AARMS-CRM Workshop on NA of SPDEs, July 2016

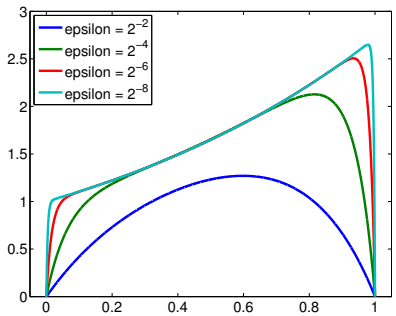
http://www.math.mun.ca/~smaclachlan/anasc_spde/

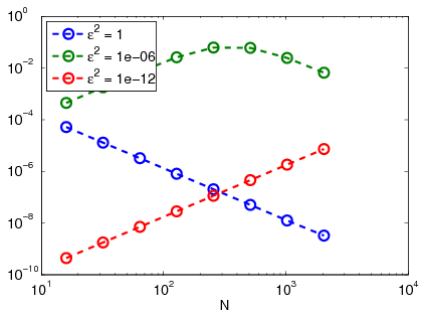
Short course on Numerical Analysis of Singularly Perturbed Differential Equations

Niall Madden, NUI Galway (Niall.Madden@NUIGalway.ie)

§2 Numerical methods for SPDEs and uniform convergence

Version 25.07.16





Handout version.

Outline

	Monday, 25 July	Tuesday, 26 July
09:00	Welcome/Coffee	
09:15	1. Introduction to singularly perturbed problems	5. PDEs (i): time-dependent problems.
10:00	Break	
10:15	2. Numerical methods and uniform convergence; FDMs and their analysis.	6. PDEs (ii): elliptic problems 7. Finite Element Methods
12:00	*Lunch*	
14:00	3. Coupled systems	8. Convection-diffusion (Stynes)
15:00	Break	
15:15	3. Coupled systems (continued)	9. Nonlinear problems (Kopteva)
16:15	4. Lab 1	10. Lab 2 (PDEs)
17:30	*Finish*	

§2 Numerical methods for SPDEs

(1:45)

Primary references

The main mathematical content of this presentation, starting at Slide 21, closely follows [Miller et al., 1996] and [Miller et al., 2012].

The definition of *parameter uniform*ity (Slides 16–18) is from [Farrell et al., 2000].

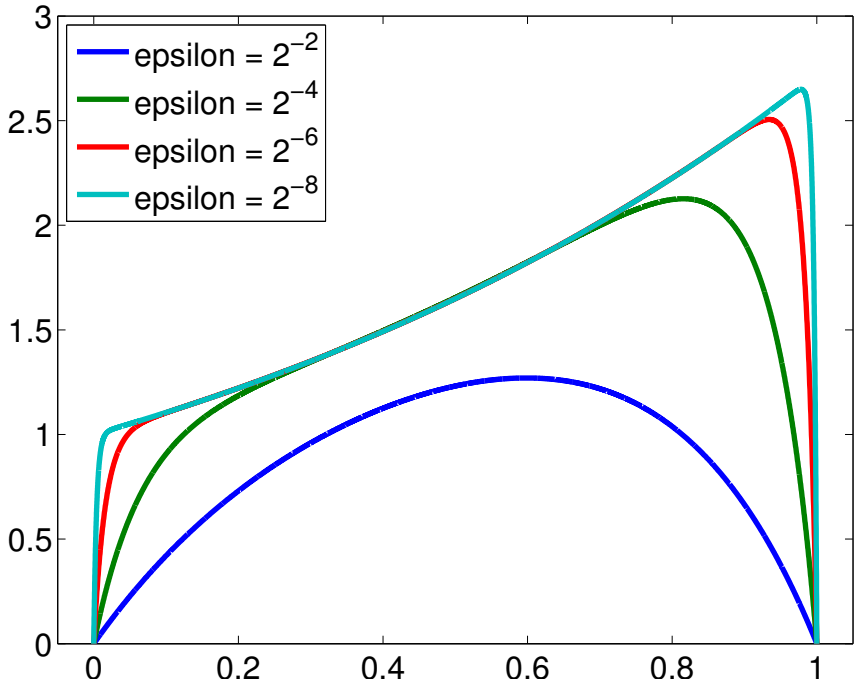
Important secondary references include [Protter and Weinberger, 1984].

