristic convergence to zero is not sharp. Of course, the direct flow map $\Phi_t(x)$ also has a similar property; to establish it one needs to look at characteristic lines moving along the vertical separatrix.

Exercise. Verify the latter claim by direct calculations. You do not have to redo the proof of Proposition 2.30, you can use symmetry to conclude the analogous asymptotic behavior for $u_2(0, x_2)$, with a different sign.

This observation shows that the bounds on the flow map in Yudovich theory are also qualitatively sharp. Finally, let us observe what sort of gradient growth one can expect in a passive scalar advected by the fluid velocity u produced by singular cross. Suppose that as in (2.44), $\partial_t \varphi + (u \cdot \nabla \varphi) = 0$, $\varphi(x, 0) = \varphi_0(x)$. Choose $\varphi_0(x)$ to be a smooth function such that $\varphi_0(0) = 0$ and $\varphi_0(\delta) = 1$ for a small number $\delta > 0$ such that $u_1(x_1, 0) \leq x_1 \log x_1$ for all $0 \leq x_1 \leq \delta$. Then as we discussed in Example 1 in Section 2.2, $\varphi(\Phi_t((\delta, 0)), t) = \varphi_0(\delta) = 1$ and $\varphi(0, t) = 0$ since the origin is a stagnation point. On the other hand, due to the above estimates, we have $\Phi_t(\delta, 0) \leq \delta^{\exp(t)}$. By the mean value theorem, we have that

$$\|\nabla \varphi(\cdot, t)\|_{L^\infty} \geq \delta^{-\exp(t)},$$

thus resulting in double exponential growth in the gradient of passively advected scalar.

One may ask how such scenario may be relevant for 2D Euler with smooth initial data. One could try to smooth out the singular cross flow, and arrange for a small perturbation of it to play the role of a passive scalar on top of singular behavior, similar to the Nadirashvili's and Denisov's examples philosophy. If one could somehow arrange for the solution to approach, in some sense, the singular cross solution of the background flow, then one could provide an example of double exponential in time growth. A similar idea was exploited by Denisov in [47] to design a finite time double exponential growth example. However, one would face serious difficulties to extend this approach to infinite time. First, to keep the background scenario stable, one needs symmetry - and odd symmetry bans nonzero perturbation right where the velocity is most capable of producing double exponential growth for all times, on the $x_2 = 0$ separatrix. Second, it is not clear how to make a smooth solution approach the "cross" in some suitable sense. Third, the perturbation will not be passive, and, for large times, will be difficult to decouple from the equation. In Denisov example, nonlinearity is something to fight; the growth of the vorticity gradient is driven by linear mechanism. To build example with double exponential growth, the nonlinearity would have to become our friend. We will consider such example in the next section. The growth of vorticity gradient in that example will be double exponential, and it will happen at the boundary of the domain. The latter is crucial for the construction. Essentially, the boundary will play a role of a separatrix in Bahouri-Chemin flow, but if voriticity has to vanish on a separatrix if we want to keep the symmetry, on the boundary it does not have to be zero.

2.4 The Kiselev-Sverak example

In this section, we will prove the following theorem [80].

Theorem 2.31 *Consider two-dimensional Euler equation on a unit disk D. There exists a smooth initial data ω_0 with $\|\nabla\omega_0\|_{L^\infty}/\|\omega_0\|_{L^\infty} > 1$ such that the corresponding solution $\omega(x,t)$ satisfies*

$$\frac{\|\nabla\omega(x,t)\|_{L^\infty}}{\|\omega_0\|_{L^\infty}} \geq \left(\frac{\|\nabla\omega_0\|_{L^\infty}}{\|\omega_0\|_{L^\infty}}\right)^{c\exp(c\|\omega_0\|_{L^\infty}t)} \tag{2.55}$$

for some $c > 0$ and for all $t \geq 0$.

As the first step towards the proof of Theorem 2.31, let us start setting up the scenario we will be considering. From now on, let D be a closed unit disk in the plane. It will be convenient for us to take the system of coordinates centered at the lowest point of the disk, so that the center of the disk is at $(0,1)$. Our initial data $\omega_0(x)$ will be odd with respect to the vertical axis: $\omega_0(x_1,x_2) = -\omega_0(-x_1,x_2)$. We checked the conservation of such symmetry in the periodic initial data case; it can be checked similarly for the case of a domain with vertical symmetry axis.

We will take smooth initial data $\omega_0(x)$ so that $\omega_0(x) \geq 0$ for $x_1 > 0$ (and so $\omega_0(x) \leq 0$ for $x_1 < 0$). This configuration makes the origin a hyperbolic fixed point of the flow; in particular, u_1 vanishes on the vertical axis. It will be clear from analysis of the Biot-Savart law. The Dirichlet Green's function for the disk is given explicitly by $G_D(x,y) = \frac{1}{2\pi}(\log|x-y| - \log|x-\bar{y}| - \log|y-e_2|)$, where with our choice of coordinates $\bar{y} = e_2 + (y-e_2)/|y-e_2|^2$, $e_2 = (0,1)$ (see e.g. [52]). Given the symmetry of ω, we have

$$u(x,t) = \nabla^\perp \int_D G_D(x,y)\omega(y,t)\,dy = \frac{1}{2\pi}\nabla^\perp \int_{D^+} \log\left(\frac{|x-y||\widetilde{x}-\bar{y}|}{|x-\bar{y}||\widetilde{x}-y|}\right)\omega(y,t)\,dy, \tag{2.56}$$

where D^+ is the half disk where $x_1 \geq 0$, and $\widetilde{x} = (-x_1,x_2)$. The following Lemma will be crucial for the proof of Theorem 2.31. Let us introduce notation $Q(x_1,x_2)$ for a region that is the intersection of D^+ and the quadrant $x_1 \leq y_1 < \infty$, $x_2 \leq y_2 < \infty$.

Lemma 2.32 *Take any γ, $\pi/2 > \gamma > 0$. Denote D_1^γ the intersection of D^+ with a sector $\pi/2 - \gamma \geq \phi \geq 0$, where ϕ is the usual angular variable. Then there exists $\delta > 0$ such that for all $x \in D_1^\gamma$ such that $|x| \leq \delta$ we have*

$$u_1(x_1,x_2,t) = -\frac{4}{\pi}x_1 \int_{Q(x_1,x_2)} \frac{y_1y_2}{|y|^4}\omega(y,t)\,dy_1dy_2 + x_1B_1(x_1,x_2,t), \tag{2.57}$$

where $|B_1(x_1,x_2,t)| \leq C(\gamma)\|\omega_0\|_{L^\infty}$.

Similarly, if we denote D_2^γ the intersection of D^+ with a sector $\pi/2 \geq \phi \geq \gamma$, then for all $x \in D_2^\gamma$ such that $|x| \leq \delta$ we have

$$u_2(x_1,x_2,t) = \frac{4}{\pi}x_2 \int_{Q(x_1,x_2)} \frac{y_1y_2}{|y|^4}\omega(y,t)\,dy_1dy_2 + x_2B_2(x_1,x_2,t), \tag{2.58}$$

where $|B_2(x_1,x_2,t)| \leq C(\gamma)\|\omega_0\|_{L^\infty}$.

Exercise The Lemma holds more generally than in the disk; perhaps the simplest proof is for the case where D is a square. The computation for that case is quite similar to Bahouri-Chemin example. Carry out the proof of the Lemma for this case.

Exercise The exclusion of a small sector does not appear to be a technical artifact. The vorticity can be arranged (momentarily) in a way that the hyperbolic picture provided by the Lemma is violated outside of D_1^γ, for example the direction of u_1 may be reversed near the vertical axis. Verify this for the case of a square.

Let us denote

$$\Omega(x_1, x_2, t) = \frac{4}{\pi} x_2 \int_{Q(x_1,x_2)} \frac{y_1 y_2}{|y|^4} \omega(y,t)\, dy_1 dy_2. \tag{2.59}$$

This term appears both in (2.57) and (2.58), and, as will become clear soon, can be thought of as the main term in these estimates in certain regime. Indeed, while the remainder in (2.57), (2.58) satisfies Lipschitz estimates, the nonlocal term $\Omega(x_1, x_2, t)$ can grow as a logarithm if the support of the vorticity approaches the origin. This growth through nonlinear feedback can lead to double exponential growth in the gradient of solution. Essentially, Lemma 2.32 makes it possible to ensure in certain regimes that the flow near the origin is hyperbolic, with fluid trajectories just hyperbolas in the main term. The speed of motion along trajectories is controlled by the nonlocal factor in (2.57), (2.58), and this factor is the same for both u_1 and u_2.

We also note a certain comparison property, monotonicity imbedded in the form of $\Omega(x_1, x_2, t)$. The size of the expression in (2.59) tends to increase as x approaches origin since the region of integration grows. This property will be important in the construction of our example.

Proof. Let us prove (2.57), the proof of (2.58) is similar. Fix a small $\gamma > 0$. Fix a point $x = (x_1, x_2) \in D_1^\gamma$, $|x| \leq \delta$. By the definition of D_1^γ, we have $x_2 \leq x_1 \cot\gamma$. Define $r = 10(1 + \cot\gamma)x_1$, then $x \in B_r(0)$. Let us assume that δ is small enough so that $r < 0.1$ whenever $|x| \leq \delta$.

Note that the contribution to u_1 from integration over $B_r(0)$ in the Biot-Savart law (2.56) does not exceed

$$C\|\omega_0\|_{L^\infty} \int_{D^+ \cap B_r(0)} \frac{1}{|x-y|}\, dy \leq C(\gamma)\|\omega\|_{L^\infty} x_1.$$

For $y \in D^+ \setminus B_r(0)$, we have $|y| \geq 10|x|$. Let us rewrite the four logarithms in (2.56) as follows:

$$\begin{aligned} &\log\left(1 - \frac{2xy}{|y|^2} + \frac{|x|^2}{|y|^2}\right) - \log\left(1 - \frac{2x\bar{y}}{|\bar{y}|^2} + \frac{|x|^2}{|\bar{y}|^2}\right) - \\ &\quad \log\left(1 - \frac{2\tilde{x}y}{|y|^2} + \frac{|x|^2}{|y|^2}\right) + \log\left(1 - \frac{2\tilde{x}\bar{y}}{|\bar{y}|^2} + \frac{|x|^2}{|\bar{y}|^2}\right). \end{aligned} \tag{2.60}$$

For small t,

$$\log(1+t) = t - \frac{t^2}{2} + O(t^3).$$

Therefore, after a direct computation the expression (2.60) leads to

$$\pi G_D(x,y) = -\frac{x_1 y_1}{|y|^2} + \frac{x_1 \bar{y}_1}{|\bar{y}|^2} - \frac{2x_1 x_2 y_1 y_2}{|y|^4} + \frac{2x_1 x_2 \bar{y}_1 \bar{y}_2}{|\bar{y}|^4} + O\left(\frac{|x|^3}{|y|^3}\right).$$

In the last term, we used that $|\bar{y}| \geq |y|$ for $y \in D^+$, something one can check by computation. A direct calculation shows that

$$\frac{\bar{y}_1}{|\bar{y}|^2} = \frac{y_1}{|y|^2}, \quad \frac{\bar{y}_2}{|\bar{y}|^2} = 1 - \frac{y_2}{|y|^2}.$$

Therefore we obtain

$$\pi G_D(x,y) = -\frac{4x_1x_2y_1y_2}{|y|^4} + \frac{2x_1x_2y_1}{|y|^2} + O\left(\frac{|x|^3}{|y|^3}\right). \tag{2.61}$$

It is not difficult to verify that the expression (2.61) can be differentiated with respect to x_2, yielding

$$\pi\frac{\partial G_D(x,y)}{\partial x_2} = -\frac{4x_1y_1y_2}{|y|^4} + \frac{2x_1y_1}{|y|^2} + O\left(\frac{|x|^2}{|y|^3}\right). \tag{2.62}$$

Now

$$\int_{D^+\setminus B_r} \frac{|x|^2}{|y|^3}\,dy \leq C|x|^2 \int_r^1 \frac{1}{s^2}\,ds \leq Cr^{-1}|x|^2 \leq C(\gamma)x_1.$$

Also,

$$\int_{D^+\setminus B_r} \frac{y_1}{|y|^2}\,dy \leq C\int_r^1 ds \leq C.$$

Therefore, the last two terms in (2.62) give regular contributions to u_1. It remains to reconcile the regions of the integration in the main term, namely to show that

$$\int_{D^+\setminus B_r} \frac{y_1y_2}{|y|^4}\omega(y)\,dy = O(1) + \int_{Q(x_1,x_2)} \frac{y_1y_2}{|y|^4}\omega(y)\,dy.$$

Indeed,

$$\int_{B_r\cap Q(x_1,x_2)} \frac{y_1y_2}{|y|^4}\,dy \leq \int_{x_1}^{Cx_1} dy_1 \int_0^{Cx_1} dy_2 \frac{y_1y_2}{|y|^4} \leq$$

$$C\int_{x_1}^{Cx_1} y_1 \int_0^{C^2x_1^2} \frac{1}{(y+y_1^2)^2}\,dy dy_1 = C\int_{x_1}^{Cx_1} dy_1 y_1^{-1} \leq C.$$

Finally, the set $D^+ \setminus (Q(x_1,x_2) \cup B_r)$ consists of two strips along x_1 and x_2 axis. The contribution of the integral over the strip along the x_1 axis does not exceed

$$\int_0^{x_1} dy_1 \int_{x_1}^1 dy_2 \frac{y_1y_2}{|y|^4} \leq \int_0^{x_1} \frac{y_1}{y_1^2+x_1^2}\,dy_1 \leq C$$

since $r >> x_1$. The integral over the strip along the x_2 axis does not exceed

$$\int_0^{x_2} dy_2 \int_{x_1}^1 dy_1 \frac{y_1y_2}{|y|^4}.$$

Since $x_2 \leq C(\gamma)x_1$, the latter integral can also be bounded by a constant via similar computation. This completes the proof of the lemma.

$\square$

Exercise. Fill in all the computations in the proof of the lemma.

Before proving Theorem 2.31, we make a simpler observation: with the aid of Lemma 2.32 it is fairly straightforward to find examples with exponential in time growth of vorticity gradient. Indeed, take smooth initial data $\omega_0(x)$ which is equal to one everywhere in D^+ except on a thin strip of width equal to δ near the vertical axis $x_1 = 0$, where $0 < \omega_0(x) < 1$ (and ω_0 vanishes on the vertical axis as it must by our symmetry assumptions). Observe that due to incompressibility, the distribution function of $\omega(x,t)$ is the same for all times. In particular, the measure of the complement of the set where $\omega(x,t) = 1$ does not exceed 2δ. In this case for every $|x| < \delta$, $x \in D^+$, we can derive the following estimate for the integral appearing in the representation (2.57):

$$\int_{Q(x_1,x_2)} \frac{y_1 y_2}{|y|^4}\omega(y,t)\,dy_1 dy_2 \geq \int_{2\delta}^{1}\int_{\pi/6}^{\pi/3} \omega(r,\phi)\frac{\sin 2\phi}{2r}\,d\phi dr \geq \frac{\sqrt{3}}{4}\int_{2\delta}^{1}\int_{\pi/6}^{\pi/3}\frac{\omega(r,\phi)}{r}\,d\phi dr.$$

The value of the integral on the right hand side is minimal when the area where $\omega(r,\phi)$ is less than one is concentrated at small values of the radial variable. Using that this area does not exceed 2δ, we obtain

$$\frac{4}{\pi}\int_{Q(x_1,x_2)} \frac{y_1 y_2}{|y|^4}\omega(y,t)\,dy_1 dy_2 \geq c_1 \int_{c_2\sqrt{\delta}}^{1}\int_{\pi/6}^{\pi/3}\frac{1}{r}\,d\phi dr \geq C_1 \log \delta^{-1}, \tag{2.63}$$

where c_1, c_2 and C_1 are positive universal constants.

Putting the estimate (2.63) into (2.57), we get that for all for $|x| \leq \delta$, $x \in D^+$ that lie on the disk boundary, we have

$$u_1(x,t) \leq -x_1(C_1 \log \delta^{-1} - C_2),$$

where $C_{1,2}$ are universal constants. We can choose $\delta > 0$ sufficiently small so that $u_1(x,t) \leq -x_1$ for all times if $|x| < \delta$. Due to the boundary condition on u, the trajectories which start at the boundary stay on the boundary for all times. Taking such a trajectory starting at a point $x_0 \in \partial D$ with $x_{0,1} \leq \delta$, we get $\Phi^1_{t,1}(x_0) \leq x_{0,1}e^{-t}$ for this characteristic curve. Since $\omega(x,t) = \omega(\Phi_t^{-1}(x))$, we see that $\|\nabla\omega(x,t)\|_{L^\infty}$ grows exponentially in time if we pick ω_0 which does not vanish identically at the boundary near the origin (for example, if $\omega_0(\delta, 1-\sqrt{1-\delta^2}) \neq 0$).

To construct examples with double exponential growth, we have to work a little harder. For the sake of simplicity, we will build our example with ω_0 such that $\|\omega_0\|_{L^\infty} = 1$.

Proof. [Proof of Theorem 2.31] We first fix some small $\gamma > 0$. We will take the smooth initial data like in the end of the previous paragraph, with $\omega_0(x) = 1$ for $x \in D^+$ apart from a narrow strip of width at most $\delta > 0$ (with δ small enough so that (2.57), (2.58) apply) near the vertical axis where $0 \leq \omega_0(x) \leq 1$. Then (2.63) holds. We will also choose δ so that $C_1 \log \delta^{-1} > 100C(\gamma)$ where $C(\gamma)$ is the constant in the bound for the error terms B_1, B_2 appearing in (2.57), (2.58).

For $0 < x_1' < x_1'' < 1$ we denote

$$\mathcal{O}(x_1', x_1'') = \left\{(x_1,x_2) \in D^+,\ x_1' < x_1 < x_1'',\ x_2 < x_1\right\}. \tag{2.64}$$

For $0 < x_1 < 1$ we let

$$\underline{u}_1(x_1,t) \quad = \quad \min_{(x_1,x_2)\in D^+\,,\,x_2<x_1} u_1(x_1,x_2,t) \tag{2.65}$$

and

$$\overline{u}_1(x_1,t) \quad = \quad \max_{(x_1,x_2)\in D^+\,,\,x_2<x_1} u_1(x_1,x_2,t)\,. \tag{2.66}$$

It is easy to see that these functions are locally Lipschitz in x_1 on $[0,1)$, with the Lipschitz constants being locally bounded in time. Hence we can define $a(t)$ by

$$\dot{a} = \overline{u}_1(a,t)\,, \quad a(0) = \varepsilon^{10} \tag{2.67}$$

and $b(t)$ by

$$\dot{b} = \underline{u}_1(b,t)\,, \quad b(0) = \varepsilon\,, \tag{2.68}$$

where $0 < \varepsilon < \delta$ is sufficiently small, its exact value to be determined later. Let

$$\mathcal{O}_t = \mathcal{O}(a(t),b(t))\,. \tag{2.69}$$

At this stage we have not yet ruled out that $\mathcal{O}_t$ perhaps might become empty for some $t > 0$. However, it is clear from the definitions that $\mathcal{O}_t$ will be non-empty at least on some non-trivial interval of time. Our estimates below show that in fact $\mathcal{O}_t$ will be non-empty for all $t > 0$.

We will choose ω_0 so that $\omega_0 = 1$ on $\mathcal{O}_0$ with smooth sharp (on a scale $\lesssim \varepsilon^{10}$) cutoff to zero into D^+. This leaves some ambiguity in the definition of $\omega_0(x)$ away from $\mathcal{O}_0$. We will see that it does not really matter how we define ω_0 there, as long as we satisfy the conditions above. For simplicity, one can think of $\omega_0(x)$ being just zero for $|x| < \delta$ away from a small neighborhood of $\mathcal{O}_0$. Using the estimates (2.57), (2.58), the estimate (2.63) and our choice of δ ensuring that $C_1 \log \delta^{-1} >> C(\gamma)$ we see that both a and b are decreasing functions of time and that near the diagonal $x_1 = x_2$ in $\{|x| < \delta\}$ we have

$$\frac{x_1(\log\delta^{-1} - C)}{x_2(\log\delta^{-1} + C)} \le \frac{-u_1(x_1,x_2)}{u_2(x_1,x_2)} \le \frac{x_1(\log\delta^{-1} + C)}{x_2(\log\delta^{-1} - C)}. \tag{2.70}$$

This means that all particle trajectories for all times are directed into the $\phi > \pi/4$ region on the diagonal. We claim that $\omega(x,t) = 1$ on $\mathcal{O}_t$. Indeed, it is clear that the "fluid particles" which at $t = 0$ are in $D^+ \setminus \overline{\mathcal{O}}_0$ cannot enter $\mathcal{O}_{t'}$ through the diagonal $\{x_1 = x_2\}$ due to (2.70) at any time $0 \le t' \le t$. Due to the very definition of $a(t), b(t)$ and $\mathcal{O}_t$, they cannot enter $\mathcal{O}_{t'}$ through the vertical segments $\{(a(t'),x_2) \in D^+\,,\ x_2 < a(t')\}$ or $\{(b(t'),x_2) \in D^+\,,\ x_2 < b(t')\}$ at any time $0 \le t' \le t$ either. Finally, they obviously cannot enter through the boundary points of D. Hence the "fluid particles" in $\mathcal{O}_t$ must have been in $\mathcal{O}_0$ at the initial time and we conclude that $\omega(\,\cdot\,,t) = 1$ in $\mathcal{O}_t$.