A reaction-diffusion problem

Let's recall our first example of a *singularly perturbed reaction-diffusion* equation.

$$-\varepsilon^2 u''(x) + b(x)u(x) = f(x), \qquad \text{on } \Omega = (0,1),$$

- ε is (still) a small parameter; it may take any value in $(0, 1]$.
- There is $\beta > 0$ such that $b(x) \geqslant \beta > 0$.
- Boundary conditions: $u(0) = u(1) = 0$



In general, one must approximate the solutions to such problems by some numerical scheme.

A "*Parameter Robust*" or "*Uniformly Convergent*" method is one that yields an approximation $\mathfrak{U}$ of $\mathfrak{u}$, such that one can prove an error estimate of the form

$$\|\mathfrak{u} - \mathfrak{U}\| \leqslant \mathrm{C}\mathrm{N}^{-\mathrm{p}}$$

where C, p ("rate of convergence") are independent of the perturbation parameter ε, and discretization parameter N. This should be valid for all $\varepsilon \in (0, 1]$ and all N.

In particular, one should *not* have to assume that, for example, $\mathrm{N} = \mathcal{O}(1/\varepsilon)$.

It is also desirable that any layer present should be resolved.

This explanation of "uniform convergence" is heuristic, (and we have not even specified $\|\cdot\|$). The concept will be will be made formal later.

The simplest numerical scheme one could apply to this problem is a second-order finite difference scheme on a uniform mesh.

- On the interval $\overline{\Omega} = [0,1]$, form a uniform mesh with N intervals:

$$\Omega^N := \{x_i\}_{i=0}^N, \qquad \text{where } x_i = i/N = ih;$$

- Approximate u'' as

$$u'' = \underbrace{\frac{1}{h^2}\big(u(x_{i-1}) - 2u(x_i) + u(x_{i+1})\big)}_{\delta^2 u(x_i)} + C\underbrace{\|u^{(4)}\|}_{\mathcal{O}(\varepsilon^{-4})} N^{-2}.$$

- Construct and solve the linear system

$$\begin{aligned} U_0 &= 0, \\ -\varepsilon^2 \delta^2 U_i + b(x_i)U_i &= f(x_i), \qquad i = 1, \ldots, N-1 \\ U_N &= 0. \end{aligned}$$

If we implement the above finite difference method, and then calculate the maximum point-wise error, we get the following results.

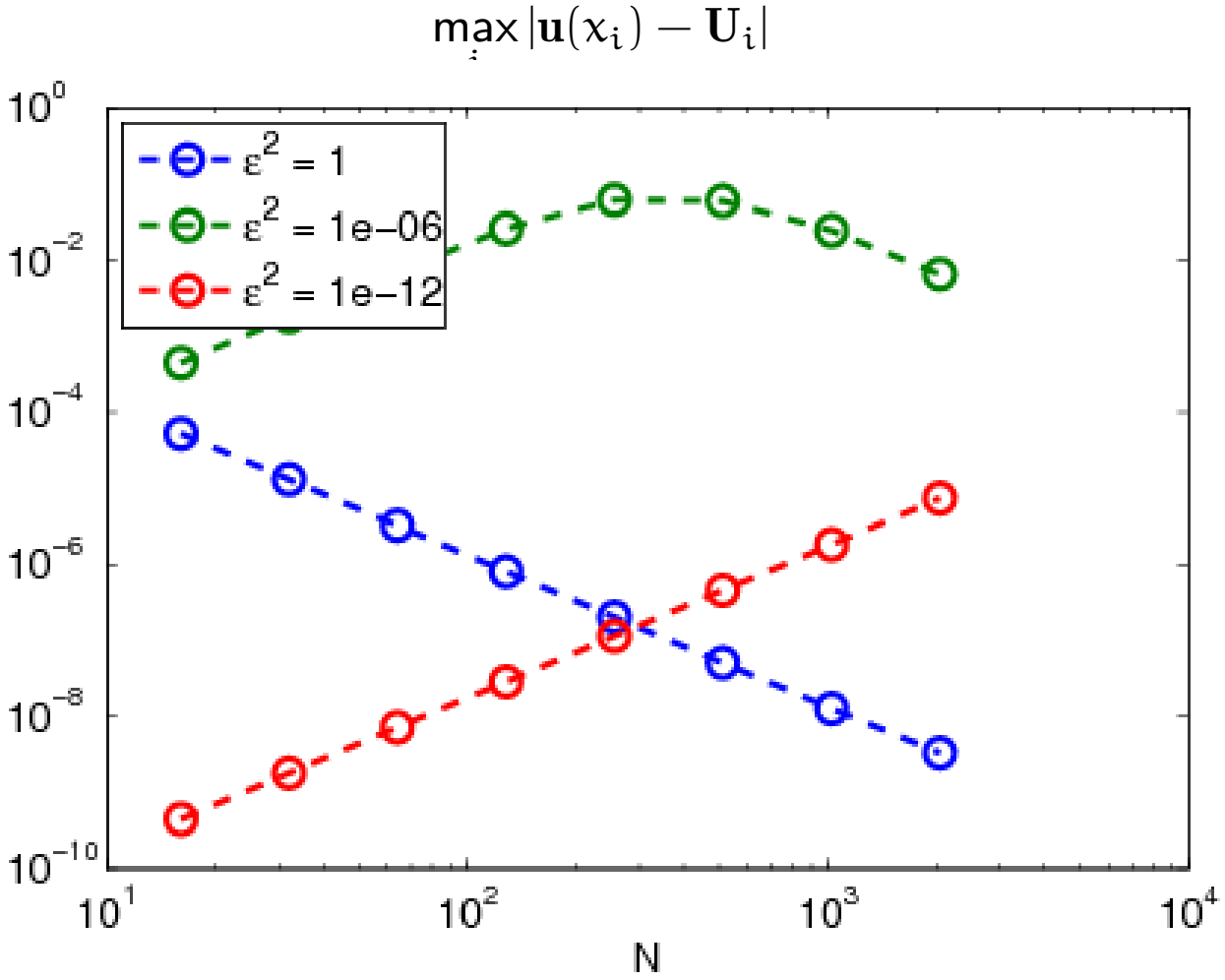
$$\max_i |\mathbf{u}(x_i) - \mathbf{U}_i| \quad \text{where } u \text{ solves } -\varepsilon^2 u'' + u = e^x.$$

ε^2	$N = 64$	$N = 128$	$N = 256$	$N = 512$
1	7.447e-06	1.861e-06	4.654e-07	1.163e-07
10^{-2}	1.023e-03	2.568e-04	6.424e-05	1.607e-05
10^{-4}	7.689e-02	2.338e-02	6.192e-03	1.583e-03
10^{-6}	1.104e-02	4.203e-02	1.033e-01	9.666e-02
10^{-8}	1.113e-04	4.452e-04	1.779e-03	7.088e-03
10^{-10}	1.113e-06	4.453e-06	1.781e-05	7.125e-05
10^{-12}	1.113e-08	4.453e-08	1.781e-07	7.125e-07