By Lemma 2.32, we have

$$\underline{u}_1(b(t),t) \ge -b(t)\,\Omega(b(t),x_2(t)) - C\,b(t),$$

for some $x_2(t) \le b(t)\,,\ (x_2(t),b(t)) \in D^+$ as $\|\omega(x,t)\|_{L^\infty} \le 1$ by our choice of the initial datum ω_0. A simple calculation shows that

$$\Omega(b(t),x_2(t)) \le \Omega(b(t),b(t)) + C.$$

Indeed, since $x_2(t) \le b(t)$ we can write

$$\int_b^2 \int_0^b \frac{y_1 y_2}{|y|^4}\, dy_2 dy_1 = \frac{1}{2}\int_b^2 y_1 \left(\frac{1}{y_1^2} - \frac{1}{y_1^2+b^2}\right) dy_1 \le b^2 \int_b^2 y_1^{-3} dy_1 \le C. \tag{2.71}$$

Thus we get

$$\underline{u}_1(b(t),t) \ge -b(t)\,\Omega(b(t),b(t)) - 2C\,b(t). \tag{2.72}$$

At the same time, for suitable $\widetilde{x}_2(t)$ with $\widetilde{x}_2(t) \le a(t)$, $(a(t),\widetilde{x}_2(t)) \in \bar{D}$ we have

$$\overline{u}_1(a(t),t) \le -a(t)\,\Omega(a(t),\widetilde{x}_2(t)) + \widetilde{C}a(t) \le -a(t)\,\Omega(a(t),0) + Ca(t),$$

by an estimate similar to (2.71) above. Observe that

$$\Omega(a(t),0) \ge \frac{4}{\pi}\int_{\mathcal{O}_t} \frac{y_1 y_2}{|y|^4}\omega(y,t)\, dy_1 dy_2 + \Omega(b(t),b(t)).$$

Since $\omega(y,t) = 1$ on $\mathcal{O}_t$,

$$\int_{\mathcal{O}_t} \frac{y_1 y_2}{|y|^4}\omega(y,t)\, dy_1 dy_2 \ge \int_\varepsilon^{\pi/4} \int_{a(t)/\cos\phi}^{b(t)/\cos\phi} \frac{\sin 2\phi}{2r}\, dr d\phi > \frac{1}{8}(-\log a(t) + \log b(t)) - C.$$

Therefore

$$\overline{u}_1(a(t),t) \le -\,a(t)\left(\frac{1}{2\pi}(-\log a(t) + \log b(t)) + \Omega(b(t),b(t))\right) + 2Ca(t). \tag{2.73}$$

Note that from estimates (2.72), (2.73) it follows that $a(t)$ and $b(t)$ are monotone decreasing in time, and by finiteness of $\|u\|_{L^\infty}$ these functions are Lipschitz in t. Hence we have sufficient regularity for the following calculations.

$$\frac{d}{dt}\log b(t) \ge -\Omega(b(t),b(t)) - 2C\,, \tag{2.74}$$

$$\frac{d}{dt}\log a(t) \le \frac{1}{2\pi}\left(\log a(t) - \log b(t)\right) - \Omega(b(t),b(t)) + 2C. \tag{2.75}$$

Subtracting (2.74) from (2.75), we obtain

$$\frac{d}{dt}\left(\log a(t) - \log b(t)\right) \le \frac{1}{2\pi}\left(\log a(t) - \log b(t)\right) + 4C. \tag{2.76}$$

From (2.76), the Gronwall lemma leads to

$$\log a(t) - \log b(t) \le \log\left(a(0)/b(0)\right)\exp(t/2\pi) + 4C\exp(t/2\pi) \le (9\log\varepsilon + 4C)\exp(t/2\pi). \tag{2.77}$$

We should choose our ε so that $-\log\varepsilon$ is larger than the constant $4C$ that appears in (2.77). In this case, we obtain from (2.77) that $\log a(t) \le 8\exp(t/2\pi)\log\varepsilon$, and so $a(t) \le \varepsilon^{8\exp(t/2\pi)}$.

Note that by the definition of $a(t)$, the first coordinate of the characteristic that originates at the point on ∂D near the origin with $x_1 = \varepsilon^{10}$ does not exceed $a(t)$. To arrive at (2.55), it remains to note that we can arrange $\|\nabla\omega_0\|_{L^\infty} \lesssim \varepsilon^{-10}$. $\square$

Let us make a few remarks regarding the construction we just completed. It is clear from the proof that the double exponential growth in our example is fairly robust. In particular, it will be present for any initial data in a sufficiently small L^∞ ball around ω_0 provided that the odd symmetry is conserved. Essentially, what we need for the construction to work is symmetry and the dominance of Ω terms in Lemma 2.32. The axial symmetry does appear crucial for the construction. It is an interesting question whether one can relax the 'symmetry requirement and get examples of double exponential growth in arbitrary smooth domain.

Another natural question is whether double exponential growth can happen in the bulk of the fluid. As we discussed, the doubly odd symmetry is very useful for controlling the solution, but also impedes growth where the fluid velocity would be most effective in creating it - on the separatrices. Recently, Zlatos [122] has proved that exponential growth of the second order derivatives of vorticity is possible starting from smooth initial data. The argument uses an upgrade of Lemma 2.32, and the geometry is similar to Bahouri-Chemin example. Yet exponential growth is still a linear phenomenon at heart, so the prospects for upgrading the result to double exponential in this framework are not clear. There could be other scenarios for double exponential growth but overall the question remains wide open.

It is also very interesting to understand how the solution looks in the long time limit in our scenario. One general philosophy of how to prove a finite time singularity in a PDE is as follows. Find a singular solution that is in some sense stable - namely, it should be such that there exists a smooth trajectory which is attracted to it in finite time. One can wonder if our scenario comes from a similar phenomenon. Of course, we know that solutions to the 2D Euler equation do not blow up in finite time. But it is not unreasonable to hypothesise that in the large time limit small areas where vorticity is less than 1 in absolute value get completely homogenized by the flow. Here we make contact with a broad and exciting field in the study of 2D Euler equation that is beyond the scope of this text - the study of long time dynamics of Euler solutions. Very little is known rigorously here, but ideas of statistical mechanics have been used to make conjectures and predictions that seem to be corroborated to some extent by numerical simulations. We refer the interested reader to the accessible note by Wayne in the Notices of the AMS [114] where more references can be found. In these theories, one relies on conserved quantities and ergodicity-type (mixing) assumptions (rather than the actual dynamics). Such assumptions are notoriously difficult to verify. In our situation this approach (conjecturally) predicts that as $t \to \infty$, the vorticity field $\omega(x,t)$ should weakly* approach a steady-state solution, which - under our symmetry assumptions - can be expected to have a discontinuity along the axis of symmetry $\{x_1 = 0\}$. Hence one can view the growth in the gradient of vorticity that we proved as a manifestation of the approach to such singular solution. Of course, proving that in our example there exists such singular steady limit is quite hard and appears to be beyond reach of current methods.

The double exponential growth example we discussed has been inspired by numerical simulations of Tom Hou and Guo Luo [73] who propose a new scenario for the development of a finite time singularity in solutions to the 3D Euler equation at a boundary. The geometry of the Kiselev-Sverak example bears resemblance to the geometry of the scenario of Hou and Luo. The problem of the finite time singularity of 3D Euler equation is one of the major open problems in fluid mechanics and PDE, and we refer the interested reader to [73] for more information and to [30] for recent analytical work explaining the connection with 3D case in more detail.

Chapter 7

More general active scalars: the Surface Quasi-Geostrophic equation and the Burgers equation

1 Introduction

The 2D Euler equations, the subject of the previous chapter are an example of a broader class of equations called "active scalars". A function $\theta(t,x)$ is called a (dissipative) active scalar if it satisfies an equation of the form

$$\theta_t + u \cdot \nabla\theta = -(-\Delta)^\alpha \theta, \;\; \theta(0,x) = \theta_0(x). \tag{1.1}$$

Here, the parameter $\alpha \geq 0$ measures the dissipation strength, and is typically taken in the range $0 \leq \alpha \leq 1$, though the case $\alpha > 1$ can also be considred. When we talk about $\alpha = 0$, we mean that the dissipative term is simply missing from the equation, and we are dealing with the inviscid case. In this chapter, we will consider equations set on the whole space $\mathbb{R}^d$ or on the torus $\mathbb{T}^d$, so that the issue of the boundaries and defining the fractional Laplacian in a bounded domain does not come into play. To define the fractional Laplacian on the torus or in the whole space, one can go to the Fourier side, where it becomes a multiplication operator. For example, in $\mathbb{R}^n$ if $f = (-\Delta)^\alpha g$, then the Fourier transforms of f and g are related by

$$\hat{f}(\xi) = |2\pi\xi|^{2\alpha}\hat{g}(\xi).$$

We will also give below an explicit formula for the fractional Laplacian in the physical space. The vector field u in (1.1) is determined by θ, hence the name "active scalar": θ itself defines how it is advected. The 2D Euler equations, written in vorticity form:

$$\omega_t + u \cdot \nabla\omega = 0, \tag{1.2}$$

are an example of inviscid active scalar – the velocity is related to the vorticity via the Biot-Savart law:

$$u = \nabla^\perp(-\Delta)^{-1}\omega, \tag{1.3}$$

where $\nabla^\perp = (\partial_2, -\partial_1)$. There are many more active scalar equations. These equations are nonlinear and the relation between the active scalar and the advecting velocity in most of them is nonlocal, as in the Biot-Savart law (1.3) for the Euler equations (1.2). Their solutions are often prone to small scale creation, just as in the 2D Euler equations. For many active scalars, the questions of finite time singularity vs. global in time regularity remain open. The scalar itself obeys the maximum principle: the maximum of $\theta(t,x)$ in space can not grow in time, as long as the advecting velocity $u(t,x)$ remains smooth. Therefore, the loss of regularity is related not to the growth of θ but to the growth of $\nabla\theta$, or other loss of spatial regularity of θ. Our main interest in this chapter is to illustrate some of techniques and results available for the study of these fundamental questions. We will focus on two main examples: the Burgers equation and the surface quasi-geostrophic equation.

The Burgers equation

This is the simplest active scalar, which has been first discussed in [10] and [64], and studied more extensively by Burgers [26] in 1948 as a model to study turbulence. The Burgers' equation is

$$\partial_t\theta + \theta\theta_x + (-\partial_x^2)^\alpha\theta = 0, \tag{1.4}$$

in one space dimension. The advecting velocity here is simply $u = \theta$. The original work of Burgers, as well as the vast majority of the consequent work, focused on $\alpha = 1$, where the Burgers equation has some special properties and, in particular, can be linearized by the Cole-Hopf transformation (see e.g. [52]). But more general values of α have also been considered. This equation is the simplest model of an interaction between the dissipative term and a fluid-type nonlinearity. It is local, as opposed to most other active scalars. Not surprisingly, as we will see, much is known for this equation due to its local nature. It is well known that when $\alpha = 0$, the solutions of the Burgers equation form shocks, which are jump discontinuity singularities, in a finite time, while no shocks can be formed and solutions remain smooth in the "diffusive case" $\alpha = 1$. In Section 1.4 below, we will discuss which value of α will be critical for the transition between the possibility of a finite time singularity and global in time regularity.

The 2D Surface Quasi-Geostrophic (SQG) equation

The outward appearance of the non-dissipative 2D Surface Quasi-Geostrophic equation is identical to that of the 2D Euler equation in the vorticity formulation (1.2):

$$\theta_t + u\cdot\nabla\theta = 0,$$

but the vector field u is one derivative less regular relative to the 2D Euler case and is given by

$$u = \nabla^\perp(-\Delta)^{-1/2}\theta, \tag{1.5}$$

or, equivalently,

$$u(x) = cP.V.\int_{\mathbb{R}^2}\theta(x-y)\left(\frac{y_2}{|y|^3}, -\frac{y_1}{|y|^3}\right)\,dy, \tag{1.6}$$

with an appropriately chosen constant c. We will explain why (1.5) and (1.6) are equivalent below. The higher singularity in the kernel of this analog of the Biot-Savart law has significant consequences for the properties of the solutions of the SQG equation. The surface quasi-geostrophic equation appears in atmospheric science, and we will sketch the derivation of this model in Section 2.

In the mathematical literature, the SQG equation was first considered by Constantin, Majda and Tabak in [33] (in the non-dissipative case $\alpha = 0$). A scenario for a finite time singularity, a closing saddle, was proposed and numerically investigated there. A close connection between the SQG equation and the 3D Euler equations was also pointed out – we will discuss this point below in more detail. It was later proved by D. Cordoba [39] that blow up does not happen in the scenario proposed by Constantin, Majda and Tabak [33]. In Section 3, we will discuss a later argument by D. Cordoba and C. Fefferman [40], which shows that the solutions to the SQG equation cannot form sharp front singularities. The general question of the finite time singularity vs global regularity remains open for the inviscid SQG equation. We will close our discussion of the SQG equation by proving the global regularity for the critical SQG equation, where the strength of dissipation is $\alpha = 1/2$. We will explain the origin of this terminology below, but for now we will simply say that, as we will see in this chapter, the proof of regularity is much more standard for $\alpha > 1/2$. This is close to the best currently available regularity result for the SQG equation.

For all active scalar equations we consider here, proving local existence and uniqueness of solutions in a sufficiently regular Sobolev space is not a problem. Moreover, if the dissipation is present, then the local in time solution is smooth for short times even if the initial condition is not C^∞, while in the inviscid case, solution is only as smooth as the initial data. We will not discuss this standard part of the argument which, while somewhat technical, is also well developed. We refer to the well known textbooks such as [87], [113] or [36] for the proofs of the local existence for related (and more complex) fluid mechanics equations. The main goal of this chapter is to outline and explain some of the themes and terminology in the field of nonlinear and, in particular, fluids PDEs that are most often used in the modern research – such as subcritical, critical, supercritical, and to show them in action. We also cover many of the most actively used methods – functional analytic estimates, a version of the comparison principle, and the Lyapunov functional-style blow up argument.

2 The derivation of the SQG equation

The SQG equation has been a focus of intense research in recent years, perhaps because it is the simplest looking, physically motivated model of fluid mechanics for which the basic question of finite time singularity vs global regularity remains open. In this section, we will sketch its derivation. We will mostly follow the exposition in the book of Majda [88], which contains most of the relevant arguments but stops one step short of explicitly writing the SQG equation. This last step is explained, for instance, in [72].

The rotating Boussinesq equations

The starting point for the derivation are the rotating Boussinesq equations

$$\rho(u_t + u \cdot \nabla u) + f e_3 \times u + \nabla p = -\rho g e_3 \tag{2.1}$$

$$\rho_t + \nabla \cdot (\rho u) = 0, \tag{2.2}$$

for the fluid density $\rho(t,x)$ and velocity $u(t,x)$, as well as the pressure $p(t,x)$. The term $fe_3 \times u$ comes from the Coriolis force due to the planetary rotation, with f the frequency of rotation, and g is the acceleration due to gravity in the direction of e_3. These equations neglect further important effects that are included into more sophisticated models, such as, for example, humidity, rain, and cloud formation. Yet, the Boussinesq equations are already a reasonable model. At the same time, one can appreciate their complexity as well as the usefulness of simpler models which can be used to gain analytical insight.

We will now obtain a simplification of the full rotating Boussinesq equations that comes from the observation that there is a significant difference between the horizontal and the vertical motion in the atmosphere and ocean: the horizontal motion is typically much more pronounced. We assume that the background fluid flow is quiescent, $\bar{u} = 0$, and the density has a background stratified profile

$$\overline{\rho} = \rho_b - b x_3,$$

with $b > 0$ to ensure that the stratification is stable. The positivity of b corresponds to the less dense material at higher altitudes. The constant ρ_b describes the dominant value of density at an altitude of interest to us, and the deviations from the background profile are assumed to be relatively small in the Boussinesq approximation framework. The background pressure profile is determined by the background density as the leading order term in (2.1):

$$\frac{\partial \overline{p}}{\partial x_3} = -\frac{g}{\rho_b}\overline{\rho}(x_3), \tag{2.3}$$

where g is the acceleration due to gravity. Note that both background profiles depend only on the vertical coordinate x_3.

Let us denote the horizontal spatial coordinate and the horizontal component of the velocity by $x_H = (x_1, x_2)$, and $u_H = (u_1, u_2)$. We will also use the notation $\nabla^H = (\partial_{x_1}, \partial_{x_2})$ for the gradient in the horizontal direction, set $u_H^\perp = (-u_2, u_1, 0)$, and

$$\frac{D}{Dt} = \partial_t + u \cdot \nabla, \quad \frac{D^H}{Dt} = \partial_t + u_H \cdot \nabla^H.$$

Note that $u_H^\perp = e_3 \times u$. With a slight abuse of notation, we denote by $\widetilde{\rho}$, $\widetilde{p}$ the full density and the pressure, respectively, and by ρ and p the deviations of the actual density and pressure from the background profiles:

$$\rho = \widetilde{\rho} - \overline{\rho}, \quad p = \widetilde{p} - \overline{p}.$$

Then, taking into account (2.3), the leading order of the momentum equation in the Boussinesq system becomes

$$\frac{D^H u_H}{Dt} + u_3 \frac{\partial u_H}{\partial x_3} + f u_H^\perp = -\rho_b^{-1} \nabla^H p, \tag{2.4}$$

$$\frac{D^H u_3}{Dt} + u_3 \frac{\partial u_3}{\partial x_3} = -\rho_b^{-1} \frac{\partial p}{\partial x_3} - \frac{g}{\rho_b}\rho. \tag{2.5}$$

Equations (2.4) and (2.5) describe the horizontal and vertical balances of the momentum. The Boussinesq approximation is used in the right sides of (2.4) and (2.5), where we divide by ρ_b instead of the full density $\widetilde{\rho}$.

As the background density is constant, the leading order term in the continuity equation (2.2) becomes the incompressibility condition for the fluid flow:

$$\mathrm{div}_H u_H + \frac{\partial u_3}{\partial x_3} = 0. \tag{2.6}$$

In the next order of the continuity equation, we obtain

$$\frac{D\rho}{Dt} = -u_3 \frac{\partial \overline{\rho}}{\partial x_3}. \tag{2.7}$$

Equations (2.4), (2.5), (2.7) and (2.6) constitute the rotating Boussinesq system. The system is set in the half-space $\mathbb{R}^3_+ = \{x \in \mathbb{R}^3,\ \ x_3 \geq 0\}$. We prescribe the no flow boundary condition at $x_3 = 0$, that is,

$$u_3(x_1, x_2, 0) = 0.$$

The non-dimensional equations

Let us non-dimensionalize the system by introducing some typical scales into the problem. This will help us determine which terms in the equations should play the key role in the phenomena we are trying to model, and which can be neglected. We have the following physical parameters and scales: L is the mean horizontal length scale, U is the mean horizontal advective velocity, $T_e = L/U$ is the eddy turnover time scale, $T_R = f^{-1}$ is the rotation time scale associated with the planetary motion. Finally, the buoyancy time scale T_N is measured by $N = T_N^{-1}$, given by

$$N = \left(-g\rho_b^{-1} \frac{\partial \overline{\rho}}{\partial x_3} \right)^{1/2}.$$

It is easy to see that T_N defined above has the dimension of time. Let us rescale the variables accordingly:

$$x' = \frac{x}{L},\quad t' = \frac{t}{T_e},\quad u' = \frac{u}{U},\quad \rho' = \frac{\rho}{\rho_b},\quad p' = \frac{p}{\overline{p}}.$$

We arrive at the non-dimensional form of rotating Boussinesq system (omitting primes in the notation):

$$\frac{D^H u_H}{Dt} + u_3 \frac{\partial u_H}{\partial x_3} + (Ro)^{-1} u_H^{\perp} = -\overline{P} \nabla^H p \tag{2.8}$$

$$\frac{D^H u_3}{Dt} + u_3 \frac{\partial u_3}{\partial x_3} = -\overline{P} \frac{\partial p}{\partial x_3} - \Gamma \rho \tag{2.9}$$

$$\frac{D\rho}{Dt} = \Gamma^{-1} (Fr)^{-2} u_3. \tag{2.10}$$

The incompressibility condition does not change. Here Ro, Fr, Γ, $\overline{P}$ are non-dimensional parameters, defined as follows:

$$Ro = \frac{T_R}{T_e} = \frac{U}{Lf}$$

is the Rossby number, the ratio of the planetary rotation time scale to the horizontal advective motion time scale. The Froude number

$$Fr = \frac{T_N}{T_e} = \frac{U}{LN}$$

is the ratio of the buoyancy time scale to the horizontal advective motion time scale. The Euler parameter

$$\overline{P} = \frac{\overline{p}}{\rho_b U^2}$$

compares the mean pressure to the pressure of the inertial forces. Finally,

$$\Gamma = \frac{gL}{U^2},$$

is the ratio of the mean potential energy to the mean kinetic energy.

Ertel's theorem

The absolute vorticity takes into account the planetary rotation and is defined as

$$\omega_a = \omega + (Ro)^{-1} e_3,$$

where $\omega =$ curl u. We will also need the non-dimensionalized total density $\widetilde{\rho}$ given by

$$\widetilde{\rho} = \rho - \Gamma^{-1}(Fr)^{-2} x_3 + 1.$$

It satisfies

$$\frac{D\widetilde{\rho}}{Dt} = 0. \tag{2.11}$$

The following observation, known as Ertel's Theorem will be useful to us in the passage to the surface quasi-geostrophic equation.

Theorem 2.1 *For a smooth solution of non-dimensionalized rotating Boussinesq system, we have*

$$\frac{D}{Dt}(\omega_a \cdot \nabla\widetilde{\rho}) = 0. \tag{2.12}$$

Proof. Note that

$$\frac{D}{Dt}(\omega_a \cdot \nabla\widetilde{\rho}) = \frac{D\omega_a}{Dt} \cdot \nabla\widetilde{\rho} + \omega_a \cdot \frac{D\widetilde{\nabla\rho}}{Dt}. \tag{2.13}$$

Let us denote by ∇u the matrix with the entries $(\nabla u)_{ij} = \partial_j u_i$. Then, applying the gradient to (2.11), we get

$$\partial_t \widetilde{\rho} + (u \cdot \nabla)\nabla\widetilde{\rho} + (\nabla u)^t \nabla\widetilde{\rho} = 0. \tag{2.14}$$

Therefore, the second term in the right side of (2.13) is

$$\omega_a \cdot \frac{D\widetilde{\nabla\rho}}{Dt} = -\omega_a \cdot (\nabla u)^t \nabla\widetilde{\rho}.$$

Next, notice that (2.8) and (2.9) can be combined as

$$\frac{Du}{Dt} + (Ro)^{-1} e_3 \times u = -\overline{P} \nabla p - \Gamma \rho e_3. \tag{2.15}$$

Let us apply the curl to (2.15). The curl of the first term in the left side is

$$\operatorname{curl}\Big(\frac{Du}{Dt}\Big) = \operatorname{curl}(u_t + u \cdot \nabla u) = \frac{D\omega}{Dt} - \omega \cdot (\nabla u),$$

as in the vorticity formulation of the 3D Euler equations (verify this computation using the vector calculus if you have not seen it before!). For the second term in the left side of (2.15), we note that

$$\operatorname{curl}(e_3 \times u) = -\Big(\frac{\partial u_1}{\partial x_3}, \frac{\partial u_2}{\partial x_3}, \frac{\partial u_3}{\partial x_3}\Big) = -(\nabla u) e_3.$$

The first term in the right side of (2.15) is a gradient, hence its curl vanishes. The curl of the last term in the right side is

$$\operatorname{curl}(\rho e_3) = \nabla_H^{\perp} \rho,$$

where $\nabla_H^{\perp} = (\partial_2, -\partial_1)$. Putting everything together, we get

$$\frac{D\omega}{Dt} - \omega \cdot (\nabla u) - (Ro)^{-1} (\nabla u) e_3 = -\Gamma \nabla_H^{\perp} \rho.$$

As $\omega \cdot \nabla u = (\nabla u)\omega$, this is

$$\frac{D\omega_a}{Dt} = (\nabla u)\omega_a - \Gamma \nabla_H^{\perp} \rho. \tag{2.16}$$

Thus, the first term in the right side of (2.13) equals

$$\frac{D\omega_a}{Dt} \cdot \nabla\widetilde{\rho} = \nabla\widetilde{\rho} \cdot (\nabla u)\omega_a - \Gamma \nabla_H^{\perp} \rho \cdot \nabla_H \rho = \nabla\widetilde{\rho} \cdot (\nabla u)\omega_a. \tag{2.17}$$

Combining (2.14) and (2.17), we see that

$$\frac{D}{Dt}(\omega_a \cdot \nabla\widetilde{\rho}) = -\omega_a \cdot (\nabla u)^t \nabla\widetilde{\rho} + \nabla\widetilde{\rho} \cdot (\nabla u)\omega_a = 0,$$

finishing the proof. □

The quasi-geostrophic equations

Now we are ready to present the assumptions that reduce the rotating Boussinesq system to a simpler system of equations. This simpler system is well known, is often used, and is called quasi-geostrophic (QG).

The Rossby number is small. In a large scale atmospheric motion, typically the Rossby number is in the range $Ro \sim 10^{-2} - 10^{-3}$. Thus, we assume that the Rossby number is small, and denote $Ro = \varepsilon$, with $\varepsilon \ll 1$. The reader should keep in mind that in a tornado, $Ro \sim 10^3$. So this assumption is not unreasonable for modeling large scale atmospheric motion, but certainly not for modeling of a tornado.

The geostrophic balance: the rotation and the pressure are in a balance, that is the Coriolis force and the mean pressure are of a similar order in the horizontal momentum equation: $\overline{P} = Ro^{-1}$.

The Froude number is small. We assume that the Froude number is both small and is of the same order as the Rossby number: $Fr = FRo$, with the constant $F = O(1)$. This assumption means that the buoyancy time is small compared to the eddy turnover time, so the fluid is highly stratified. In the rest of this section we will assume that $F = 1$. Otherwise, it can be normalized this way in the quasi-geostrophic system below by rescaling the x_3 variable.

The kinetic energy is small: we assume that $\Gamma = (Fr)^{-1}$.

To summarize, the assumptions are

$$Ro = \varepsilon \ll 1, \quad \overline{P} = \varepsilon^{-1}, \quad Fr = \varepsilon, \quad \Gamma = \varepsilon^{-1}.$$

With these assumptions, the rotating Boussinesq system turns into

$$\frac{D^H u_H^\varepsilon}{Dt} + u_3^\varepsilon \frac{\partial u_H^\varepsilon}{\partial x_3} = -\varepsilon^{-1}(u_H^{\varepsilon,\perp} + \nabla^H p^\varepsilon) \tag{2.18}$$

$$\frac{D^H u_3^\varepsilon}{Dt} + u_3^\varepsilon \frac{\partial u_3^\varepsilon}{\partial x_3} = -\varepsilon^{-1}\left(\frac{\partial p^\varepsilon}{\partial x_3} + \rho^\varepsilon\right) \tag{2.19}$$

$$\frac{D\rho^\varepsilon}{Dt} = \varepsilon^{-1} u_3^\varepsilon, \ \ \mathrm{div}_H u_H^\varepsilon + \frac{\partial u_3^\varepsilon}{\partial x_3} = 0. \tag{2.20}$$

Also, Ertel's Theorem gives

$$\frac{D}{Dt}\left((\omega + \varepsilon^{-1} e_3)\cdot \nabla(\rho - \varepsilon^{-1}x_3)\right) = -\varepsilon^{-1}\frac{D}{Dt}\left(\omega_3 - \frac{\partial \rho}{\partial x_3} - \varepsilon\omega\cdot\nabla\rho\right) = 0. \tag{2.21}$$

We will think of ε as a small parameter, and will consider a formal asymptotic expansion of the solution into a series in powers of ε :

$$u_H^\varepsilon = u_H^{(0)} + \varepsilon u_H^{(1)} + O(\varepsilon^2); \ \ u_3^\varepsilon = u_3^{(0)} + \varepsilon u_3^{(1)} + O(\varepsilon^2)$$
$$\rho^\varepsilon = \rho^{(0)} + \varepsilon\rho^{(1)} + O(\varepsilon^2); \ \ p^\varepsilon = p^{(0)} + \varepsilon p^{(1)} + O(\varepsilon^2).$$

Let us substitute these expansions into the system (2.18), (2.19) and (2.20). The leading terms will be of the order ε^{-1}.

From (2.18), we obtain

$$u_H^{(0)} = -\nabla_H^\perp p^{(0)}, \tag{2.22}$$

the geostrophic balance. It implies, in particular, that

$$\Delta_H p^{(0)} = \omega_3^{(0)}. \tag{2.23}$$

The vertical momentum equation gives

$$\frac{\partial p^{(0)}}{\partial x_3} = -\rho^{(0)}. \tag{2.24}$$

This equality is known as the hydrostatic balance.

The density equation leads to $u_3^{(0)} = 0$. The incompressibility condition is then automatically satisfied to the leading order:

$$\mathrm{div}_H u_H^{(0)} + \frac{\partial u_3^{(0)}}{\partial x_3} = \mathrm{div}_H \nabla_H^{\perp} p^{(0)} = 0.$$

Next, since $u_3^{(0)} = 0$, the Ertel theorem , to the leading order, gives

$$\frac{D^H}{Dt}\left(\omega_3^{(0)} - \frac{\partial \rho^{(0)}}{\partial x_3}\right) = 0. \tag{2.25}$$

According to (2.24), we have

$$\frac{\partial \rho^{(0)}}{\partial x_3} = -\frac{\partial^2 p^{(0)}}{\partial x_3^2}.$$

Substituting this identity and (2.23) into (2.25), we obtain

$$\frac{D^H}{Dt}\left(\Delta_H p^{(0)} + \frac{\partial^2 p^{(0)}}{\partial x_3^2}\right) = 0. \tag{2.26}$$

The system of equations (2.23), (2.24) and (2.26):

$$\Delta_H p^{(0)} = \omega_3^{(0)} \tag{2.27}$$

$$\frac{\partial p^{(0)}}{\partial x_3} = -\rho^{(0)} \tag{2.28}$$

$$\frac{D^H}{Dt}\left(\Delta_H p^{(0)} + \frac{\partial^2 p^{(0)}}{\partial x_3^2}\right) = 0, \tag{2.29}$$

is called the quasi-geostrophic (QG) system, and the quantity in the parentheses in (2.26) is called the "potential vorticity". One can, in fact, prove the convergence of the solutions of the rotating Boussinesq system to the solutions of the QG equation as $\varepsilon \to 0$ [88] (for a finite time while solutions are known to remain smooth).

The surface quasi-geostrophic equation

To obtain the surface quasi-geostrophic equation, we make one more assumption: the potential vorticity is originally zero, that is,

$$\Delta p^{(0)}(0, x) = 0.$$

See [72] for a discussion of this assumption. It follows from (2.26) that this remains true for all times, so $p^{(0)}(x, t)$ is a harmonic function:

$$\Delta p^{(0)}(t, x) = 0, \quad x \in \mathbb{R}^3_+,\ t \geq 0. \tag{2.30}$$

Consider the boundary of our half-space, the plane $x_3 = 0$. On this plane, according to (2.20), and since u_3 vanishes on the boundary, the density $\rho^{(0)}(t, x_H, 0)$ satisfies a closed-form advection equation

$$\partial_t \rho^{(0)} + u_H \cdot \nabla^H \rho^{(0)} = 0. \tag{2.31}$$

In addition, we get from (2.24)

$$\left.\frac{\partial p^{(0)}}{\partial x_3}\right|_{x_3=0} = -\rho^{(0)}(x_H, 0). \tag{2.32}$$

Let us recall the following fact.

Lemma 2.2 *Suppose that a function p is harmonic and bounded in $\mathbb{R}^3_+$, and is continuous up to the boundary along with the first and second order derivatives. Then*

$$\left.\frac{\partial p}{\partial x_3}\right|_{x_3=0} = -(-\Delta_H)^{1/2} p(x_1, x_2, 0), \tag{2.33}$$

where Δ_H denotes the Laplacian in horizontal variables x_1 and x_2.

That is, the normal derivative at the boundary of a function harmonic in the half-space $\mathbb{R}^3_+$ is equal to the planar fractional 1/2-Laplacian of the trace of this harmonic function on the boundary.

Let us postpone the proof of the lemma and finish the derivation of the SQG equation. Given the result of Lemma 2.2, (2.32) leads to the relation between the density and the pressure on the surface $\{x_3 = 0\}$:

$$\rho^{(0)}(t, x_H, 0) = (-\Delta_H)^{1/2} p^{(0)}(t, x_H, 0).$$

But then, by geostrophic balance (2.22), we have on the boundary $x_3 = 0$

$$u_H^{(0)}(t, x_H, 0) = -\nabla_H^\perp p^{(0)}(t, x_H, 0) = -\nabla_H^\perp (-\Delta_H)^{-1/2} \rho^{(0)}(t, x_H, 0). \tag{2.34}$$

If we now drop the sub-script H and set $\theta(t, x) = \rho^{(0)}(t, x, 0)$, then equations (2.31) and (2.34) together become

$$\theta_t + u \cdot \nabla\theta = 0, \tag{2.35}$$

$$u(t, x) = -\nabla^\perp (-\Delta_H)^{-1/2} \theta(t, x), \tag{2.36}$$

which is exactly the inviscid SQG equation (**modulo the sign change but such is life**). One can solve this equation on the plane and then recover $p^{(0)}$ on the half-space from the value of its normal derivative, thus solving the QG system in the leading order for the particular class of initial data with vanishing potential vorticity.

Let us also mention that another version of the SQG equation, which is well motivated from the physical point of view is the critical SQG equation, with the dissipative term described by the fractional Laplacian of the power 1/2:

$$\partial_t \rho^{(0)} + u_H \cdot \nabla^H \rho^{(0)} + (-\Delta_H)^{1/2} \rho^{(0)} = 0. \tag{2.37}$$

Note that, since by our assumptions the function $p^{(0)}$ is harmonic in $\mathbb{R}^3_+$, the relationship (2.32) implies that $\rho^{(0)}$ is harmonic in $\mathbb{R}^3_+$ as well. Then, Lemma 2.2 implies that

$$(-\Delta_H)^{1/2} \rho^{(0)}(t, x_H, 0) = -\frac{\partial \rho^{(0)}}{\partial x_3}(t, x_H, 0).$$

One may use this identity tp show that the additional dissipative term in (2.37) describes the heat loss due to a boundary layer effect and models the so-called Ekman pumping. We refer to, for example, [102] for further discussion of this phenomenon.

It remains to prove Lemma 2.2. The formal reason why it holds is quite simple: taking the Fourier transform in the horizontal variables of the Laplace equation

$$\Delta_H p + \frac{\partial^2 p}{\partial x_3^2} = 0,$$

gives

$$\frac{\partial^2 \hat{p}(\xi, x_3)}{\partial x_3^2} = 4\pi^2 |\xi|^2 \hat{p}(\xi, x_3).$$

Boundedness of $\hat{p}(\xi, x_3)$ as $x_3 \to +\infty$ implies that

$$\hat{p}_3(\xi, x_3) = \hat{p}(\xi, 0) e^{-2\pi|\xi| x_3},$$

so that

$$\frac{\partial \hat{p}(\xi, 0)}{\partial x_3} = -2\pi|\xi|\hat{p}(\xi, 0),$$

which is the Fourier transform form of (2.33).

To give a more careful proof, recall that the two dimensional Poisson kernel is given by

$$\mathcal{P}_h^2(x) = c_2 h(|x|^2 + h^2)^{-3/2},$$

where c_2 is a positive constant. Observe that $\mathcal{P}_h^2(x)$ is the derivative of the three dimensional Laplacian Green's function $(|x|^2 + h^2)^{-1/2}$ with respect to h. We first claim that

$$p(x_1, x_2, h) = \mathcal{P}_h^2 * p = c_2 \int_{\mathbb{R}^2} \frac{h p(y_1, y_2, 0)}{(|x-y|^2 + h^2)^{3/2}} dy_1 dy_2. \tag{2.38}$$

Indeed, the link of $\mathcal{P}_h^2$ and the Laplacian's Green's function allows to verify that the function in the right side of (2.38) is harmonic in $\mathbb{R}^3_+$. Fruthermore, one can show that for the right choice of c_2, the Poisson kernel $\mathcal{P}_h^2$ is an approximation of identity as $h \to 0$. Therefore, the right hand side of (2.38) converges to $p(x_1, x_2, 0)$ as $h \to 0$. Well known uniqueness properties of harmonic functions imply then that the equality (2.38) holds.

Exercise 2.3 Verify the claims in the preceding paragraph via a direct computation.

In the next section we will discuss the Poisson kernels in more detail. In particular, we will see that the Fourier transform of the Poisson kernel $\hat{\mathcal{P}_h^2}(k)$ is $e^{-h|k|}$, as expected from the formal computation at the start of this proof. This helps explain the convergence of $\mathcal{P}_h^2$ to the δ-function as $h \to 0$.

Now, we have

$$\frac{\partial p}{\partial h}(x_1, x_2, h) = c_2 \int_{\mathbb{R}^2} \left(\frac{p(y_1, y_2, 0)}{(|x-y|^2 + h^2)^{3/2}} - \frac{3h^2 p(y_1, y_2, 0)}{(|x-y|^2 + h^2)^{5/2}} \right) dy_1 dy_2.$$

A direct computation which is best carried out in the polar coordinates shows that the kernel

$$\frac{1}{(|x|^2+h^2)^{3/2}} - \frac{3h^2}{(|x|^2+h^2)^{5/2}}$$

is mean zero in $\mathbb{R}^2$, so we may write

$$\frac{\partial p}{\partial h}(x_1,x_2,h) = c_2 \int\limits_{\mathbb{R}^2} \Big(\frac{1}{(|x-y|^2+h^2)^{3/2}} - \frac{3h^2}{(|x-y|^2+h^2)^{5/2}}\Big)(p(y,0)-p(x,0))dy. \quad (2.39)$$

Observe that, due to the symmetry of the kernel, the first order terms in the Taylor expansion of $p(y,0)$ near $(x,0)$ integrate out to zero, so that

$$\begin{aligned} \Big|\int_{\mathbb{R}^2} \frac{3h^2}{(|x-y|^2+h^2)^{5/2}}(p(y,0)-p(x,0))\,dy\Big| &\le \int_{\mathbb{R}^2} \frac{Ch^2\min(|x-y|^2,1)}{(|x-y|^2+h^2)^{5/2}}\,dy \\ &\le C\left(h^2 + \int_{|x-y|\le 1} \frac{h^{1/2}}{|x-y|^{3/2}}\,dy\right) \to 0, \qquad (2.40) \end{aligned}$$

as $h \to 0$. On the other hand, one can show that

$$\lim_{h\to 0}\int_{\mathbb{R}^2} \frac{1}{(|x-y|^2+h^2)^{3/2}}(p(y,0)-p(x,0))\,dy = P.V.\int_{\mathbb{R}^2} \frac{1}{|x-y|^3}(p(y,0)-p(x,0))\,dy, \quad (2.41)$$

where the principal value integral is defined as

$$P.V.\int_{\mathbb{R}^2} \frac{1}{|x-y|^3}(p(y,0)-p(x,0))\,dy = \lim_{\varepsilon\to 0}\int_{|x-y|\ge\varepsilon} \frac{1}{|x-y|^3}(p(y,0)-p(x,0))\,dy. \quad (2.42)$$

Exercise 2.4 Verify (2.41).

In Lemma 4.1 below, we will see that (2.42) is exactly equal to $(-\Delta_H)^{1/2}p(x_1,x_2,0)$. This completes the proof of the lemma. □

3 Ruling out the front singularity

In the paper of Constantin, Majda and Tabak, where the SQG equation was essentially first introduced in the mathematical literature, a scenario for a finite time singularity formation has been proposed. The scenario involves a particular initial condition, where the level sets of the temperature (density) contain a hyperbolic saddle point. Numerical simulations showed that the arms of the saddle tend to close in a finite time, producing a sharp front along a curve. Later, more careful numerical studies of Ohkitani-Yamada [101] and Constantin-Nie-Schorghofer [37] suggested that instead of the finite time singularity, the derivatives of the temperature grow as a double exponential in time. Analytically, the closing hyperbolic saddle scenario has been studied by D. Cordoba in [39], where he showed that the angle of the saddle cannot decrease faster than a double exponential in time. Our goal in this section is to provide a proof of the result due to D. Cordoba and C. Fefferman [36], which shows that the solutions of the SQG equation cannot in general form a front-like singularity along a curve – under certain mild conditions on the geometry of the singularity formation. We will mostly follow the arguments of [36] in this section.

Assume that a smooth decaying function $\theta(t,x)$ solves the inviscid SQG equation

$$\partial_t\theta + (u\cdot\nabla)\theta = 0, \tag{3.1}$$
$$u = (\partial_2(-\Delta)^{-1/2}\theta, -\partial_1(-\Delta)^{-1/2}\theta), \tag{3.2}$$

with the initial condition $\theta(0,x) = \theta_0(x)$, for $t\in[0,T)$. Let us define the stream function

$$\psi = (-\Delta)^{-1/2}\theta,$$

so that $u = \nabla^\perp\psi = (\partial_2\psi, -\partial_1\psi)$. The Fourier transforms of the functions ψ and θ are related by

$$\hat\psi(\xi) = \frac{1}{2\pi|\xi|}\hat\theta(\xi).$$

The function $\hat g(\xi) = 1/|\xi|$, in dimension $n=2$, is a Schwartz distribution, which is radially symmetric and homogeneous of degree (-1). Therefore, its inverse Fourier transform is a Schwartz distribution $g(x)$ which is homogeneous of degree $-2-(-1) = -1$, and is also radially symmetric. Hence, $g(x) = c/|x|$, with a constant c that may be computed but is of no particular importance so we will not dwell on its value. Thus, the stream function is given by

$$\psi(x) = (-\Delta)^{-1/2}\theta(x) = c_1\int_{\mathbb{R}^2}\frac{\theta(y)}{|x-y|}\,dy, \tag{3.3}$$

with an appropriate constant c_1.

Let us now assume that a level curve of the function $\theta(t,x)$ can be parametrized by

$$x_2 = \phi_\rho(t,x_1) \text{ for } x_1\in[a,b], \tag{3.4}$$

where $\phi_\rho\in C^1([0,T),[a,b])$, in the sense that

$$\theta(t,x_1,\phi_\rho(t,x_1)) = \rho \text{ for } x_1\in[a,b],\ \ t\in[0,T). \tag{3.5}$$

We obtain from (3.4) and (3.5):

$$\frac{\partial\theta}{\partial x_1} + \frac{\partial\theta}{\partial x_2}\frac{\partial\phi_\rho}{\partial x_1} = 0,\ \ \frac{\partial\theta}{\partial t} + \frac{\partial\theta}{\partial x_2}\frac{\partial\phi_\rho}{\partial t} = 0, \tag{3.6}$$

when evaluated at $(t,x_1,\phi_\rho(t,x_1))$. The SQG equation (3.1) for $\theta(t,x)$ used in (3.6), together with the expression $u = \nabla^\perp\psi$, gives

$$\frac{\partial\phi_\rho}{\partial t} = -\frac{\partial\theta/\partial t}{\partial\theta/\partial x_2} = \frac{(u\cdot\nabla\theta)}{\partial\theta/\partial x_2} = \frac{\partial\psi}{\partial x_2}\frac{\partial\theta/\partial x_1}{\partial\theta/\partial x_2} - \frac{\partial\psi}{\partial x_1} = -\frac{\partial\psi}{\partial x_2}\frac{\partial\phi_\rho}{\partial x_1} - \frac{\partial\psi}{\partial x_1}. \tag{3.7}$$

On the other hand, we also have

$$\frac{\partial}{\partial x_1}(\psi(t,x_1,\phi_\rho(t,x_1))) = \frac{\partial\psi}{\partial x_1} + \frac{\partial\psi}{\partial x_2}\frac{\partial\phi_\rho}{\partial x_1}, \tag{3.8}$$

and therefore,

$$\frac{\partial \phi_\rho}{\partial t} = -\frac{\partial}{\partial x_1}(\psi(x_1, \phi_\rho(x_1,t),t)). \tag{3.9}$$

Equation (3.9) can be used to compute the change in the area between two level curves of θ over the interval $x_1 \in [a,b]$:

$$\begin{aligned}&\frac{d}{dt}\int_a^b (\phi_{\rho_2}(t,x_1) - \phi_{\rho_1}(t,x_1))\, dx_1 = -\psi(t,b,\phi_{\rho_2}(t,b)) + \psi(t,a,\phi_{\rho_2}(t,a))\\ &+\psi(t,b,\phi_{\rho_1}(t,b)) - \psi(t,a,\phi_{\rho_1}(t,a))\\ &= \psi(t,b,\phi_{\rho_1}(t,b)) - \psi(t,b,\phi_{\rho_2}(t,b)) - (\psi(t,a,\phi_{\rho_1}(t,a)) - \psi(t,a,\phi_{\rho_2}(t,a)).\end{aligned} \tag{3.10}$$

Let us define the local thickness of the front $\delta(x_1,t)$ by

$$\delta(t,x_1) = |\phi_{\rho_2}(t,x_1) - \phi_{\rho_1}(t,x_1)|.$$

We say that the solution exhibits a semi-uniform collapse on a curve if (3.4) and (3.5) hold (for fixed values ρ_1 and ρ_2), if

$$\min_{x_1\in[a,b]}\delta(t,x_1) \to 0, \text{ as } t \to T,$$

and if the thickness of the front $\delta(t,x_1)$ satisfies

$$\min_{x_1\in[a,b]}\delta(t,x_1) \geq c\max_{x_1\in[a,b]}\delta(t,x_1) \tag{3.11}$$

for all $t \leq T$, with a constant $c > 0$ that is independent of t. We call the value $b-a$ the length of the front. The first condition above means that the front collapses and $\theta(t,x)$ becomes singular at the time T, and the second means that the front collapse is uniform, so that the singularity at the time T happens along a curve.