We observe that,

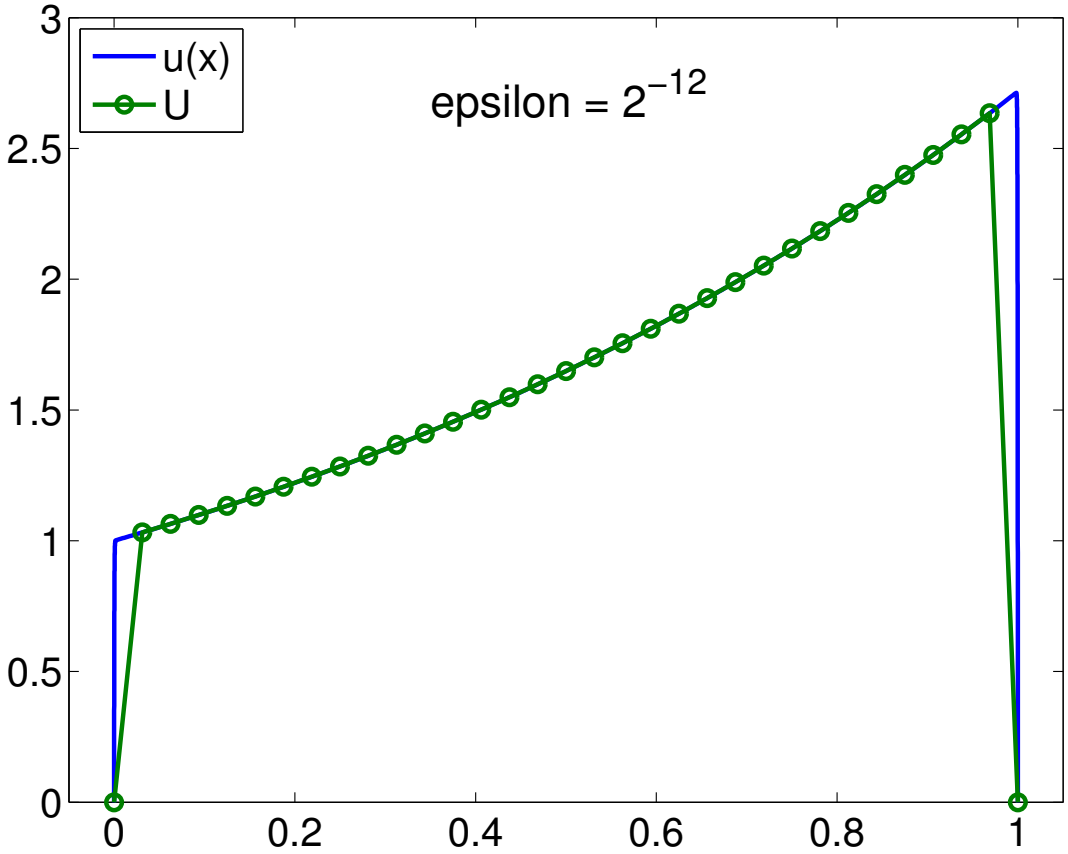
- for small fixed N the error decreases as ε *decreases* (counter-intuitive)
- for small fixed ε, the error *increases* as N increases (i.e., not converging)

Comparing "convergence" for different values of ε.



Why, for small fixed N, does the error appear to decrease as ε is reduced?