The following theorem shows that semi-uniform collapse along a front of a positive length is not possible for the solutions of the SQG equation.

Theorem 3.1 *For a solution of the SQG equation with a semi-uniform front, the thickness of the front must satisfy*

$$\delta(t) \equiv \frac{1}{b-a}\int_a^b (\phi_{\rho_2}(x_1,t) - \phi_{\rho_1}(x_1,t))\, dx_1 \geq e^{-e^{At+B}} \tag{3.12}$$

for all t for which the semi-uniform front persists. Here, the constants A and B may depend only on the length of the front, the constant c in (3.11)*, the initial thickness of the front, and on the L^1 and L^∞ norms of the initial data θ_0.*

Proof. Since the fluid velocity in the SQG equation is incompressible, it is straightforward to check that any L^p, $1 \leq p \leq \infty$ norm of smooth solution does not change in time. From (3.10), we see that

$$\left|\frac{d}{dt}\delta(t)\right| \leq \frac{C}{b-a}\sup_{x_1\in[a,b]}|\psi(t,x_1,\phi_{\rho_2}(t,x_1)) - \psi(t,x_1,\phi_{\rho_1}(t,x_1))|. \tag{3.13}$$

Next, using (3.3), we can write

$$\psi(t,z_2)-\psi(t,z_1)=c_1\int_{\mathbb{R}^2}\theta(t,y)\left(\frac{1}{|y-z_2|}-\frac{1}{|y-z_1|}\right)\,dy. \tag{3.14}$$

Since we are interested in the regime where the thickness of the front is small, and the front collapse is semi-uniform, we can assume without loss of generality that $\tau=|z_2-z_1|<1/2$. Let us split the integral in the right side of (3.14) into three regions:

$$|y-z_1|\le 2\tau,\quad 2\tau\le|y-z_1|\le 1,\quad \text{and } |y-z_1|\ge 1.$$

We denote the contributions from these three regions by I_1, I_2 and I_3. To estimate I_1, note that

$$\begin{aligned}|I_1|&\le C\|\theta_0\|_{L^\infty}\int_{|z_1-y|\le 2\tau}\left|\frac{1}{|y-z_2|}-\frac{1}{|y-z_1|}\right|\,dy\\&\le C\|\theta_0\|_{L^\infty}\int_{|z_1-y|\le 2\tau}\Big(\frac{1}{|y-z_2|}+\frac{1}{|y-z_1|}\Big)dy\le C\tau.\end{aligned} \tag{3.15}$$

The term I_3 is also easy to bound: as $|z_1-z_2|\le 1/2$, we have

$$|I_3|\le c_1\int_{|z_1-y|\ge 1}|\theta(y)|\left|\frac{z_2-z_1}{|y-z_1||y-z_2|}\right|\,dy\le 2c_1\|\theta_0\|_{L^1}\tau\le C\tau. \tag{3.16}$$

Finally, let us estimate I_2. The mean value theorem implies that there is a point $s(y)$ on the line connecting z_1 and z_2 such that

$$\left|\frac{1}{|y-z_2|}-\frac{1}{|y-z_1|}\right|=\left|\frac{(y-s)\cdot(z_2-z_1)}{|y-s|^3}\right|.$$

As we are considering the region $2\tau\le|y-z_1|\le 1$, for any point s on the line connecting z_1 and z_2, we have

$$|y-s|\ge|y-z_1|-|z_1-s|\ge|y-z_1|-\tau\ge\frac{1}{2}|y-z_1|.$$

Taking this into account, we obtain

$$\begin{aligned}|I_2|&\le C\|\theta_0\|_{L^\infty}\int_{2\tau\le|y-z_1|\le 1}\left|\frac{(y-s(y))\cdot(z_2-z_1)}{|y-s(y)|^3}\right|dy\le C\|\theta_0\|_{L^\infty}\tau\int_{2\tau\le|y-z_1|\le 1}\frac{1}{|y-z_1|^2}\,dy\\&\le C\tau\log\tau^{-1}.\end{aligned} \tag{3.17}$$

Combining (3.15), (3.16) and (3.17), we get

$$|\psi(z_2,t)-\psi(z_1,t)|\le M|z_2-z_1|\log|z_2-z_1|^{-1}, \tag{3.18}$$

whenever $|z_2-z_1|<1/2$, with the constant M that depends only on the initial data θ_0.

It follows from (3.13) and (3.18) that

$$\left|\frac{d}{dt}\delta(t)\right| \leq \frac{M}{b-a}\sup_{x_1\in[a,b]}\left(|\phi_{\rho_2}(x_1,t)-\phi_{\rho_1}(x_1,t)|\log|\phi_{\rho_2}(x_1,t)-\phi_{\rho_1}(x_1,t)|^{-1}\right)$$
$$\leq \frac{CM}{b-a}\delta(t)\log\delta(t)^{-1}. \tag{3.19}$$

The last inequality follows from the definition of $\delta(t)$, and the assumption that the front is semi-uniform. By standard arguments, the differential inequality (3.19) implies that

$$\delta(t) \geq \delta(0)^{e^{\frac{CM}{b-a}t}},$$

from which the conclusion of the theorem follows. □

In general, the question whether solutions of the inviscid SQG equation develop a singularity in a finite time remains open. There is no clear candidate scenario for the singularity formation in the case of the smooth initial data, but, on the other hand, the current analytical methods fall far short from providing a general global regularity proof. In the next section, we will consider the SQG equation with the dissipation in the form of a fractional Laplacian. As we have mentioned, the fractional Laplacian of the power 1/2 is physically motivated. More general powers are then natural to consider for mathematical reasons. Interestingly, as we will see in the next section, the 1/2 power also carries a very special mathematical meaning – it is critical. This means, informally, that at this power the dissipation precisely balances the strength of the nonlinearity. While one can expect the global regularity when the dissipation is stronger than the nonlinearity, the critical case is more subtle and can not be decided in a simple way. In the supercritical case, when the dissipation is weak, one expects that the finite time singularity vs global regularity question should be decided by the nonlinearity. However, things may be more complicated than that – for many supercritical PDE this basic question remains open.

4 Global regularity for the subcritical SQG equation

We will discuss below the dissipative SQG equation

$$\frac{\partial\theta}{\partial t}+(u\cdot\nabla)\theta+(-\Delta)^{\alpha}\theta=0,\;\; u=\nabla^{\perp}(-\Delta)^{1/2}\theta,\;\; \theta(x,0)=\theta_0(x), \tag{4.1}$$

with $\alpha > 0$. We will start with the maximum principle that applied to solutions of the dissipative SQG equation with all $\alpha > 0$ and shows that $\|\theta(t)\|_{L^\infty} \leq \|\theta_0\|_{L^\infty}$. Next, we will use the maximum principle and a Gagliardo-Nirenberg inequality to prove the global in time existence of smooth solutions in the subcritical case $\alpha > 1/2$. The "truly difficult" case $\alpha = 1/2$ requires a very different approach and is considered in the next section.

The maximum principle

The first step in the analysis of the dissipative SQG equation is to establish the L^∞ maximum principle. Let us begin by deriving an explicit formula for the fractional Laplacian. The argument below is taken from [41].

Lemma 4.1 *Let $0 < \alpha < 1$, and let $f(x)$ be a $C_0^\infty(\mathbb{R}^d)$ function, then*

$$(-\Delta)^\alpha f(x) = C_\alpha P.V. \int_{\mathbb{R}^2} \frac{f(x) - f(y)}{|x-y|^{d+2\alpha}}\, dy. \tag{4.2}$$

Here, $C_\alpha > 0$ is a constant that can be computed explicitly. The same formula remains valid in the periodic case if $f \in C^\infty(\mathbb{T}^d)$, given that we periodize $f(x)$ and regard the integral (4.2) in this sense.

Proof. Recall the following formula which generalizes (3.3) to all $\alpha \in (0,1)$:

$$r(x) = (-\Delta)^{-\alpha} f(x) = c_\alpha P.V. \int_{\mathbb{R}^d} \frac{f(y)}{|x-y|^{d-2\alpha}}\, dy. \tag{4.3}$$

This is proved as (3.3): the Fourier transform of the function $r(x)$ is

$$\hat{r}(\xi) = \frac{1}{(2\pi|\xi|)^{2\alpha}} \hat{f}(\xi).$$

The inverse Fourier transform of $\hat{g}(\xi) = 1/|\xi|^{2\alpha}$ is a radially symmetric function which is homogeneous of degree $-d - (-2\alpha)$. Thus, $g(x) = c_\alpha/|x|^{d-2\alpha}$. Then, (4.3) follows from the fact that the multiplication operation on the Fourier side corresponds to a convolution in the physical space.

We now can write

$$\begin{aligned}(-\Delta)^\alpha f(x) &= (-\Delta)^{\alpha-1}(-\Delta) f(x) = -c_\alpha P.V. \int_{\mathbb{R}^d} \frac{\Delta f(y)}{|x-y|^{d-2+2\alpha}}\, dy \\ &= -c_\alpha \lim_{\varepsilon \to 0} \int_{|x-y| \geq \varepsilon} \frac{\Delta_y (f(y) - f(x))}{|x-y|^{d-2+2\alpha}}\, dy := \lim_{\varepsilon \to 0} (-\Delta)^\alpha_\varepsilon f(x).\end{aligned}$$

An application of Green's theorem gives us (here, we denote by ν_y the inward normal to the sphere $|x-y| = \varepsilon$ at the point y)

$$\begin{aligned}(-\Delta)^\alpha_\varepsilon f(x) = C_\alpha \int_{|x-y| \geq \varepsilon} \frac{f(x) - f(y)}{|x-y|^{d+2\alpha}}\, dy + \int_{|x-y|=\varepsilon} (f(y) - f(x)) \frac{\partial}{\partial \nu_y} \left[|x-y|^{-d+2-2\alpha}\right] d\sigma(y) \\ - \int_{|x-y|=\varepsilon} |x-y|^{-d+2-2\alpha} \frac{\partial (f(y) - f(x))}{\partial \nu_y}\, d\sigma(y) = I_1 + I_2 + I_3.\end{aligned} \tag{4.4}$$

The last two terms can be estimated as

$$I_2 = \varepsilon^{-d+1-2\alpha} \int_{|x-y|=\varepsilon} (f(y) - f(x))\, d\sigma(y) = O(\varepsilon^{2-2\alpha}),$$

and

$$I_3 = \varepsilon^{-d+2-2\alpha} \int_{|x-y|=\varepsilon} \frac{\partial (f(x) - f(y))}{\partial \nu_y}\, d\sigma(y) = O(\varepsilon^{2-2\alpha}).$$

Therefore, passing to the limit as $\varepsilon \to 0$, we obtain (4.2). The periodic analog can be derived similarly to how it was done in Lemma 2.20 in the previous chapter. □

The formula (4.2) is very convenient for the derivation of the L^∞ maximum principle.

Lemma 4.2 *Suppose that $f(x) \in C_0^\infty(\mathbb{R}^2)$, or $f \in C^\infty(\mathbb{T}^2)$ and mean zero, $0 \le \alpha \le 1$, and $1 \le p < \infty$, then*

$$\int |f(x)|^{p-2} f(x)(-\Delta)^\alpha f(x)\, dx \ge 0. \tag{4.5}$$

Proof. The cases $\alpha = 0$ and $\alpha = 1$ are easy to check directly. For $0 < \alpha < 1$, we have

$$\int |f|^{p-2} f(-\Delta)^\alpha f\, dx = \lim_{\varepsilon\to 0} \int |f|^{p-2} f(-\Delta)^\alpha_\varepsilon f\, dx = \lim_{\varepsilon\to 0} \int |f|^{p-2} f I_1\, dx,$$

where I_1 is the same as above in (4.4). Observe that

$$\begin{aligned}\int |f|^{p-2} f I_1\, dx &= c_\alpha \int\int_{|x-y|\ge\varepsilon} |f(x)|^{p-2} f(x) \frac{(f(x)-f(y))}{|x-y|^{2+2\alpha}}\, dxdy \\ &= -c_\alpha \int\int_{|x-y|\ge\varepsilon} |f(y)|^{p-2} f(y) \frac{(f(x)-f(y))}{|x-y|^{2+2\alpha}}\, dxdy\end{aligned}$$

by relabeling the variables. Symmetrizing, we get

$$\int |f|^{p-2} f I_1\, dx = \frac{c_\alpha}{2} \int\int_{|x-y|\ge\varepsilon} \left(|f(x)|^{p-2} f(x) - |f(y)|^{p-2} f(y)\right) \frac{(f(x)-f(y))}{|x-y|^{2+2\alpha}}\, dxdy \ge 0,$$

so that (4.5) holds. □

Now the L^∞ maximum principle follows easily and in more generality.

Corollary 4.3 (L^∞ Maximum Principle) *Let $\theta(t,x)$ and $u(t,x)$ be smooth functions on either $\mathbb{R}^2$ or $\mathbb{T}^2$ (rapidly decaying in the $\mathbb{R}^2$ case) satisfying*

$$\theta_t + u\cdot\nabla\theta + (-\Delta)^\alpha\theta = 0,$$

with $0 \le \alpha \le 1$, and either $\nabla\cdot u = 0$ or $u_i = \partial_i G(\theta)$, with some smooth function $G(\theta)$. Then for $1 \le p \le \infty$ we have

$$\|\theta(\cdot,t)\|_{L^p} \le \|\theta_0\|_{L^p}.$$

Proof. We compute

$$\frac{d}{dt}\int |\theta|^p\, dx = p\int |\theta|^{p-2}\theta(-u\cdot\nabla\theta - (-\Delta)^\alpha\theta)\, dx = -p\int |\theta|^{p-2}\theta(-\Delta)^\alpha\theta\, dx \le 0,$$

where we integrated by parts, used that $\nabla\cdot u = 0$ or $u_i = \partial_i G(\theta)$, and applied Lemma 4.2. □

The regularity for the subcritical SQG equation

The L^∞ maximum principle makes the dissipation exponent $\alpha = 1/2$ critical for the SQG equation (as well as for the Burgers equation). This means that for $\alpha > 1/2$ the dissipation term is strong enough to control the nonlinearity and prevent the singularity of the solutions in a "simplistic fashion". One can see this intuitively by comparing the relative sizes of the nonlinear term $u \cdot \nabla\theta$ and the dissipative term $(-\Delta)^\alpha\theta$. We know that $\|\theta\|_{L^\infty}$ is controlled. Then, since for the SQG equation the fluid flow

$$u = \nabla^\perp((-\Delta)^{1/2}\theta),$$

is dimensionally the same as θ, we can expect a similar control of u. This is not quite true: u is the Riesz transform of θ and, as we have discussed in the previous chapter, the Riesz transform is bounded on L^p for $1 < p < \infty$ but not on L^∞. But let us for now proceed with a rough heuristic argument. Assuming we can control $\|u\|_{L^\infty}$, the balance is between the terms $\|u\|_{L^\infty}|\nabla\theta|$ and $(-\Delta)^\alpha\theta$. If there is a sharp front of width h then the first term is of the size $\|u\|_{L^\infty}h^{-1}$, while the second is of the size $h^{-2\alpha}$. Because of the (presumed) bound on $\|u\|_{L^\infty}$, the dissipation should win if $\alpha > 1/2$. If $\alpha = 1/2$, then, naively, one would expect nonlinearity to be able to dominate for large initial data since $\nabla\theta$ is, roughly, of the same size as $(-\Delta)^{1/2}\theta$ and $\|u\|_{L^\infty}$ would be "large". Somewhat surprisingly, this is not the case, and global regularity remains true for $\alpha = 1/2$. We will first sketch global regularity proof for the subcritical case $\alpha > 1/2$, which is a fairly standard argument. We will then give a proof of the critical regularity, which is also simple but uses novel ideas.

Let us begin by stating the analog of the Biot-Savart formula for the SQG equation velocity.

Proposition 4.4 *The velocity u for the SQG equation can be determined from θ by the following formula*

$$u(x) = \frac{1}{2\pi} P.V. \int_{\mathbb{R}^2} \frac{(x-y)^\perp}{|x-y|^3}\theta(y)\,dy. \tag{4.6}$$

The formula is valid for $\theta \in C^1 \cap L^1(\mathbb{R}^2)$ or $\theta \in C^1(\mathbb{T}^2)$ with zero mean. In the latter case, we extend θ periodically to all $\mathbb{R}^2$ in the integral (4.6)*.*

Proof. Consider the $\mathbb{R}^2$ case first. According to (4.3), we have

$$(-\Delta)^{-1/2}\theta(x) = c\int_{\mathbb{R}^2} \frac{f(y)}{|x-y|}\,dy,$$

and the explicit constant $c = (2\pi)^{-1}$ can be found in [110]. Computing the distributional derivatives of the above expression, we obtain (4.6). The periodic case can be handled in the manner similar to the 2D Euler case in Lemma 2.20. □

Exercise 4.5 Work out carefully the details of the sketched argument, justifying the differentiation, using the definition of the principal value.

The main result in the subcritical case is the following theorem.

Theorem 4.6 *Let $\theta_0(x) \in C^\infty(\mathbb{T}^2)$, and $\alpha > 1/2$. Then there exists a unique global smooth solution $\theta(t,x)$ of* (4.1) *with the initial condition θ_0.*

Proof. We choose to present the $\mathbb{T}^2$ case but the whole space argument is similar. We also do not pursue the most general assumptions on the initial data. As we have discussed in the introduction, the local in time existence of a smooth solution can be established by standard means. The key for the global in time existence argument is to obtain global a priori bounds on the sufficiently high order Sobolev norms of solution. This would allow us to use the local in time argument repeatedly and to continue the solution up to an arbitrary time.

Let us multiply (4.1) by $(-\Delta)^s\theta$ and integrate. We get

$$\frac{1}{2}\frac{d}{dt}\|\theta\|_{H^s}^2 + \int (-\Delta)^s\theta(-\Delta)^\alpha\theta dx = \int_{\mathbb{T}^2} (u\cdot\nabla)\theta(-\Delta)^s\theta\, dx.$$

As

$$\int_{\mathbb{T}^2} (-\Delta)^s\theta(-\Delta)^\alpha\theta dx = \int_{\mathbb{T}^2} |(-\Delta)^{(\alpha+s)/2}\theta|^2 dx,$$

we deduce that

$$\frac{1}{2}\frac{d}{dt}\|\theta\|_{H^s}^2 + \|\theta\|_{H^{s+\alpha}}^2 = \int_{\mathbb{T}^2} (u\cdot\nabla)\theta(-\Delta)^s\theta\, dx. \tag{4.7}$$

The second term in the left side of (4.7) is the dissipation that helps us, while the term in the right side, which comes from the nonlinearity, is "dangerous".

Note that $(-\Delta)^s\theta$ is a sum of derivatives of the order $2s$ of the form $\partial_{i_1}\partial_{i_1}\dots\partial_{i_s}\partial_{i_s}\theta$. Let us integrate by parts, transferring exactly half of the matching derivatives to the term $(u\cdot\nabla)\theta$. We obtain a sum of terms of the form

$$\int_{\mathbb{T}^2} \partial_{i_1}\dots\partial_{i_s}((u\cdot\nabla)\theta)\partial_{i_1}\dots\partial_{i_s}\theta\, dx.$$

Let us apply the Leibnitz rule in the first factor. Note that when all derivatives fall on θ, we obtain

$$\frac{1}{2}\int_{\mathbb{T}^2} (u\cdot\nabla)(\partial_{i_1}\dots\partial_{i_s}\theta)^2\, dx. \tag{4.8}$$

This term is potentially very dangerous – a priori it depends on the derivatives of θ of the order $s+1$. Hence, it would be impossible to control it by the dissipative term $\|\theta\|_{H^{s+\alpha}}$ unless $\alpha > 1$, while we only assume that $\alpha > 1/2$. Fortunately, this term vanishes after integration by parts due to the incompressibility of u, and is therefore not a problem. All other summands can be written in the form

$$\int_{\mathbb{T}^2} (D^l u)(D^{s-l+1})(\theta D^s\theta)\, dx,$$

where each D stands for some partial derivative and l may vary from 1 to s. Let us estimate every integral using the Hölder inequality

$$\left|\int_{\mathbb{T}^2} (D^l u)(D^{s-l+1}\theta)(D^s\theta)\, dx\right| \le \|D^l u\|_{L^{p_l}}\|D^{s-l+1}\theta\|_{L^{q_l}}\|D^s\theta\|_{L^{n_l}}, \tag{4.9}$$

with

$$\frac{1}{p_l} + \frac{1}{q_l} + \frac{1}{n_l} = 1. \tag{4.10}$$

The powers p_l, q_l and n_l will be chosen soon. At the moment we obtain

$$\frac{d}{dt}\|\theta\|_{H^s}^2 \le C\sum_{l=1}^{s} \|D^l u\|_{L^{p_l}}\|D^{s-l+1}\theta\|_{L^{q_l}}\|D^s\theta\|_{L^{n_l}} - \|\theta\|_{H^{s+\alpha}}^2. \tag{4.11}$$

Observe that

$$\|D^l u\|_{L^{p_l}} \leq C\|D^l\theta\|_{L^{p_l}},$$

since $D^l u$ is a Riesz transform of $D^l\theta$, and $1 < p_l < \infty$. Singular integral operators such as the Riesz transform are known to be bounded on L^p spaces with $1 < p < \infty$, see, for example, [109] for the proof. Thus, we actually have

$$\frac{d}{dt}\|\theta\|_{H^s}^2 \leq C\sum_{l=1}^{s}\|D^l\theta\|_{L^{p_l}}\|D^{s-l+1}\theta\|_{L^{q_l}}\|D^s\theta\|_{L^{n_l}} - \|\theta\|_{H^{s+\alpha}}^2. \tag{4.12}$$

Our goal is to beat the first term in the right side of (4.11), which may produce growth by the last term, which comes with the negative sign. To this end, we will need to use the following version of the Gagliardo-Nirenberg inequality.

Theorem 4.7 *Suppose that $f \in C^\infty(\mathbb{T}^d)$ and m is an integer such that $1 \leq m \leq r$, then*

$$\|D^m f\|_{L^{2r/m}} \leq C\|f\|_{L^\infty}^{1-m/r}\|(-\Delta)^{r/2}f\|_{L^2}^{m/r}. \tag{4.13}$$

Here, D^m denotes an arbitrary partial derivative of order m.

The inequality (4.13) is well known; it is contained, for example, in the encyclopedic reference [91]. However we are not aware of the source containing a simple and easily accessible proof. For this reason, we include a sketch of the proof of (4.13) in Section 6.1 at the end of this chapter. Let us just mention here that the inequality (4.13) is dimensionally correct: if x has the dimension of length L, then the left side of (4.13) has the dimension

$$(L^{-m(2r/m)}L^d)^{m/(2r)} = L^{-m+dm/(2r)},$$

and the right side of this inequality has the dimension

$$(L^{-2r}L^d)^{m/(2r)},$$

which is the same. Often, the Gagliardo-Nirenberg inequalities that are dimensionally correct, are actually true, but see Section 6.1 for an actual proof.

The reason we choose this particular Gagliardo-Nirenberg inequality is clear – we have control over $\|\theta\|_{L^\infty}$, so using (4.13) would reduce the overall power of the potentially dangerous cubic term in the right side (4.11). The price in the form of the H^r-Sobolev norm that we need to pay is not bad, as long as $r < s+\alpha$, which is the good term in the right side of (4.11).

We will use Theorem 4.7 to bound each term in the cubic nonlinearity in the right side of (4.12), always using the same r. This will give us, for each $1 \leq l \leq s$:

$$p_l = \frac{2r}{l}, \quad q_l = \frac{2r}{s-l+1}, \quad n_l = \frac{2r}{s}.$$

Condition (4.10) becomes then

$$\frac{l}{2r} + \frac{s-l+1}{2r} + \frac{s}{2r} = 1,$$

or $r = s + 1/2$. It is exactly this little calculation that determines the critical $\alpha = 1/2$ – the dissipation involves the Sobolev norm $H^{s+\alpha}$, and to have any hope to beat the nonlinearity with such dissipation, we need $s + \alpha > r$, meaning $\alpha > 1/2$. With this choice, the cubic term in (4.12), for each $1 \leq l \leq s$, can be estimated using (4.13) as

$$\|D^l u\|_{L^{p_l}} \|D^{s-l+1}\theta\|_{L^{q_l}} \|D^s\theta\|_{L^{n_l}} \leq C\|\theta\|_{L^\infty}^a \|\theta\|_{H^{s+1/2}}^b, \tag{4.14}$$

with

$$a = 1 - \frac{2}{p_l} + 1 - \frac{2}{q_l} + 1 - \frac{2}{n_l} = 1,$$

and

$$b = \frac{2}{p_l} + 2\frac{2}{q_l} + \frac{2}{n_l} = 2,$$

so that

$$\|D^l u\|_{L^{p_l}} \|D^{s-l+1}\theta\|_{L^{q_l}} \|D^s\theta\|_{L^{n_l}} \leq C\|\theta\|_{L^\infty} \|\theta\|_{H^{s+1/2}}^2. \tag{4.15}$$

Putting these estimates into (4.11), together with the L^∞ maximum principle, we get

$$\frac{d}{dt}\|\theta\|_{H^s}^2 \leq C\|\theta\|_{L^\infty}\|\theta\|_{H^{s+1/2}}^2 - \|\theta\|_{H^{s+\alpha}}^2 \leq C\|\theta_0\|_{L^\infty}\|\theta\|_{H^{s+1/2}}^2 - \|\theta\|_{H^{s+\alpha}}^2. \tag{4.16}$$

It remains to observe that, as $\alpha > 1/2$, we may use Hölder's inequality to get

$$\|\theta\|_{H^{s+1/2}}^2 = C\sum_{n\in\mathbb{Z}^2} |n|^{2s+1}|\hat{\theta}_n|^2 \leq C\Big(\sum_{n\in\mathbb{Z}^2} |n|^{(2s+1)p}|\hat{\theta}_n|^{(2-a)p}\Big)^{1/p}\Big(\sum_{n\in\mathbb{Z}^2} |\hat{\theta}_n|^{aq}\Big)^{1/q}, \tag{4.17}$$

for any

$$0 < a < 2 \text{ and } \frac{1}{p} + \frac{1}{q} = 1.$$

As $\alpha > 1/2$, we may choose

$$p = \frac{2(s+\alpha)}{2s+1} > 1,$$

and q and a so that

$$\frac{1}{p} = 1 - \frac{a}{2}, \quad \frac{1}{q} = \frac{a}{2}.$$

With this choice of the parameters, (4.17) becomes

$$\|\theta\|_{H^{s+1/2}}^2 \leq C\|\theta\|_{H^{s+\alpha}}^{(2s+1)/(s+\alpha)}\|\theta\|_{L^2}^{(2\alpha-1)/(s+\alpha)} \leq C\|\theta_0\|_{L^2}^{(2\alpha-1)/(s+\alpha)}\|\theta\|_{H^{s+\alpha}}^{(2s+1)/(s+\alpha)}. \tag{4.18}$$

Thus, (4.16) can be written as a differential inequality

$$\frac{d}{dt}\|\theta\|_{H^s}^2 \leq C_1\|\theta\|_{H^{s+\alpha}}^\beta - C_2\|\theta\|_{H^{s+\alpha}}^2, \tag{4.19}$$

with the constants C_1 and C_2 that depend on the initial data and

$$\beta = \frac{2s+1}{s+\alpha} < 2,$$

for $\alpha > 1/2$. Therefore, there exists $M > 0$ which depends on C_1 and C_2 so that

$$\frac{d}{dt}\|\theta(t)\|_{H^s} \leq 0 \quad \text{if } \|\theta(t)\|_{H^{s+\alpha}} > M. \tag{4.20}$$

As θ has mean zero, we also know that

$$\|\theta\|_{H^{s+\alpha}} \geq \|\theta\|_{H^s}.$$

Therefore, we know from (4.20) that $\|\theta(t)\|_{H^s}$ is decreasing if $\|\theta(t)\|_{H^s} > M$. These two observations imply that $\|\theta\|_{H^s}$ is bounded globally in time if $\alpha > 1/2$. The argument clearly fails if $\alpha = 1/2$. □

Exercise 4.8 We have used the same value $r = s + 1/2$ above, as we applied the Gagliardo-Nirenberg inequality to estimate each of $\|D^l u\|_{L^{p_l}}$, $\|D^{s-l+1}\theta\|_{L^{q_l}}$ and $\|D^s\theta\|_{L^{n_l}}$. Show that using a different $r_{1,2,3}$ for each of these terms, would not allow us to improve the value of the critical $\alpha = 1/2$.

5 The regularity of the critical SQG equation

For the critical case $\alpha = 1/2$, we need new ideas. The critical SQG equation has the form

$$\begin{aligned} &\theta_t + u \cdot \nabla\theta = -(-\Delta)^{1/2}\theta, \\ &u = \nabla^{\perp}(-\Delta)^{-1/2}\theta, \\ &\theta(0,x) = \theta_0(x). \end{aligned} \tag{5.1}$$

The problem of the regularity of its solutions has been solved a few years ago independently by Kiselev, Nazarov and Volberg [77], and by Caffarelli and Vasseur [29] using very different methods. By now, there are several more proofs [34, 35, 79]. The proof of Constantin, Tarfulea and Vicol [35] shows most clearly how dissipation controls the nonlinearity. We will follow the original proof in [77], which is less explicit but elegant.

Theorem 5.1 *The critical surface quasi-geostrophic equation with periodic smooth initial data $\theta_0(x)$ has a unique global in time smooth solution. Moreover, the following estimate holds for all times:*

$$\|\nabla\theta(t,\cdot)\|_\infty \leq C\|\nabla\theta_0\|_\infty \exp\exp\{C\|\theta_0\|_\infty\}\,. \tag{5.2}$$

In fact, the key to the proof of Theorem 5.1 is the a priori estimate (5.2), as the following proposition implies.

Proposition 5.2 *Assume that $\theta(t,x)$ is a smooth, local in time solution to the critical SQG equation with smooth initial data. Suppose that $T > 0$ is the first time such that some Sobolev norm of the solution blows up: $\|\theta(t,\cdot)\|_{H^s} \to \infty$ as $t \to T$. Then we also have*

$$\lim_{t\to T}\|\nabla\theta(x,t)\|_{L^\infty} \to \infty.$$

Exercise 5.3 Prove Proposition 5.2 using the following strategy. Suppose, for the sake of a contradiction, that some Sobolev norm of the solution blows up at $t = T$ while $\|\nabla\theta(t)\|_{L^\infty}$ remains bounded. We can follow the proof of Theorem 4.6 but now we have a much better control of the function $\theta(t,x)$ because of the a priori information that $\|\nabla\theta(t)\|_{L^\infty}$ is uniformly bounded in time. In particular, we can base the Gagilardo-Nirenberg inequalities on $\|\nabla\theta\|_{L^\infty}$ instead of $\|\theta\|_{L^\infty}$. This will give a lower power of $\|\theta\|_{s+1/2}$ in the estimate of the nonlinearity than in the dissipative term even for $\alpha = 1/2$, leading to a differential inequality for $\|\theta\|_{H^s}$ that will show that $\|\theta\|_{H^s}$ is decreasing if it is large. Thus, if $\|\nabla\theta\|_{L^\infty} \le C < \infty$ on $[0,T]$, then we have a bound on any H^s norm of $\theta(x,t)$ preventing it from going to infinity at $t = T$. Fill in the (technical) details.

We note that sharper results are available in the literature. The L^∞ norm is barely not enough to handle the critical case, so the control of any positive Hölder norm of the solution should suffice to prove the global in time regularity. Indeed, it has been shown by Constantin and Wu [38] that controlling C^β, $\beta > 1 - 2\alpha$, norm of the solution is sufficient to prove regularity for the supercritical SQG equation with $(-\Delta)^\alpha$ dissipation.

The modulus of continuity

The main idea of the proof will be to try to estimate a certain well chosen modulus of continuity of the solution.

Definition 5.4 *A function $\omega : \mathbb{R}^+ \mapsto \mathbb{R}^+$ is a modulus of continuity if $\omega(0) = 0$, ω is continuous, increasing and concave. We will also require that ω is piecewise C^1 on $(0,\infty)$. That is, its derivative is continuous, apart from possibly a finite number of points where the one-sided derivatives exist but may not be equal. A function f obeys a modulus of continuity ω if*

$$|f(x) - f(y)| < \omega(|x-y|) \text{ for all } x \neq y.$$

We say that an evolution equation for $\theta(t,x)$ preserves ω if $\theta(t,x)$ obeys a modulus of continuity ω for all times $t > 0$ provided that the initial data $\theta_0(x)$ obeys ω.

A classical example of a modulus of continuity is $\omega(\xi) = \xi^\beta$, $0 < \beta < 1$, corresponding to functions of the Hölder classes.

The flow term $u \cdot \nabla\theta$ in the dissipative quasi-geostrophic equation can potentially make the modulus of continuity of θ worse while the dissipation term $(-\Delta)^\alpha\theta$ tends to make it better. Our aim is to construct some special moduli of continuity for which the dissipation term always prevails, and such that every periodic C^∞-function θ_0 obeys one of these special moduli of continuity.

The critical ($\alpha = 1/2$) SQG equation has a simple scaling invariance: if $\theta(t,x)$ is a solution, then so is $\theta(Bt, Bx)$, for any $B > 0$. This means that if we manage to find one modulus of continuity ω that is preserved by the evolution for all periodic solutions (that is, with arbitrary lengths and spacial orientations of the periods), then the whole family

$$\omega_B(\xi) = \omega(B\xi)$$

of moduli of continuity will also be preserved for all periodic solutions.

Note that if a modulus of continuity ω is unbounded, then any given C^∞ periodic function has the modulus of continuity ω_B for a sufficiently large $B > 0$. Also, if the modulus of continuity ω has a finite derivative at 0 and θ obeys ω, then $\|\nabla\theta\|_\infty \le \omega'(0)$. Thus, due to Proposition 5.2, our task in the proof of global in time regularity of the solutions reduces to constructing an unbounded modulus of continuity with a finite derivative at 0 that is preserved by the critical SQG evolution.

Exercise 5.5 Given an unbounded modulus of continuity ω and smooth periodic initial data $\theta_0(x)$, find explicitly B_0 such that for $B > B_0$, the function θ_0 obeys ω_B. The constant B_0 should only depend on ω, $\|\theta_0\|_{L^\infty}$, and $\|\nabla\theta_0\|_{L^\infty}$.

From now on, we will also assume that in addition to unboundedness and the condition $\omega'(0) < +\infty$ our modulus of continuity will satisfy

$$\lim_{\xi\to 0+} \omega''(\xi) = -\infty.$$

This assumption is of a purely technical nature but will be very useful for us. One simple observation is the following

Lemma 5.6 *If a smooth periodic function f obeys a modulus of continuity ω with*

$$\omega'(0) < +\infty \ \textit{and} \ \lim_{\xi\to 0+} \omega''(\xi) = -\infty, \tag{5.3}$$

then

$$\|\nabla f\|_\infty < \omega'(0)\,. \tag{5.4}$$

The key here is the strict inequality: the less or equal bound in (5.4) follows easily without the extra assumptions (5.3) on $\omega''(\xi)$.

Proof. Indeed, take a point $x \in \mathbb{R}^d$ at which $\max|\nabla f|$ is attained and consider the point

$$y = x + \xi e,$$

where

$$e = \frac{\nabla f}{|\nabla f|}.$$

Then we must have

$$f(y) - f(x) \le \omega(\xi), \quad \text{for all } \xi \ge 0. \tag{5.5}$$

The left side above is at least

$$f(y) - f(x) \ge |\nabla f(x)|\xi - \frac{\|\nabla^2 f\|_\infty}{2}\xi^2,$$

while the right side in (5.5) can be represented as

$$\omega(\xi) = \omega'(0)\xi - \rho(\xi)\xi^2,$$

with $\rho(\xi) \to +\infty$ as $\xi \to 0+$. Thus, we have

$$|\nabla f(x)| \le \omega'(0) - (\rho(\xi) - \frac{\|\nabla^2 f\|_\infty}{2})\xi \text{ for all } \xi > 0,$$

and it remains to choose small enough $\xi > 0$ ensuring $\rho(\xi) > |\nabla^2 f\|_\infty/2$, showing that (5.4) holds, with a strict inequality. □

Modulus of continuity of the velocity

Before we begin the analysis of the SQG dynamics, we link the moduli of continuity obeyed by θ and by u.

Lemma 5.7 *If the function f obeys a modulus of continuity ω, then*

$$u = (\partial_2(-\Delta)^{-1/2} f, -\partial_1(-\Delta)^{1/2} f)$$

has the modulus of continuity

$$\Omega(\xi) = A\left(\int_0^\xi \frac{\omega(\eta)}{\eta}\, d\eta + \xi \int_\xi^\infty \frac{\omega(\eta)}{\eta^2}\, d\eta\right), \tag{5.6}$$

with some universal constant $A > 0$.

Observe that for $\omega(\xi)$ that is Hölder near zero, $\omega(\xi) = C\xi^\alpha$ with $0 < \alpha < 1$, the modulus of continuity $\Omega(\xi)$ given by (5.6) has the Hölder behavior with the same exponent. On the other hand, for $\omega(\xi)$ that is Lipschitz near zero, we get an extra logarithm in $\Omega(\xi)$, coming from the second term in (5.6). This is an illustration of a well known fact (see e.g. [109]) that the Riesz transforms are bounded on C^α but not on the space of Lipschitz functions.

Proof of Lemma 5.7. The Riesz transforms

$$R_{1,2} = \partial_{1,2}(-\Delta)^{-1/2}$$

are singular integral operators with explicit kernels of the form

$$K(r, \zeta) = r^{-2}\Omega(\zeta),$$

where (r, ζ) are the polar coordinates (see (4.6)). The function Ω is smooth and

$$\int_{S^1} \Omega(\zeta) d\sigma(\zeta) = 0. \tag{5.7}$$

Assume that the function f satisfies

$$|f(x) - f(y)| < \omega(|x - y|)$$

for some modulus of continuity ω. Take any x, y with $|x - y| = \xi$, and consider the difference

$$P.V. \int K(x - z) f(z)\, dz - P.V. \int K(y - z) f(z)\, dz. \tag{5.8}$$

The mean zero property (5.7) of Ω allows us to write

$$\left|P.V. \int_{|x-z|\le 2\xi} K(x - z) f(z)\, dz\right| = \left|P.V. \int_{|x-z|\le 2\xi} K(x - z)(f(z) - f(x))\, dz\right| < C \int_0^{2\xi} \frac{\omega(r)}{r}\, dr.$$

Since ω is concave and $\omega(0) = 0$, the function $\omega(r)/r$ is decreasing, thus we have

$$\int_0^{2\xi} \frac{\omega(r)}{r}\, dr \le 2 \int_0^\xi \frac{\omega(r)}{r}\, dr.$$

A similar estimate holds for the second integral in (5.8) for the region $|y - z| \le 2\xi$. Next, let $\widetilde{x} = (x + y)/2$ and note that if $|x - z| > 2\xi$, then

$$|\widetilde{x} - z| \ge |x - z| - |\widetilde{x} - x| \ge \frac{3\xi}{2},$$

and similarly for z such that $|y - z| \ge 2\xi$. In addition, whenever $|\widetilde{x} - z| > 3\xi$, we have both

$$|x - z| \ge 2\xi, \text{ and } |y - z| \ge 2\xi.$$

Thus, we can write, once again using (5.7):

$$\begin{aligned}
&\Big|\int_{|x-z|\ge 2\xi} K(x-z)f(z)\,dz - \int_{|y-z|\ge 2\xi} K(y-z)f(z)\,dz\Big| \\
&= \Big|\int_{|x-z|\ge 2\xi} K(x-z)(f(z)-f(\widetilde{x}))\,dz - \int_{|y-z|\ge 2\xi} K(y-z)(f(z)-f(\widetilde{x}))\,dz\Big| \\
&\le \int_{|\widetilde{x}-z|\ge 3\xi} |K(x-z)-K(y-z)||f(z)-f(\widetilde{x})|\,dz \\
&+ \int_{3\xi/2\le|\widetilde{x}-z|\le 3\xi} (|K(x-z)|+|K(y-z)|)|f(z)-f(\widetilde{x})|\,dz = I + II.
\end{aligned}$$

Since

$$|K(x-z) - K(y-z)| \le C\frac{|x-y|}{|\widetilde{x}-z|^3}$$

when $|\widetilde{x} - z| \ge 3\xi$, the first integral is estimated by

$$I \le C\xi \int_{3\xi}^{\infty} \frac{\omega(r)}{r^2}\,dr.$$

The second integral is estimated by

$$II \le C\omega(3\xi)\int_{3\xi/2}^{3\xi} \frac{r dr}{\xi^2} \le C\omega(3\xi).$$

Once again, as the function ω is concave and $\omega(0) = 0$, we have $\omega'(r) \le \omega(r)/r$, thus

$$II \le 3C\int_0^{3\xi} \frac{\omega(r)}{r}\,dr \le 9C\int_0^{\xi} \frac{\omega(r)}{r}\,dr.$$

This completes the proof of Lemma 5.7. □

Breaking a modulus of continuity

A breakthrough scenario

Next, we determine what must happen for the solution $\theta(t, x)$ of the critical SQG equation to break a modulus ω that θ_0 obeys.