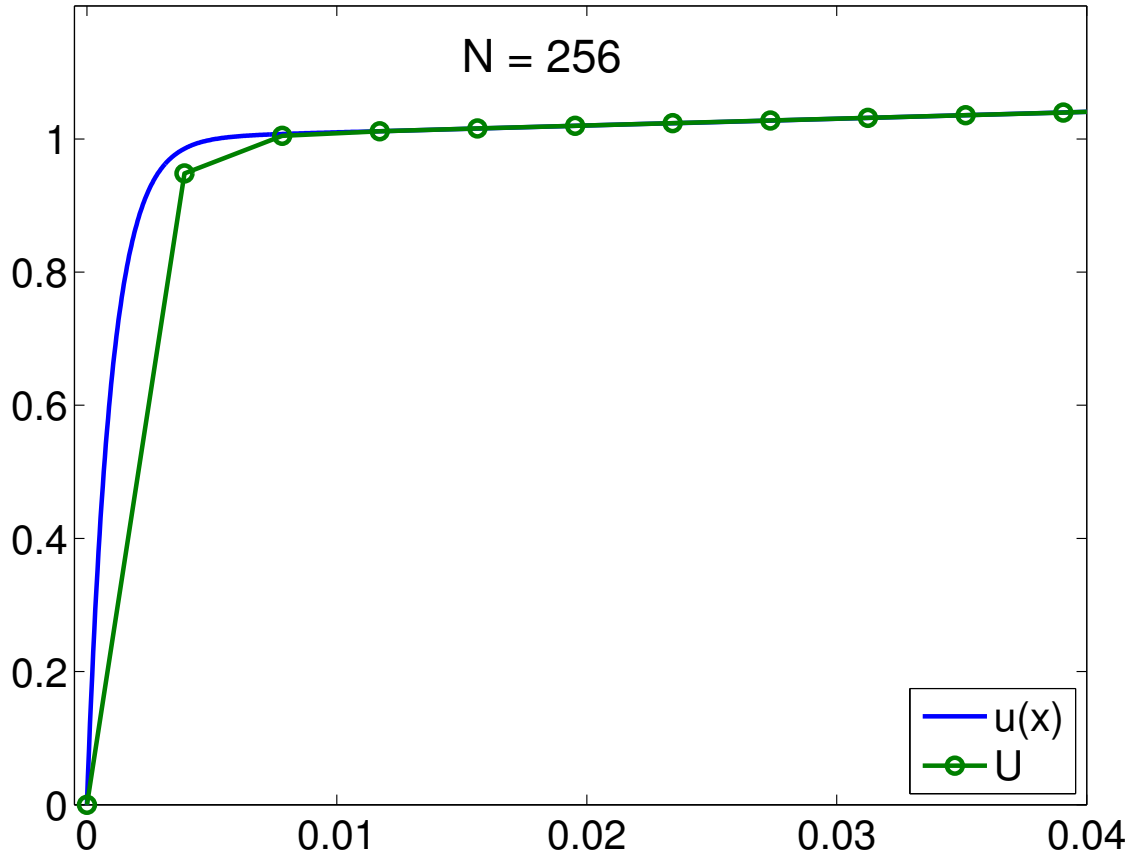
We fix $N = 32$ and take $\varepsilon = 10^{-2}, 10^{-4}, \ldots, 10^{-10}$.



The pointwise errors are small because the layer is not resolved.

Why, for small fixed ε, does the error appear to increase as N is increased?

We fix $\varepsilon = 2^{-10}$ and take $N = 32, 64, 128, \ldots$. As N approaches ε^{-1}, the method begins to resolve the layer, and so the computed pointwise error increases.



Motivated by the previous graphs, we compute the difference between the true solution and the *piecewise linear interpolant* to the approximation.

$$\max_{0 \leqslant x \leqslant 1} |\mathbf{u}(x) - \bar{\mathbf{U}}(x)|$$

ε^2	N = 64	N = 128	N = 256	N = 512
1	3.75e-01	3.75e-01	3.75e-01	3.75e-01
1e-02	3.77e-01	3.75e-01	3.75e-01	3.75e-01
1e-04	4.62e-01	4.06e-01	3.84e-01	3.78e-01
1e-06	7.30e-01	6.86e-01	5.94e-01	4.89e-01
1e-08	7.50e-01	7.49e-01	7.47e-01	7.37e-01
1e-10	7.50e-01	7.50e-01	7.50e-01	7.50e-01
1e-12	7.50e-01	7.50e-01	7.50e-01	7.50e-01

We can conclude from this that the method given here is not suitable for this problem.

Most of the remainder of this section will be given over to deriving and analysing a method that *is* suitable.

The differential equation is $-\varepsilon^2 u'' + bu = f$.
From this we see that $\|u''\|$ is $\mathcal{O}(\varepsilon^{-2})$.
If b is constant, by differentiating the DE, we get that $\|u^{(4)}\|$ is $\mathcal{O}(\varepsilon^{-4})$. (We will do this more carefully for variable b later).

If standard arguments based on the truncation error are employed, one will find that

$$\|u - U\|_\infty \leqslant CN^{-2}(1 + \varepsilon^{-2}).$$

(This bound suggests that this method is inappropriate for this problem, which is true; but it is not sharp. We will return to this point later).