Lemma 5.8 (The breakthrough lemma) *Suppose that the smooth periodic initial condition θ_0 obeys the modulus of continuity ω, but the solution $\theta(T, x)$ no longer obeys it at some $T > 0$. Then, there must exist a time t_1 such that for all $t < t_1$, $\theta(t, x)$ obeys ω, while at $t = t_1$ there exist $x \neq y$ such that*

$$\theta(t_1, x) - \theta(t_1, y) = \omega(|x - y|).$$

Proof. Suppose that θ obeys the modulus of continuity ω at a time $t_0 \geq 0$:

$$|\theta(t_0, x) - \theta(t_0, y)| < \omega(|x - y|) \text{ for all } x \neq y.$$

We claim that then θ obeys then modulus of continuity ω for all $t > t_0$ sufficiently close to t_0. Indeed, by Lemma 5.6, at the moment t_0 we have

$$\|\nabla\theta\|_\infty < \omega'(0).$$

By the continuity of the derivatives, this inequality also holds for $t > t_0$ close to t_0, which guarantees the inequality

$$|\theta(t, x) - \theta(t, y)| < \omega(|x - y|) \text{ for small } |x - y|.$$

Also, since ω is unbounded and $\|\theta\|_\infty$ doesn't grow with time, we automatically have

$$|\theta(t, x) - \theta(t, y)| < \omega(|x - y|) \text{ for large } |x - y|.$$

The final observation is that, due to the periodicity of θ, for each $y \in \mathbb{T}^2$ fixed, it suffices to check the inequality

$$|\theta(t, x) - \theta(t, y)| < \omega(|x - y|)$$

for x in a compact set $K \subset \mathbb{R}^2$. Thus, we are left with the task to show that, if

$$|\theta(t_0, x) - \theta(t_0, y)| < \omega(|x - y|) \text{ for all } x \in K,\ \delta \leq |x - y| \leq \delta^{-1},$$

with some fixed $\delta > 0$, then the same inequality remains true for a short time beyond t_0. But this immediately follows from the uniform continuity of θ. This proves that the set S of t for which θ obeys ω is open. As $\theta(T, x)$ does not obey ω, the set of such times is bounded from above. Consider its supremum t_1. By continuity, we must have

$$|\theta(t_1, x) - \theta(t_1, x)| \leq \omega(|x - y|) \text{ for all } x \neq y.$$

On the other hand, as the set S is open, $t_1 \notin S$,hence the function θ does not obey ω at t_1. Thus, we can find $x \neq y$ as in the statement of the lemma. □

An unbreakable modulus of continuity

The main idea now is to construct a modulus of continuity for which the breakthrough scenario is impossible. Let ω be a modulus of continuity obeyed by θ_0 and let t_1 be the first time when the modulus is touched: there exist some "breakthrough points" $x, y \in \mathbb{T}^2$ so that

$$\theta(t_1, x) - \theta(t_1, y) = \omega(|x - y|). \tag{5.9}$$

We are going to choose ω so that if (5.9) holds but

$$|\theta(t,x_1)-\theta(t,x_2)|\le\omega(|x_1-x_2|),\ \text{for all } 0\le t\le t_1 \text{ and all } x_1,x_2\in\mathbb{T}^2, \tag{5.10}$$

then

$$\partial_t(\theta(x,t)-\theta(y,t))\big|_{t_1}<0, \tag{5.11}$$

which is a contradiction. Therefore, for this choice of ω we must have

$$|\theta(t,x)-\theta(t,y)|<\omega(|x-y|), \tag{5.12}$$

if

$$|\theta_0(x)-\theta_0(y)|<\omega(|x-y|). \tag{5.13}$$

In addition, we will choose ω so that the above property holds not only for ω itself but also for its rescaled versions $\omega_B(\xi)=\omega(B\xi)$. As we have mentioned, given any initial data we are able to find $B>0$ so that

$$|\theta_0(x)-\theta_0(y)|<\omega_B(|x-y|). \tag{5.14}$$

As a consequence, we will have

$$|\theta(t,x)-\theta(t,y)|<\omega(|x-y|). \tag{5.15}$$

It will follow then that

$$\|\nabla\theta(t)\|_{L^\infty}\le\omega_B'(0),\qquad \text{for all } t\ge 0,$$

providing an a priori L^∞-bound on $\theta(t,x)$. Proposition 5.2 will then imply that $\theta(t,x)$ is regular for all $t\ge 0$.

We start the search for such ω by computing

$$\begin{aligned}\partial_t(\theta(t,x)-\theta(t,y))\big|_{t=t_1}&=-(u\cdot\nabla\theta)(t_1,x)+(u\cdot\nabla\theta)(t_1,y)\\&\quad-(-\Delta)^{1/2}\theta(t_1,x)+(-\Delta)^{1/2}\theta(t_1,y).\end{aligned} \tag{5.16}$$

Here, x and y are the breakthrough points, and t_1 is the breakthrough time. The two terms in the right side of (5.15) will be estimated separately.

An estimate for the flow term

We first look at the flow term in the right side of (5.16) and estimate how dangerous it is. Observe that

$$(u\cdot\nabla\theta)(t_1,x)=\frac{d}{dh}\theta(t_1,x+hu(t_1,x))\big|_{h=0},$$

and similarly for y, so that

$$\begin{aligned}&(u\cdot\nabla\theta)(t_1,x)-(u\cdot\nabla\theta)(t_1,y)\\&=\lim_{h\to 0}\frac{1}{h}(\theta(t_1,x+hu(t_1,x))-\theta(t_1,x)-\theta(t_1,y+hu(t_1,y))+\theta(t_1,y)).\end{aligned}$$

Lemma 5.7 implies that

$$\begin{aligned}|\theta(t_1,x+hu(t_1,x))-\theta(t_1,y+hu(t_1,y))|&\le\omega(|x-y|+h|u(t_1,x)-u(t_1,y)|)\\&\le\omega(\xi+h\Omega(\xi)),\end{aligned}$$

where, as in (5.6),

$$\Omega(\xi) = A\left(\int_0^\xi \frac{\omega(\eta)}{\eta}\,d\eta + \xi\int_\xi^\infty \frac{\omega(\eta)}{\eta^2}\,d\eta\right).$$

We have set here $\xi = |x-y|$. Since

$$\theta(t_1,x) - \theta(t_1,y) = \omega(\xi),$$

we conclude that

$$(u\cdot\nabla\theta)(t_1,x) - (u\cdot\nabla\theta)(t_1,y) \le \lim_{h\to 0}\frac{1}{h}\left(\omega(\xi+h\Omega(\xi)) - \omega(\xi)\right) \le \Omega(\xi)\omega'(\xi). \tag{5.17}$$

Switching the role of x and y, we obtain

$$|(u\cdot\nabla\theta)(t_1,x) - (u\cdot\nabla\theta)(t_1,y)| \le \Omega(\xi)\omega'(\xi). \tag{5.18}$$

An estimate for the dissipative term

According to Lemma 2.2, the fractional Laplacian of a smooth function θ can be represented as

$$-(-\Delta)^{1/2}\theta(x) = \lim_{h\to 0}\frac{1}{h}\left(\mathcal{P}_h * \theta(x) - \theta(x)\right), \tag{5.19}$$

where $\mathcal{P}_h(x)$ is the 2-dimensional Poisson kernel

$$\mathcal{P}_h(x) = c_2 h(|x|^2 + h^2)^{-3/2},$$

with an appropriately chosen constant c_2 to ensure that $P_h(x) \to \delta(x)$ as $h \to 0$, in the sense of distributions. The formula (5.19) is easy to verify on the Fourier side. In fact, it is straightforward to use the Fourier transform to compute the Poisson kernel in any dimension:

$$\mathcal{P}_h^d(x) = c_d h(|x|^2 + h^2)^{-\frac{d+1}{2}}. \tag{5.20}$$

The computation below will use the following dimensional reduction formula:

$$\int_{\mathbb{R}} \mathcal{P}_h^2(x_1,x_2)\,dx_2 = \mathcal{P}_h^1(x_1). \tag{5.21}$$

These formulae hold for all smooth periodic functions regardless of the lengths and spatial orientation of the periods, which allows us to freely use the scaling, translation and rotation.

Exercise 5.9 Compute explicitly, with the help of the Fourier inversion that

$$\mathcal{P}_h^1(x) = \frac{h}{\pi(x^2+h^2)}, \tag{5.22}$$

and the relation (5.21). Finally, verify the formula (5.20) for the Poisson kernel in a general dimension.

Thus, our task is to estimate

$$(\mathcal{P}_h * \theta)(x) - (\mathcal{P}_h * \theta)(y),$$

under the assumption that θ obeys a modulus of continuity ω. Since everything is translationally and rotationally invariant, we may assume that $x = (\xi/2, 0)$ and $y = (-\xi/2, 0)$. We will show, after a somewhat lengthy computation, that

$$\begin{aligned}(\mathcal{P}_h * \theta)(x) - (\mathcal{P}_h * \theta)(y) - \omega(|x-y|) \leq \frac{1}{\pi}\int_0^{\xi/2} \frac{\omega(\xi+2\eta)+\omega(\xi-2\eta)-2\omega(\xi)}{\eta^2}\,d\eta \\ + \frac{1}{\pi}\int_{\xi/2}^{\infty} \frac{\omega(2\eta+\xi)-\omega(2\eta-\xi)-2\omega(\xi)}{\eta^2}\,d\eta.\end{aligned} \tag{5.23}$$

Then, taking into account that x and y are such that

$$\theta(t_1, x) - \theta(t_1, y) = \omega(|x-y|),$$

together with (5.19), the corresponding term in (5.16) can be estimated as

$$\begin{aligned}&-(-\Delta)^{1/2}\theta(t_1,x) + (-\Delta)^{1/2}\theta(t_1,y) \\ &= \lim_{h\to 0}\frac{1}{h}\left(\mathcal{P}_h * \theta(t_1,x) - \mathcal{P}_h * \theta(t_1,x) - (\theta(t_1,x) - \theta(t_1,y))\right) \\ &\leq \frac{1}{\pi}\int_0^{\xi/2} \frac{\omega(\xi+2\eta)+\omega(\xi-2\eta)-2\omega(\xi)}{\eta^2}\,d\eta \\ &+\frac{1}{\pi}\int_{\xi/2}^{\infty} \frac{\omega(2\eta+\xi)-\omega(2\eta-\xi)-2\omega(\xi)}{\eta^2}\,d\eta.\end{aligned} \tag{5.24}$$

Note that both terms in the right side of (5.24) are strictly negative. For the first term, this follows immediately from the concavity of ω:

$$\frac{1}{2}[\omega(\xi+2\eta)+\omega(\xi-2\eta)] \leq \omega(\xi).$$

For the second term, note that

$$\omega(2\eta+\xi) - \omega(2\eta-\xi) = 2\omega'(\zeta)\xi,$$

with some $\zeta \in (2\eta-\xi, 2\eta+\xi)$. In the domain of integration of the second term in the right side of (5.23), we have $\eta > \xi/2$, thus

$$\zeta > 2\eta - \xi > \xi.$$

Concavity of ω implies then that

$$\omega'(\zeta) < \omega'(\xi) \leq \frac{\omega(\xi)}{\xi},$$

and thus

$$\omega(2\eta+\xi) - \omega(2\eta-\xi) = 2\omega'(\zeta)\xi \leq 2\omega(\xi). \tag{5.25}$$

Therefore, both terms in the right side of (5.23) are negative, so that the dissipation may help us to ensure that (5.11) holds, that is,

$$\partial_t(\theta(x,t)-\theta(y,t))\big|_{t_1}<0, \tag{5.26}$$

if

$$\begin{aligned}&|\theta(t,x_1)-\theta(t,x_2)|\leq\omega(|x_1-x_2|),\ \text{for all } 0\leq t\leq t_1 \text{ and all } x_1,x_2\in\mathbb{T}^2,\\ &\text{and } \theta(t_1,x)-\theta(t_1,y)=\omega(|x_1-x_2|),\end{aligned} \tag{5.27}$$

which is our ultimate goal.

To start the computation leading to (5.23), let us write (omitting the time dependence)

$$\begin{aligned}(\mathcal{P}_h*\theta)(x)-(\mathcal{P}_h*\theta)(y)&=\int_{\mathbb{R}^2}[\mathcal{P}_h(\frac{\xi}{2}-\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}-\eta,\nu)]\theta(\eta,\nu)\,d\eta d\nu\\ &=\int_{\mathbb{R}}d\nu\int_0^\infty[\mathcal{P}_h(\frac{\xi}{2}-\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}-\eta,\nu)]\theta(\eta,\nu)d\eta\\ &+\int_{\mathbb{R}}d\nu\int_0^\infty[\mathcal{P}_h(\frac{\xi}{2}+\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}+\eta,\nu)]\theta(-\eta,\nu)\,d\eta.\end{aligned}$$

As $\mathcal{P}_h(x_1,x_2)$ is even in x_1, we may re-write the last expression as

$$(\mathcal{P}_h*\theta)(x)-(\mathcal{P}_h*\theta)(y)=\int_{\mathbb{R}}d\nu\int_0^\infty[\mathcal{P}_h(\frac{\xi}{2}+\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}+\eta,\nu)](\theta(\eta,\nu)-\theta(-\eta,\nu)\,d\eta.$$

Since $\xi>0$, $\eta>0$, and, once again, $\mathcal{P}_h(x_1,x_2)$ is even in x_1, and monotonically decreasing in $|x_1|$, we have

$$\mathcal{P}_h(\frac{\xi}{2}+\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}+\eta,\nu)\geq 0,$$

and thus

$$(\mathcal{P}_h*\theta)(x)-(\mathcal{P}_h*\theta)(y)\leq\int_{\mathbb{R}}d\nu\int_0^\infty[\mathcal{P}_h(\frac{\xi}{2}-\eta,\nu)-\mathcal{P}_h(-\frac{\xi}{2}-\eta,\nu)]\omega(2\eta)\,d\eta.$$

Now, we may use the dimension reduction formula (5.21) to integrate out the ν-variable:

$$(\mathcal{P}_h*\theta)(x)-(\mathcal{P}_h*\theta)(y)\leq\int_0^\infty[\mathcal{P}_h^1(\frac{\xi}{2}-\eta)-\mathcal{P}_h^1(-\frac{\xi}{2}-\eta)]\omega(2\eta)\,d\eta.$$

Using the symmetry of $\mathcal{P}_h^1(x_1)$ one more time, this becomes

$$(\mathcal{P}_h*\theta)(x)-(\mathcal{P}_h*\theta)(y)\leq\int_0^\xi\mathcal{P}_h^1(\frac{\xi}{2}-\eta)\omega(2\eta)\,d\eta+\int_0^\infty\mathcal{P}_h^1(\frac{\xi}{2}+\eta)[\omega(2\eta+2\xi)-\omega(2\eta)]\,d\eta. \tag{5.28}$$

The same symmetry means that the first integral above can be re-written as

$$\int_0^\xi\mathcal{P}_h^1(\frac{\xi}{2}-\eta)\omega(2\eta)\,d\eta=\int_{-\xi/2}^{\xi/2}\mathcal{P}_h^1(\eta)\omega(\xi-2\eta)\,d\eta=\int_0^{\xi/2}\mathcal{P}_h^1(\eta)[\omega(\xi+2\eta)+\omega(\xi-2\eta)]\,d\eta.$$

Thus, (5.29) becomes

$$(\mathcal{P}_h * \theta)(x) - (\mathcal{P}_h * \theta)(y) \le \int_0^{\xi/2} \mathcal{P}_h^1(\eta)[\omega(\xi + 2\eta) + \omega(\xi - 2\eta)]\, d\eta \qquad (5.29)$$
$$+ \int_{\xi/2}^{\infty} \mathcal{P}_h^1(\eta)[\omega(2\eta + \xi) - \omega(2\eta - \xi)]\, d\eta\,.$$

As

$$\int_0^\infty \mathcal{P}_h^1(\eta)\, d\eta = \frac{1}{2},$$

we conclude that

$$(\mathcal{P}_h * \theta)(x) - (\mathcal{P}_h * \theta)(y) - \omega(\xi) \le \int_0^{\xi/2} \mathcal{P}_h^1(\eta)[\omega(\xi + 2\eta) + \omega(\xi - 2\eta) - 2\omega(\xi)]\, d\eta$$
$$+ \int_{\xi/2}^{\infty} \mathcal{P}_h^1(\eta)[\omega(2\eta + \xi) - \omega(2\eta - \xi) - 2\omega(\xi)]\, d\eta\,.$$

Using the explicit formula (5.23) for $\mathcal{P}_h^1$, dividing by h and passing to the limit $h \to 0+$, we finally conclude that the contribution of the dissipative term to our derivative is bounded from above by

$$(\mathcal{P}_h * \theta)(x) - (\mathcal{P}_h * \theta)(y) - \omega(|x-y|) \le \frac{1}{\pi} \int_0^{\xi/2} \frac{\omega(\xi + 2\eta) + \omega(\xi - 2\eta) - 2\omega(\xi)}{\eta^2}\, d\eta$$
$$+ \frac{1}{\pi} \int_{\xi/2}^{\infty} \frac{\omega(2\eta + \xi) - \omega(2\eta - \xi) - 2\omega(\xi)}{\eta^2}\, d\eta \qquad (5.30)$$

which is (5.23).

The choice of the modulus of continuity

Summarizing the estimates (5.18) and (5.24) for the fluid and dissipation terms, respectfully, in (5.16), we have so far shown that for any concave modulus of continuity we have

$$\partial_t(\theta(t,x) - \theta(t,y))|_{t=t_1} \le \Omega(\xi)\omega'(\xi) + \frac{1}{\pi} \int_0^{\xi/2} \frac{\omega(\xi + 2\eta) + \omega(\xi - 2\eta) - 2\omega(\xi)}{\eta^2}\, d\eta$$
$$+ \frac{1}{\pi} \int_{\xi/2}^{\infty} \frac{\omega(2\eta + \xi) - \omega(2\eta - \xi) - 2\omega(\xi)}{\eta^2}\, d\eta. \qquad (5.31)$$

Recall that x and y realize the modulus of continuity at the time t_1:

$$\theta(t_1, x) - \theta(t_1, y) = \omega(|x - y|),$$

and the function $\Omega(\xi)$ is given by (5.6):

$$\Omega(\xi) = A\left(\int_0^\xi \frac{\omega(\eta)}{\eta}\, d\eta + \xi \int_\xi^\infty \frac{\omega(\eta)}{\eta^2}\, d\eta\right), \qquad (5.32)$$

As we have mentioned, we will construct a special modulus of continuity so as to make the right side of (5.31) negative. This will contradict the assumption that t_1 is the first time when a pair of point realizes the modulus of continuity ω, and will prove the claim that this modulus of continuity survives forever, so that the solution of the critical SQG equation is regular globally in time.

To define the modulus of continuity, consider two small positive numbers $\delta > \gamma > 0$ to be determined later and define the continuous function ω by

$$\omega(\xi) = \xi - \xi^{\frac{3}{2}} \qquad \text{when } 0 \le \xi \le \delta$$

and

$$\omega'(\xi) = \frac{\gamma}{\xi(4 + \log(\xi/\delta))} \qquad \text{when } \xi > \delta\,.$$

Note that, for small δ, the left derivative of ω at δ is

$$\omega'(\delta^-) = 1 - \frac{3}{2}\delta^{1/2},$$

while the right derivative equals

$$\omega'(\delta^+) = \frac{\gamma}{4\delta} < \frac{1}{4} < \omega'(\delta^-).$$

Therefore, the function ω is concave if δ is small enough. It is clear that

$$\omega'(0) = 1, \qquad \lim_{\xi \to 0+} \omega''(\xi) = -\infty,$$

and that ω is unbounded (it grows at infinity as a double logarithm). The hard part, of course, is to show that, for this ω, the negative contribution to the right side of (5.31) coming from the dissipative term prevails over the positive contribution coming from the flow term. More precisely, we have to check the inequality

$$\begin{aligned} A\left[\int_0^\xi \frac{\omega(\eta)}{\eta}\,d\eta + \xi\int_\xi^\infty \frac{\omega(\eta)}{\eta^2}\,d\eta\right]\omega'(\xi) + \frac{1}{\pi}\int_0^{\xi/2} \frac{\omega(\xi+2\eta) + \omega(\xi-2\eta) - 2\omega(\xi)}{\eta^2}\,d\eta \\ + \frac{1}{\pi}\int_{\xi/2}^\infty \frac{\omega(2\eta+\xi) - \omega(2\eta-\xi) - 2\omega(\xi)}{\eta^2}\,d\eta < 0 \qquad \text{for all } \xi > 0\,. \end{aligned} \tag{5.33}$$

As we have mentioned, this inequality together with (5.31) would imply that

$$\partial_t(\theta(x,t) - \theta(y,t))|_{t_1} < 0, \tag{5.34}$$

if

$$\begin{aligned} &|\theta(t,x_1) - \theta(t,x_2)| \le \omega(|x_1 - x_2|), \text{ for all } 0 \le t \le t_1 \text{ and all } x_1, x_2 \in \mathbb{T}^2, \\ &\text{and } \theta(t_1,x) - \theta(t_1,y) = \omega(|x_1 - x_2|). \end{aligned} \tag{5.35}$$

This is a maximum principle for the modulus continuity, ensuring that $\theta(t,x)$ obeys ω for all $t \ge 0$ if θ_0 obeys ω, and that would finish the proof of Theorem 5.1.

Chcking the inequality (5.33)

We will refer to

$$P = A\left[\int_0^{\xi}\frac{\omega(\eta)}{\eta}\,d\eta + \xi\int_{\xi}^{\infty}\frac{\omega(\eta)}{\eta^2}\,d\eta\right]\omega'(\xi) > 0,$$

as the positive part of the left side of (5.33), and to

$$N = \frac{1}{\pi}\int_0^{\xi/2}\frac{\omega(\xi+2\eta)+\omega(\xi-2\eta)-2\omega(\xi)}{\eta^2}\,d\eta$$
$$+\frac{1}{\pi}\int_{\xi/2}^{\infty}\frac{\omega(2\eta+\xi)-\omega(2\eta-\xi)-2\omega(\xi)}{\eta^2}\,d\eta = N_1+N_2$$

as the negative part. Recall that the integrands both in N_1 and N_2 are negative because of the concavity of ω.

Let is first assume that $0\le\xi\le\delta$. Since $\omega(\eta)\le\eta$ for all $\eta\ge 0$, we have

$$\int_0^{\xi}\frac{\omega(\eta)}{\eta}\,d\eta\le\xi,$$

and

$$\int_{\xi}^{\delta}\frac{\omega(\eta)}{\eta^2}\,d\eta\le\log\frac{\delta}{\xi}.$$

We also have

$$\int_{\delta}^{\infty}\frac{\omega(\eta)}{\eta^2}\,d\eta = \frac{\omega(\delta)}{\delta}+\gamma\int_{\delta}^{\infty}\frac{1}{\eta^2(4+\log(\eta/\delta))}\,d\eta\le 1+\frac{\gamma}{4\delta}<2\,.$$

Observing that $\omega'(\xi)\le 1$, we conclude that the positive part is bounded by

$$P\le A\xi(3+\log\frac{\delta}{\xi}).$$

To estimate the negative part, we just use N_1. Note that

$$\omega(\xi+2\eta)\le\omega(\xi)+2\omega'(\xi)\eta$$

due to the concavity of ω, and

$$\omega(\xi-2\eta)\le\omega(\xi)-2\omega'(\xi)\eta-2\omega''(\xi)\eta^2,$$

due to the second order Taylor formula and monotonicity of ω'' on $[0,\xi]$. Inserting these inequalities into the integral in N_1, we get the bound

$$N_1\le\frac{1}{\pi}\int_0^{\xi/2}\frac{\omega(\xi+2\eta)+\omega(\xi-2\eta)-2\omega(\xi)}{\eta^2}\,d\eta\le\frac{1}{\pi}\xi\omega''(\xi) = -\frac{3}{4\pi}\xi^{1/2}\,.$$

Therefore, we have

$$P+N\le P+N_1\le A\xi(3+\log\frac{\delta}{\xi})-\frac{3}{4\pi}\xi^{1/2}<0,\ \text{for }\xi\in(0,\delta],$$

if δ is small enough.

Finally, let us assume that $\xi \geq \delta$. In this case, we use the bounds

$$\omega(\eta) \leq \eta \text{ for } 0 \leq \eta \leq \delta,$$

and

$$\omega(\eta) \leq \omega(\xi) \text{ for } \delta \leq \eta \leq \xi.$$

Hence, we have, for the first integral in the positive part

$$\int_0^\xi \frac{\omega(\eta)}{\eta}\, d\eta \leq \delta + \omega(\xi) \log\frac{\xi}{\delta} \leq \omega(\xi)\left(2 + \log\frac{\xi}{\delta}\right)$$

because

$$\omega(\xi) \geq \omega(\delta) > \frac{\delta}{2} \text{ if } \delta \text{ is small enough.}$$

The second integral in the positive part can be bounded as

$$\int_\xi^\infty \frac{\omega(\eta)}{\eta^2}\, d\eta = \frac{\omega(\xi)}{\xi} + \gamma \int_\xi^\infty \frac{d\eta}{\eta^2(4+\log(\eta/\delta))} \leq \frac{\omega(\xi)}{\xi} + \frac{\gamma}{\xi} \leq \frac{2\omega(\xi)}{\xi}$$

if $\gamma < \delta/2$ and δ is small enough. Thus, the positive part is bounded from above by

$$P \leq A\omega(\xi)\left(4 + \log\frac{\xi}{\delta}\right)\omega'(\xi) = A\gamma\frac{\omega(\xi)}{\xi}, \quad \text{for } \xi \geq \delta.$$

To estimate the negative term, we will now rely on N_2. Due to the concavity of ω, we have (see (5.25))

$$\omega(2\eta+\xi) - \omega(2\eta-\xi) \leq \omega(2\xi), \text{ for all } \eta \geq \frac{\xi}{2}.$$

In addition, for $\xi \geq \delta$, we have

$$\omega(2\xi) = \omega(\xi) + \gamma \int_\xi^{2\xi} \frac{d\eta}{\eta(4+\log(\eta/\delta))} \leq \omega(\xi) + \frac{\gamma \log 2}{4} \leq \frac{3}{2}\omega(\xi)$$

if $\gamma < \delta/10$. Therefore, N_2 can be bounded as

$$N_2 = \frac{1}{\pi}\int_{\xi/2}^\infty \frac{\omega(2\eta+\xi) - \omega(2\eta-\xi) - 2\omega(\xi)}{\eta^2}\, d\eta \leq -\frac{1}{2\pi}\int_{\xi/2}^\infty \frac{\omega(\xi)}{\eta^2}\, d\eta = -\frac{1}{\pi}\frac{\omega(\xi)}{\xi}\,.$$

It follows that

$$P + N \leq P + N_2 \leq \frac{\omega(\xi)}{\xi}(A\gamma - \frac{1}{\pi}) < 0,$$

if γ is small enough. This completes the proof of Theorem 5.1. □

6 Finite time blow up: the Burgers equations

No story in nonlinear PDE is complete without a finite time blow up example. Such example is not available for the SQG equation or the classical 3D equations of fluid mechanics. We will therefore consider a more basic but still quite interesting equation - the Burgers equation. The lack of nonlocality in the nonlinear term makes this equation much more amenable to analysis.

In this section, we will prove finite time blow up for the supercritical dissipative Burgers equation given by

$$\partial_t \theta + \theta\theta_x + (-\Delta)^\alpha \theta = 0, \;\; \theta(x,0) = \theta_0(x). \tag{6.1}$$

We will consider equation set in one dimension with periodic boundary conditions. The main result we will prove is

Theorem 6.1 *Suppose that $\alpha \geq 1/2$, and θ_0 is smooth. Then there exists a unique smooth solution $\theta(x,t)$ of (6.1).*

Suppose that $\alpha < 1/2$. Then there exist smooth θ_0 such that the corresponding solution blows up in finite time.

The proof of the first part of the Theorem is similar to the SQG equation case. In fact the argument is simpler due to the advection velocity being equal to just θ.

Exercise. Carry out the global regularity proof for the critical Burgers equation. What would you get in place of the estimate (5.2)?

For the Burgers equation without dissipation, finite time shock formation is well known and can be shown by the method of characteristics (see e.g. [52]). This method does not work for the dissipative part of the equation. The equation (6.1) has two terms each of which is relatively easy to understand: local nonlinear term and fractional heat equation dissipative term. One way to analyze equation with such structure is to use a method called time splitting. The origin of this method is the Trotter formula which says that

$$e^{A+B} = \lim_{n\to\infty} \left(e^{A/n}e^{B/n}\right)^n$$

where A, B are for example bounded self-adjoint operators (see e.g. [105], [112]). This formula can be generalized to nonlinear setting, and time splitting approach was used in [76] to prove the second part of Theorem 6.1. However this method is fairly technical. We will use a different, less direct approach based on construction of Lyapunov functional (following [50]). Let us first illustrate this approach by providing a proof of loss of regularity of solutions to inviscid Burgers. This method will be easier to generalize to the viscous case than the characteristics approach.

Suppose that the period is equal to 2. Consider θ_0 which is odd, and $\theta_0(x) \geq 0$ for $-1 \leq x \leq 0$. It is not difficult to show, using uniqueness of solutions, that the solution $\theta(x,t)$ is odd, too, at least while it remains smooth. This is similar to the proofs of symmetry conservation for the solutions of 2D Euler equation that we carried out in the previous chapter. Let

$$L(t) = -\int_0^1 \frac{\theta(x,t)}{x^\sigma}\, dx.$$

Note that if $\sigma < 2$, the integral defining $L(t)$ is convergent while $\theta(x,t)$ remains smooth, due to its oddness. Now while $\theta(x,t)$ is smooth,

$$L'(t) = -\int_0^1 \frac{\partial_t \theta(x,t)}{x^\sigma}\,dx = \frac{1}{2}\int_0^1 \frac{\partial_x(\theta(x,t))^2}{x^\sigma}\,dx = \frac{\sigma}{2}\int_0^1 \frac{\theta(x,t)^2}{x^{1+\sigma}}\,dx.$$

By Hölder inequality,

$$\left|\int_0^1 \frac{\theta(x,t)}{x^\sigma}\,dx\right| \le \left(\int_0^1 \frac{\theta(x,t)^2}{x^{1+\sigma}}\,dx\right)^{1/2}\left(\int_0^1 x^{1-\sigma}\,dx\right)^{1/2}.$$

Thus $L'(t) \ge CL(t)^2$, and $L(0) > 0$ by choice of initial data. Hence $L(t)$ has to blow up in finite time, and therefore $\theta(x,t)$ has to lose regularity in finite time.

The proof of the second part of Theorem 6.1 follows a similar idea.

Proof. [Proof of Theorem 6.1] Suppose again that the period is equal to 2. Take as before θ_0 to be odd and $\theta_0(x) \ge 0$ for $-1 \le x \le 0$. Let us introduce a weight function

$$w(x) = \begin{cases} \operatorname{sgn}(x)\left(|x|^{-\delta} - 1\right), & |x| \in (0,1), \\ 0 & x \notin (-1,1), \end{cases} \tag{6.2}$$

$0 < \delta < 2$. Consider $L(t) = -\int_{-1}^1 \theta(x,t)w(x)\,dx$. Note that $L(0) \ge 0$. Then

$$L'(t) = \frac{1}{2}\int_{-1}^1 \partial_x(\theta(x,t)^2)w(x)\,dx + \int_{-1}^1 (-\Delta)^\alpha \theta(x,t)w(x)\,dx. \tag{6.3}$$

Assuming that $\theta(x,t)$ remains smooth, we can integrate by parts in the first integral in (6.3), obtaining

$$\frac{1}{2}\theta(x,t)^2 w(x)\Big|_{-1}^{1} + \frac{\delta}{2}\int_{-1}^1 \theta(x,t)^2|x|^{-1-\delta}\,dx,$$

where the first term vanishes due to (6.2). On the other hand,

$$-\int_{-1}^1 \theta(x,t)w(x)\,dx \le \left(\int_{-1}^1 \theta(x,t)^2|x|^{-1-\delta}\,dx\right)^{1/2}\left(\int_{-1}^1 |x|^{1-\delta}\right)^{1/2}.$$

Therefore, similarly to the inviscid computation,

$$\int_{-1}^1 \theta(x,t)^2|x|^{-1-\delta}\,dx \ge CL(t)^2.$$

Next, consider

$$\int_{\mathbb{R}} (-\Delta)^\alpha \theta(x,t)w(x)\,dx.$$

Recall that

$$(-\Delta)^\alpha \theta(x) = P.V.\int_{\mathbb{R}} \frac{\theta(x) - \theta(y)}{|x-y|^{1+2\alpha}}\,dy,$$

where $\theta(y)$ is extended periodically to $\mathbb{R}$.

First, we claim that

$$\int_{\mathbb{R}} (-\Delta)^\alpha \theta(x,t) w(x)\, dx = \int_{\mathbb{R}} \theta(x,t) (-\Delta)^\alpha w(x)\, dx. \tag{6.4}$$

Indeed,

$$\int_{\mathbb{R}} P.V. \int_{\mathbb{R}} \frac{\theta(x)-\theta(y)}{|x-y|^{1+2\alpha}}\, dy w(x)\, dx = \lim_{\varepsilon\to 0} \int_{|x-y|\geq \varepsilon} \frac{(\theta(x)-\theta(y))w(x)}{|x-y|^{1+2\alpha}}\, dydx =$$
$$\frac{1}{2}\lim_{\varepsilon\to 0} \int_{|x-y|\geq \varepsilon} \frac{(\theta(x)-\theta(y))w(x) + (\theta(y)-\theta(x))w(y)}{|x-y|^{1+2\alpha}}\, dxdy =$$
$$\frac{1}{2}\lim_{\varepsilon\to 0} \int_{|x-y|\geq \varepsilon} \frac{\theta(x)(w(x)-w(y)) + \theta(y)(w(y)-w(x))}{|x-y|^{1+2\alpha}}\, dxdy = \lim_{\varepsilon\to 0} \int_{|x-y|\geq \varepsilon} \theta(x) \frac{w(x)-w(y)}{|x-y|^{1+2\alpha}}\, dydx.$$

To estimate (6.4), we need to control $(-\Delta)^\alpha w(x)$.

Lemma 6.2 *Assume* $0 < \delta < 1$. *Then*
(i) If $|x| \geq 1$, *we have* $|(-\Delta)^\alpha w(x)| \leq C|x|^{-2-2\alpha}$.
(ii) If $|x| < 1$, *we have* $|(-\Delta)^\alpha w(x)| \leq C|x|^{-\delta-2\alpha}$.
In particular, if $\delta \in (0, 1-2\alpha)$, *then* $\int_{\mathbb{R}} |(-\Delta)^\alpha w(x)|\, dx \leq C$.

Let us finish the proof of Theorem 6.1 using Lemma 6.2. Given $\alpha < 1/2$, choose δ so that $1-2\alpha > \delta > 0$. With such choice of parameters, it follows from Lemma 6.2 and our earlier computations that $L'(t) \geq C_1 L(t)^2 - C_2$. If $L(0)$ is sufficiently large, this differential inequality implies finite time blow.

Exercise. Verify carefully the last claim.

Note that $L(t)$ cannot diverge; for our parameters the integrand is absolutely integrable even if $\theta(x,t)$ is not smooth but just bounded. Rather, we obtain a contradiction with our assumption that $\theta(x,t)$ remains smooth for all times. The estimates we got from integration by parts no longer hold true after some finite time due to loss of regularity. □

Proof. [Proof of Lemma 6.4] Since $(-\Delta)^\alpha w(x)$ is odd, it is enough to consider $x \geq 0$. For $x \geq 1$, we have

$$(-\Delta)^\alpha w(x) = -\int_{-1}^{1} \frac{w(y)}{|x-y|^{1+2\alpha}}\, dy = -\int_0^1 (y^{-\delta}-1)((x-y)^{-1-2\alpha} - (x+y)^{-1-2\alpha})\, dy. \tag{6.5}$$

If $x \geq 2$, the right hand side of (6.5) does not exceed

$$Cx^{-2-2\alpha} \int_0^1 (y^{-\delta}-1) y\, dy \leq Cx^{-2-2\alpha}$$

by mean value theorem. If $1 \leq x \leq 2$, observe that $(x-y)^{-1-2\alpha} - (x+y)^{-1-2\alpha}$ is decreasing in x if $x \geq 1$ and $y \in [0,1]$. So the expression is maximal when $x = 1$:

$$\left| \int_0^1 (y^{-\delta}-1)((x-y)^{-1-2\alpha} - (x+y)^{-1-2\alpha})\, dy \right| \leq \left| \int_0^1 y^{-\delta}(1-y^\delta)(1-y)^{-1-2\alpha}\, dy \right| \leq C$$

if $2\alpha < 1$, $\delta < 1$. This completes the proof of (i).

Consider next $1 > x \geq 0$. Set $r = y - x$, then

$$(-\Delta)^\alpha \theta(x) = P.V. \int_{\mathbb{R}} \frac{w(x) - w(x+r)}{r^{1+2\alpha}}\, dr = P.V. \int_{-1-x}^{1-x} \frac{w(x) - w(x+r)}{r^{1+2\alpha}}\, dr + \qquad (6.6)$$
$$w(x) \left| \int_{-\infty}^{-1-x} r^{-1-2\alpha}\, dr + \int_{1-x}^{\infty} r^{-1-2\alpha}\, dr \right|.$$

Look at the last term. The first integral is controlled by $C(x+1)^{-2\alpha} w(x) \leq Cx^{-\delta}$. The second integral gives

$$(x^{-\delta} - 1) \int_{1-x}^{\infty} r^{-1-2\alpha}\, dr \leq Cx^{-\delta}(1 - x^\delta)(1-x)^{-2\alpha} \leq Cx^{-\delta}$$

if $2\alpha < 1$.

Consider the first term on the right hand side of (6.6),

$$P.V. \int_{-1-x}^{1-x} \frac{w(x) - w(x+r)}{r^{1+2\alpha}}\, dr = \int_{-1-x}^{-x} \frac{x^{-\delta} - 2 + |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr + P.V. \int_{-x}^{1-x} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr. \qquad (6.7)$$

In the first integral on the right hand side,

$$\left| \int_{-1-x}^{-x} \frac{1}{r^{1+2\alpha}}\, dr \right| \leq Cx^{-2\alpha},$$

so only $\int_{-1-x}^{-x} \frac{|x+r|^{-\delta}}{r^{1+2\alpha}}\, dr$ needs estimating, the rest can be bounded by $Cx^{-2\alpha-\delta}$. Now

$$\int_{-1-x}^{-x} \frac{|x+r|^{-\delta}}{r^{1+2\alpha}}\, dr = \int_{x}^{1+x} \frac{1}{r^{1+2\alpha}|x-r|^\delta}\, dr = \int_{x}^{3x/2} \frac{1}{r^{1+2\alpha}|x-r|^\delta}\, dr +$$
$$\int_{3x/2}^{1+x} \frac{1}{r^{1+2\alpha}|x-r|^\delta}\, dr \leq Cx^{-1-2\alpha} \int_0^{x/2} y^{-\delta}\, dy + Cx^{-\delta} \int_{3x/2}^{1+x} r^{-1-2\alpha}\, dr \leq Cx^{-2\alpha-\delta}.$$

So it is left to estimate the last term on the right hand side of (6.7),

$$P.V. \int_{-x}^{1-x} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr = \int_{-x}^{-x/2} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr + P.V. \int_{-x/2}^{1-x} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr.$$

The first integral on the right hand side does not exceed

$$Cx^{-1-2\alpha} \left(x^{1-\delta} + \int_0^{x/2} y^{-\delta}\, dy \right) \leq Cx^{-2\alpha-\delta}.$$

For the second integral, note that $|x+r| \geq x/2$ in the range of integration, so $x^{-\delta} - |x+r|^{-\delta} \leq Cx^{-\delta-1} r$. Then

$$P.V. \int_{-x/2}^{1-x} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr \leq Cx^{-\delta-1} \int_{-x/2}^{1-x} r^{-2\alpha}\, dr \leq Cx^{-\delta-2\alpha}$$

if $1 - x \leq x$. If $1 - x > x$, we need to also estimate

$$\int_{x}^{1-x} \frac{x^{-\delta} - |x+r|^{-\delta}}{r^{1+2\alpha}}\, dr \leq x^{-\delta} \int_{x}^{1-x} r^{-1-2\alpha}\, dr \leq Cx^{-\delta-2\alpha}.$$

□

6.1 Appendix: A Gagliardo-Nirenberg inequality

In this section we will prove a Gagliardo-Nirenberg inequality

Theorem 6.3 *Suppose $f \in C_0^\infty(\mathbb{R}^d)$. Let $1 \le m \le s$, with m integer. Then*

$$\|D^m f\|_{L^{2s/m}} \le C \|f\|_{L^\infty}^{1-m/s} \|(-\Delta)^{s/2} f\|_{L^2}^{m/s} \tag{6.8}$$

Here D^m denotes an arbitrary partial derivative of order m.

Exercise. Given this inequality for functions in $\mathbb{R}^d$, prove the analogous estimate for the $\mathbb{T}^d$ case.

We note that this is just one of the larger family of Gagliardo-Nirenberg inequalities. When checking which inequality of this sort may be possible, the first test to apply is scaling: change $f(x)$ to $f(\lambda x)$ and see if both sides scale the same. Another test to verify is whether the strength of the derivative on the right hand side controls that on the left hand side (when powers of the norms are taken into account). The inequality (6.3) verifies the scaling and is sharp for the second check: $m = (2s/2)m/s$. Now let us prove it.

First, recall Littlewood-Paley decomposition. Take $\varphi \in C_0^\infty(\mathbb{R}^d)$, φ radial, radially decreasing, non-negative, $\varphi(\xi) = 1$ if $|\xi| \le 1$, and $\varphi(\xi) = 0$ if $|\xi| \ge 2$. Set $\psi(\xi) = \varphi(\xi/2) - \varphi(\xi)$. Denote $\psi_l(\xi) = \psi(2^{-l}\xi)$, $l \in \mathbb{Z}$. Then $\sum_{l \in \mathbb{Z}} \psi_l(\xi) = 1$ for every $\xi \ne 0$. Define Littlewood-Paley projections of a function f as $P_l f(x) = (\psi_l(\xi)\hat{f}(\xi))\check{} = Q_l * f(x)$, where $\hat{f}(\xi)$ denotes the Fourier transform of f, and $\check{g}$ denotes the inverse Fourier transform of the function g. Observe that for a smooth function f, $f(x) = \sum_{l \in \mathbb{Z}} P_l f(x)$. We will use a short hand notation $f_l(x) \equiv P_l f(x)$.

Note that we can also write $f_l(x) = Q_l * f(x)$, where $Q_l(x) = \check{\psi}_l(x) = 2^{ld} Q_1(2^l x) = 2^{ld}\check{\psi}(2^l x)$. In particular, $\|Q_l\|_{L^1}$ does not depend on l.

Lemma 6.4 *Let $D^m f$ denote any partial derivative of $f \in C_0^\infty(\mathbb{R}^d)$ of order m. Then*

$$\|D^m f\|_{L^2}^2 \le C \sum_{l \in \mathbb{Z}} 2^{2lm} \|f_l\|_{L^2}^2. \tag{6.9}$$

Also, for every $s \in \mathbb{R}$,

$$\|(-\Delta)^s f\|_{L^2}^2 \sim \sum_{l \in \mathbb{Z}} 2^{2ls} \|f_l\|_{L^2}^2. \tag{6.10}$$

Here $\sim$ means both upper and lower bounds with some universal positive constants.

Proof. After taking Fourier transform, we see that both (6.9) and (6.10) would follow from

$$\int_{\mathbb{R}^d} |\xi|^{2m} |\hat{f}(\xi)|^2 \, d\xi \sim \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2 \sum_{l \in \mathbb{Z}} 2^{2lm} |\psi_l(\xi)|^2.$$

But from definition of ψ_l it follows that $1 \ge \sum_{l \in \mathbb{Z}} |\psi_l(\xi)|^2 \ge c > 0$ for all $\xi \ne 0$. This together with support location of ψ_l finishes the proof of (6.9) and (6.10). $\square$

Proof. [Proof of Theorem 6.3] Without loss of generality, we can assume that $\|f\|_{L^\infty} = 1$. Write $f(x) = \sum_{l\in\mathbb{Z}} f_l(x)$. Observe that $\|f_l\|_{L^\infty} \le \|Q_1\|_{L^1}$. Let us write the left hand side of (6.8) as

$$\int_{\mathbb{R}^d} \left| \sum_{l\in\mathbb{Z}^d} 2^{lm} \left(2^{-lm} D^m f_l\right) \right|^{2s/m} dx. \tag{6.11}$$

Set $g_l = 2^{-lm} D^m f_l$. Denote $\eta(\xi) = \psi(\xi)\xi_{i_1}\dots\xi_{i_m}$, where ξ_{i_j} in the product corresponding to partial derivatives in D^m. Observe that $\eta \in C_0^\infty$, so that $\check{\eta} \in L^1$. Then

$$\|g_l\|_{L^\infty} \le \|(2^{-lm}\xi_{i_1}\dots\xi_{i_m}\psi_l)\check{}\|_{L^1} = \|(\eta(2^{-l}\xi)\check{}\|_{L^1} = \|2^{ld}\check{\eta}(2^l x)\|_{L^1} = \|\check{\eta}\|_{L^1}.$$

Coming back to (6.11), consider $\left|\sum_{l\in\mathbb{Z}} 2^{lm} g_l(x)\right|^{2s/m}$ at some fixed x. Look at subsums over all l such that $2^{-k}\|\check{\eta}\|_{L^1} \le g_l(x) \le 2^{-k+1}\|\check{\eta}\|_{L^1}$, $k \in \mathbb{N}$. Denote such l by notation $g_l \sim 2^{-k}$. This set certainly may depend on x, but we will suppress this in notation. Then

$$\sum_{l\in\mathbb{Z}} 2^{lm} g_l(x) = \sum_{k=1}^{\infty} \sum_{g_l\sim 2^{-k}} 2^{lm} g_l(x).$$

Lemma 6.5 *For every x, we have $g_l(x) \to 0$ as $l \to \pm\infty$. Therefore, each $g_l \sim 2^{-k}$ set is finite.*

Proof. Recall that

$$g_l(x) = 2^{ld} \int_{\mathbb{R}} \check{\eta}(2^l(x-y))g(y)\, dy.$$

The fact that $g_l(x) \to 0$ as $l \to -\infty$ follows immediately from $g \in C_0^\infty \subset L^1$. For the $l \to \infty$ case, note that $\int_{\mathbb{R}^d} \check{\eta}(y)\, dy = 0$ (since $\eta(0) = 0$). The result then follows from the following statement which is left as an exercise.

Exercise. Suppose that $f, g \in C_0^\infty$, and $\int_{\mathbb{R}^d} f\, dx = 0$. Then $\int_{\mathbb{R}^d} 2^{ld} f(2^l(x-y))g(y)\, dy \to 0$ as $l \to 0$.

□

Consider now $\sum_{g_l\sim 2^{-k}} 2^{lm} g_l(x)$. Denote l_{max} the maximal value of l such that $g_l \sim 2^{-k}$. By Lemma 6.4, such maximal value exists. Then

$$\sum_{g_l\sim 2^{-k}} 2^{lm}|g_l(x)| \le 8\|\check{\eta}\|_{L^1} 2^{l_{max}m} 2^{-k} \le 8\|\check{\eta}\|_{L^1} \left(\sum_{g_l\sim 2^{-k}} 2^{2sl}|g_l(x)|^2 \right)^{m/2s} 2^{-k(s-m)/s}. \tag{6.12}$$

Indeed, the sum on the right hand side is simply a geometric progression (up to a factor of two), perhaps with gaps, and as such it is dominated by the largest term in the series. Summing up estimates (6.12) in k and using Hölder inequality, we get

$$\sum_{l\in\mathbb{Z}} 2^{lm}|g_l(x)| \le 8\|\check{\eta}\|_{L^1} \sum_{k=1}^{\infty} \left(\left(\sum_{g_l\sim 2^{-k}} 2^{2lm}|g_l(x)|^2 \right)^{m/2s} 2^{-k(s-m)/s} \right) \le C \left(\sum_{l\in\mathbb{Z}} 2^{2ml}|g_l(x)|^2 \right)^{m/2s}.$$

Then

$$\|D^m f\|_{L^{2s/m}}^{2s/m} \leq \int_{\mathbb{R}^d} \left(\sum_{l\in\mathbb{Z}} 2^{lm} |g_l(x)| \right)^{2s/m} dx \leq C \int_{\mathbb{R}^d} \sum_{l\in\mathbb{Z}} 2^{2sl} |g_l(x)|^2 \, dx \leq$$
$$C \int_{\mathbb{R}^d} \sum_{l\in\mathbb{Z}} 2^{2(s-m)l} |\xi|^{2m} |\psi_l(\xi)|^2 |\hat{f}(\xi)|^2 \, d\xi \leq C\|f\|_s^2.$$

The last step follows similarly to proof of (6.10). Since $\|f\|_{L^\infty}$ was normalized to be one, this is exactly (6.8). ⌑

Bibliography

[1] S. Agmon, On positivity and decay of solutions of second order elliptic equations on Riemannian manifolds, pp. 19-52, in: Methods of Functional Analysis and Theory of Elliptic Equations, D. Greco, ed., Liguori Ed.Napoli, 1983.

[2] G. Alberti, L. Ambrosio and X. Cabré, On a long-standing conjecture of E. De Giorgi: symmetry in 3D for general nonlinearities and a local minimality property. Acta Appl. Math. **65**, 2001, 9–33.

[3] L. Ambrosio and X. Cabré, Entire solutions of semilinear elliptic equations in $\mathbb{R}^3$ and a conjecture of de Giorgi, Jour. Amer. Math. Soc. **13**, 2000, 725–739.

[4] H. Bahouri, J.-Y. Chemin, *Équations de transport relatives á des champs de vecteurs nonLipschitziens et mécanique des uides. (French) [Transport equations for non-Lipschitz vector Fields and fluid mechanics],* Arch. Rational Mech. Anal., **127** (1994), no. 2, 159–181

[5] I.J. Bakelman, *Convex analysis and nonlinear geometric elliptic equations*, Springer-Verlag, Berlin, 1994

[6] G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi*, Mathématiques & Applications, **17** Springer-Verlag, Paris, 1994.

[7] G. Barles and B. Perthame, Discontinuous solutions of deterministic optimal stopping time problems, M2AN, **21**, 1987, 557–579.

[8] R.F. Bass, *Probabilistic Techniques in Analysis*, Springer, 1995.

[9] R.F. Bass, *Diffusions and Elliptic Operators*, Springer, 1998.

[10] H. Bateman, *Some recent researches on the motion of fluids,* Monthly Weather Review **43** (1915), 163-170

[11] J.T. Beale, T. Kato and A. Majda, *Remarks on the breakdown of smooth solutions for the 3D Euler equations,* Commun. Math. Phys. **94** (1984), 61–66

[12] A. Bensoussan, J.L. Lions and G. Papanicoalou, *Asymptotic Analysis for Periodic Structures*, AMS, 2011.

[13] H. Berestycki, L. Caffarelli and L. Nirenberg, Monotonicity for elliptic equations in unbounded Lipschitz domains, Comm. Pure Appl. Math., **50**, 1997, 1089–1111.

[14] H. Berestycki and F. Hamel, *Reaction-diffusion Equations And Propagation Phenomena*, Springer, 2014, to appear.

[15] H. Berestycki, F. Hamel and R. Monneau, One-dimensional symmetry of bounded entire solutions of some elliptic equations, Duke Math. Jour., **103**, 2000, 375-396.

[16] H. Berestycki, F. Hamel and N. Nadirashvili, Elliptic eigenvalue problems with large drift and applications to nonlinear propagation phenomena. Comm. Math. Phys. **253**, 2005, 451–480.

[17] H. Berestycki, F. Hamel and G. Nadin, Asymptotic spreading in heterogeneous diffusive excitable media. J. Funct. Anal. **255**, 2008, 2146–2189.

[18] H. Berestycki, F. Hamel and L. Rossi, Liouville-type results for semilinear elliptic equations in unbounded domains, Ann. Mat. Pura Appl. (4) **186**, 2007, 469-507.

[19] H. Berestycki, A. Kiselev, A. Novikov and L. Ryzhik, The explosion problem in a flow, Jour. d'Anal. Mathématique, **110**, 2010, 31–65.

[20] H. Berestycki and L. Nirenberg, On the method of moving planes and the sliding method, Bol. Soc. Brasil. Mat. **22**, 1991, 1–37.

[21] H. Berestycki and L. Nirenberg, Traveling fronts in cylinders, Annales de l'IHP, Analyse non linéare, **9**, 1992, 497-572.

[22] H. Berestycki, L. Nirenberg and S.R.S Varadhan, The principal eigenvalue and maximum principle for second-order elliptic operators in general domains, Comm. Pure Appl. Math., **47**, 1994, 47–92.

[23] H. Berestycki and L. Rossi, On the principal eigenvalue of elliptic operators in $\mathbb{R}^N$ and applications. J. Eur. Math. Soc. (JEMS) **8**, 2006, 195–215.

[24] P. Bernard, Existence of $C^{1,1}$ critical sub-solutions of the Hamilton-Jacobi equation on compact manifolds. Ann. Sci. cole Norm. Sup. (4) **40**, 2007, 445-452.

[25] P. Bernard, Smooth critical sub-solutions of the Hamilton-Jacobi equation, Math. Res. Lett., **14**, 2007, 503-511.

[26] J.M. Burgers, *A mathematical model illustrating the theory of turbulence,* Adv. Appl. Mech **1** (1948), 171–199

[27] X. Cabre, Partial differential equations, geometry and stochastic control (in Catalan), Butl. Soc. Catalana Mat. **15**, 200, 7–27.

[28] X. Cabre, Elliptic PDEs in probability and geometry: symmetry and regularity of solutions, Discr. Cont. Sys. B, **20**, 2008, 425–457.

[29] L. Caffarelli and A. Vasseur. *Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation*, Ann. of Math. (2) **171** (2010), 1903–1930

[30] K. Choi, T. Hou, A. Kiselev, G. Luo, V. Sverak and Y. Yao, *On the Finite-Time Blowup of a 1D Model for the 3D Axisymmetric Euler Equations,* preprint arXiv:1407.4776

[31] A.J. Chorin and J.E. Marsden, *A Mathemtical Introduction to Fluid Mechanics,* Springer, 1993

[32] P. Constantin, A. Kiselev, L. Ryzhik and A. Zlatos, Diffusion and mixing in fluid flow. Ann. of Math., **168**, 2008, 643–674.

[33] P. Constantin, A. Majda and E. Tabak. *Formation of strong fronts in the 2D quasi-geostrophic thermal active scalar*. Nonlinearity, 7, (1994), 1495–1533

[34] P. Constantin and V. Vicol, *Nonlinear maximum principles for dissipative linear nonlocal operators and applications,* Geometric and Functional Analysis, **22** (2012), 1289-1321

[35] P. Constantin, A. Tarfulea, and V. Vicol, *Long time dynamics of forced critical SQG,* preprint arXiv:1308.0640

[36] P. Constantin and C. Foias, *Navier-Stokes Equations,* University of Chicago Press, 1988

[37] P. Constantin, Q. Nie, and N. Schorghofer, *Nonsingular surface-quasi-geostrophic flow,* Phys. Lett. A, **24** (1998), 168–172

[38] P. Constantin and J. Wu, *Regularity of Hölder continuous solutions of the supercritical quasi-geostrophic equation,* Ann. Inst. H. Poincare Anal. Non Linearie, **25** (2008), 1103–1110

[39] D. Cordoba, *Nonexistence of simple hyperbolic blow up for the quasi-geostrophic equation*, Ann. of Math., **148**, (1998), 1135–1152

[40] D. Cordoba and C. Fefferman, *Growth of solutions for QG and 2D Euler equations,* Journal of the AMS, **15**, 665–670

[41] A. Cordoba and D. Cordoba, *A maximum principle applied to to quasi-geostrophic equations,* Commun. Math. Phys. **249** (2004), 511–528

[42] M.G. Crandall and P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations, Trans. Amer. Math. Soc., **277**, 1983, 1–42.

[43] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Vol. 3: Spectral Theory and Applications, Springer, 2000.

[44] M. del Pino, M. Kowalczyk, and J. Wei, On a conjecture by De Giorgi in dimensions 9 and higher. Symmetry for elliptic PDEs, 115–137, Contemp. Math., 528, Amer. Math. Soc., Providence, RI, 2010.

[45] S. Denisov, *Infinite superlinear growth of the gradient for the two-dimensional Euler equation,* Discrete Contin. Dyn. Syst. A, **23** (2009), 755–764

[46] S. Denisov, *Double-exponential growth of the vorticity gradient for the two-dimensional Euler equation,* to appear in Proceedings of the AMS, preprint arXiv:1201.1771

[47] S. Denisov, *The sharp corner formation in 2D Euler dynamics of patches: infinite double exponential rate of merging,* preprint arXiv:1201.2210

[48] C.R. Doering and J.D. Gibbon, *Applied Analysis of the Navier-Stokes Equations,* Cambridge Texts in Applied Mathematics (Book 12), Cambridge University Press, 1995

[49] H. Dong and D. Du, *Global well-posedness and a decay estimate for the critical quasi-geostrophic equation,* Discrete Contin. Dyn. Syst., **21** (2008) no. 4, 1095–1101

[50] H. Dong, D. Du and D. Li, *Finite time singularities and global well-posedness for fractal Burgers' equation,* Indiana Univ. Math. J., **58** (2009), 807-

[51] L. Euler, *Principes généraux du mouvement des fluides,* Mémoires de L'Académie Royale des Sciences et des Belles-Lettres de Berlin **11** (4 September 1755, printed 1757), 217–273; reprinted in Opera Omnia ser. 2 **12**, 219–250

[52] L.C. Evans, Partial Differential Equations, Second Edition, AMS, 2010.

[53] L.C. Evans and R. Gariepy, *Measure Theory and Fine Properties of Funcitons*, CRC Press, 1992.

[54] L.C. Evans and P. Souganidis, A PDE approach to certain large deviation problems for systems of parabolic equations. Analyse non linéaire (Perpignan, 1987). Ann. Inst. H. Poincaré Anal. Non Linéaire **6** (1989), suppl., 229–258.

[55] L.C. Evans and P. Souganidis, A PDE approach to geometric optics for certain semilinear parabolic equations. Indiana Univ. Math. J. **38**, 1989, 141–172.

[56] E.B. Fabes and D.W. Stroock, A new proof of Moser's parabolic Harnack inequality using the old ideas of Nash, Arch. Rational Mech. Anal. **96**, 1986, 327–338.

[57] A. Fathi, Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens, C. R. Acad. Sci. Paris Sér. I Math. **324**, 1997, 1043–1046.

[58] A. Fathi, Solutions KAM faibles conjuguées et barrières de Peierls, C. R. Acad. Sci. Paris Sr. I Math. **325**, 1997, 649–652.

[59] A. Fathi, Orbites hétéroclines et ensemble de Peierls, C. R. Acad. Sci. Paris Sér. I Math. **326**, 1998, 1213–1216.

[60] A. Fathi, Sur la convergence du semi-groupe de Lax-Oleinik, C. R. Acad. Sci. Paris Sér. I Math. **327**, 1998, 267–270.

[61] A. Fathi, *Weak KAM Theory*, Cambridge University Press.

[62] A. Fathi and A. Siconolfi, Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation, Invent. Math. **155**, 2004, 363–388.

[63] R.A. Fisher, The wave of advance of advantageous genes, Ann. Eugenics, **7**, 1937, 355-369.

[64] A.R. Forsyth, *Theory of Differential Equations. Part 4. Partial Differential Equations* (Vol. 5-6), 1906

[65] B. Franke, Integral inequalities for the fundamental solutions of diffusions on manifolds with divergence-free drift. (English summary) Math. Z. **246**, 2004, n 373–403.

[66] M. Freidlin and J. Gártier, The propagation of concentration waves in periodic and random media. (Russian) Dokl. Akad. Nauk SSSR **249**, 1979, 521–525.

[67] M. Freidlin, Functional integration and partial differential equations. Annals of Mathematics Studies, 109. Princeton University Press, Princeton, NJ, 1985.

[68] U. Frisch, *Turbulence,* Cambridge University Press, 1999

[69] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of the Second Order*, Springer-Verlag, Berlin, Heidelberg, 1983.

[70] N. Ghoussoub and C. Gui, On the De Giorgi conjecture in dimensions 4 and 5, Ann. Math., **157**, 2003, 313-334.

[71] Q. Han and F.Lin, *Elliptic Partial Differential equations*, AMS, 1997.

[72] I. Held, R. Pierrehumbert, S. Garner and K. Swanson. *Surface quasi-geostrophic dynamics*, J. Fluid Mech., 282, (1995), 1–20

[73] T. Hou and G. Luo, *Potentially Singular Solutions of the 3D Incompressible Euler Equations,* preprint arXiv:1310.0497

[74] T. Kato, *Remarks on the Euler and Navier-Stokes equations in* $\mathbb{R}^2$*,* Proc. Symp. Pure Math. **45** (1986), 1–7

[75] Y. Katznelson, *An Introduction to Harmonic Analysis*, Cambridge University Press, 2004.

[76] A. Kiselev, F. Nazarov and R. Shterenberg, *On blow up and regularity in dissipative Burgers equation,* Dynamics of PDEs, **5** (2008), 211–240

[77] A. Kiselev, F. Nazarov and A. Volberg, *Global well-posedness for the critical* 2*D dissipative quasi-geostrophic equation,* Inventiones Math. **167** (2007) 445–453

[78] A. Kiselev, *Some recent results on the critical surface quasi-geostrophic equation: a review. Hyperbolic problems: theory, numerics and applications,* 105122, Proc. Sympos. Appl. Math., 67, Part 1, Amer. Math. Soc., Providence, RI, 2009

[79] A. Kiselev and F. Nazarov, *A variation on a theme of Caffarelli and Vasseur,* Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **370** (2009)

[80] A. Kiselev and V. Sverak, *Small scale creation for solutions of the incompressible two dimensional Euler equation,* arXiv:1310.4799 [math.AP], to appear at Annals Math.

[81] H. Koch, *Transport and instability for perfect fluids,* Math. Ann. **323** (2002), 491–523

[82] A.N. Kolmogorov, I.G. Petrovskii and N.S. Piskunov, Étude de l'équation de la chaleurde matière et son application à un problème biologique, Bull. Moskov. Gos. Univ. Mat. Mekh. **1** (1937), 1-25. (see [103] pp. 105-130 for an English transl.)

[83] J.- M. Lasry, P.- L. Lions, A Remark on Regularization in Hilbert Spaces, Israel Math. J. **55**, 1996, 257–266.

[84] M. Ledoux, *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[85] P.-L. Lions, *Generalized solutions of Hamilton-Jacobi Equations,* Research Notes in Mathematics, Pitman, 1983.

[86] P.-L. Lions, G. Papanicolaou and S.R.S. Varadhan, Homogenization of Hamilton-Jacobi equations, Preprint.

[87] A. Majda and A. Bertozzi, *Vorticity and Incompressible Flow,* Cambridge University Press, 2002

[88] A. Majda, *Introduction to PDEs and Waves for the Atmoshpere and Ocean,* Courant Lecture Notes in Mathematics, AMS 2003

[89] C. Marchioro and M. Pulvirenti, *Mathematical Theory of Incompressible Nonviscous Fluids,* Applied Mathematical Sciences Series (Springer-Verlag, New York), **96**, 1994

[90] J.N. Mather, Variational construction of connecting orbits, Ann. Inst. Fourier **43**, 1993, 1349–1386.

[91] V. Maz'ja, *Sobolev Spaces,* Springer-Verlag, Berlin Heidelberg, 1985

[92] A. Morgulis, A. Shnirelman and V. Yudovich, *Loss of smoothness and inherent instability of 2D inviscid fluid flows,* Comm. Partial Differential Equations **33** (2008), no. 4-6, 943-968

[93] J. Moser, *On the volume elements on a manifold,* Trans. Amer. Math. Soc., **120**, 1965, 286–294.

[94] J.D. Murray, Mathematical biology. I. An introduction. Third edition. Interdisciplinary Applied Mathematics, 17. Springer-Verlag, New York, 2002.

[95] J.D. Murray, Mathematical biology. II. Spatial models and biomedical applications. Third edition. Interdisciplinary Applied Mathematics, 18. Springer-Verlag, New York, 2003

[96] N.S. Nadirashvili, *Wandering solutions of the two-dimensional Euler equation (Russian),* Funktsional. Anal. i Prilozhen. **25** (1991), 70–71; translation in Funct. Anal. Appl. **25** (1991), 220–221 (1992)

[97] J. Nash, Continuity of solutions of parabolic and elliptic equations, Amer. Jour. Math., **80**, 1958, 931-954.

[98] J.R. Norris, Long-time behaviour of heat flow: global estimates and exact asymptotics. Arch. Rational Mech. Anal. **140**, 1997, 161–195.

[99] R. Nussbaum and Y. Pinchover, On variational principles for the generalized principal eigenvalue of second order elliptic operators and some applications, J. Anal. Math. **59**, 1992, 161-177.

[100] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 2010.

[101] K. Ohkitani and M. Yamada, *Inviscid and inviscid-limit behavior of a surface quasi-geostrophic flow,* Phys. Fluids, **9** (1997), 876–882

[102] J. Pedlosky, *Geophysical Fluid Dynamics,* Springer, New York, 1987

[103] Dynamics of curved fronts, P. Pelcé, Ed., Academic Press, 1988.

[104] Y. Pinchover and J. Rubinstein, *An Introduction to Partial Differential Equations*, Cambridge University Press, 2005.

[105] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis,* Academic Press, New York, 1972

[106] R.T. Rockafellar, *Convex analysis,* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997.

[107] J.-M. Roquejoffre, Eventual monotonicity and convergence to travelling fronts for the solutions of parabolic equations in cylinders. Ann. Inst. H. Poincaré Anal. Non Linéaire, **14**, 1997, 499–552.

[108] O. Savin, Regularity of flat level sets in phase transitions, Ann. of Math. **169**, 2009, 41–78.

[109] E. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals,* Princeton University Press, 1993

[110] E. Stein, *Singular Integrals and Differentiability Properties of Functions,* Princeton University Press, 1970

[111] S. Sternberg, *On differential equations on the torus,* Amer. J. Math., **79**, 1957, 397–402.

[112] M. Taylor, *Partial Differential Equations. II. Qualitative Studies of Linear Equations,* Applied Mathematical Sciences **116**, Springer-Verlag, New York 1996

[113] R. Temam, *Navier-Stokes Equations: Theory and Numerical Analysis,* AMS Chelsea Pub., 2001

[114] C.E. Wayne, *Vortices and two-dimensional fluid motion,* Notices of the AMS, **58**, no. 1, 10–19

[115] H. Weinberger, On spreading speeds and traveling waves for growth and migration models in a periodic habitat. J. Math. Biol. **45**, 2002, 511–548. Erratum: J. Math. Biol. **46**, 2003, p. 190.

[116] W. Wolibner, *Un theorème sur l'existence du mouvement plan d'un fluide parfait, homogène, incompressible, pendant un temps infiniment long. (French)* Mat. Z., **37** (1933), 698726

[117] J. Xin, Analysis and modelling of front propagation in heterogeneous media, SIAM Rev., **42**, 2000, 161-230.

[118] J. Xin, *An Introduction to Fronts in Random Media*, Springer, 2009.

[119] V.I. Yudovich, *The loss of smoothness of the solutions of the Euler equation with time (Russian),* Dinamika Sploshn. Sredy **16** Nestacionarnye Probelmy Gidordinamiki (1974), 71–78

[120] V.I. Yudovich, *On the loss of smoothness of the soltions to the Euler equation and the inhereht instability of flows of an ideal fluid,* Chaos **10** (2000), 705–719

[121] V. I. Yudovich, *Non-stationary flows of an ideal incompressible ?uid,* Zh Vych Mat, **3** (1963), 1032–1066

[122] A. Zlatos, *Exponential growth of the vorticity gradient for the Euler equation on the torus,* preprint arXiv:1310.6128