A convection-diffusion problem

Before we study the reaction-diffusion problem in greater depth, we take a detour to point out that things could be much, much *worse*.

The numerical method presented above yields a reasonable solution to the reaction-diffusion problem *away from layers*.

If we apply the method to the obvious *convection-diffusion problem*, the resulting solution can be unstable.

Again we start with the uniform mesh with N intervals:

$$\Omega^N := \{x_i\}_{i=0}^N, \qquad \text{where } x_i = i/N = ih;$$

And again approximate u'' as $u'' = \delta^2 u(x_i) + K_2 N^{-2}$.

We approximate u' by the corresponding second-order central difference scheme

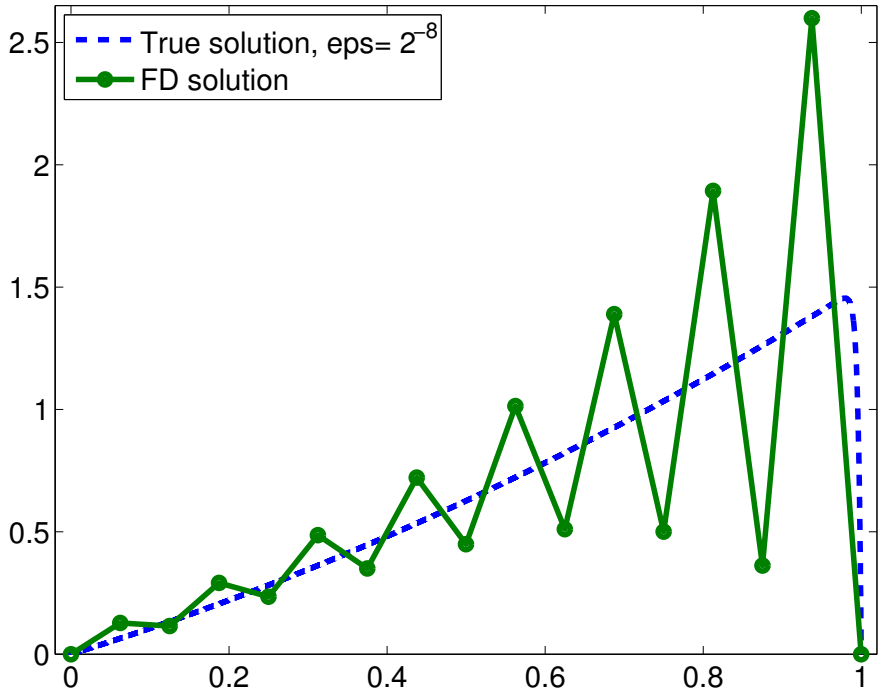
$$u' = \underbrace{\frac{1}{2h}\big(-u(x_i) + u(x_{i+1})\big)}_{D^0 u(x_i)} + K_1 N^{-2}.$$

The finite difference method is then

$$-\varepsilon\delta^2 U_i + a(x_i) D^0 U_i = f(x_i), \qquad i = 1, \ldots, N-1.$$

A convection-diffusion problem

$$-\varepsilon u''(x) + u'(x) = 1 + x, \quad \text{on } \Omega = (0,1), \text{ with } u(0)=u(1)=0$$



Uniform convergence

We have seen that the two simplest methods for reaction-diffusion and convection-diffusion problems are inadequate.

Before constructing a method that *is* adequate, we need a concept of what "adequate" means.

Although it is possible to design a method that gives a "reasonable" solution for a fixed, small ε, we want to investigate schemes which are accurate for *all* $\varepsilon \in (0, 1]$.

Furthermore, the scheme should not rely on choosing some large $N = N(\varepsilon)$ in order to ensure accuracy.

That is...

[Farrell et al., 2000, p10]

... we undertake ... the task of constructing numerical methods that generate numerical solutions which converge uniformly for all values of the parameter ε in the range $(0, 1]$, and that require a parameter-uniform amount of computational work to compute each numerical solution.
Such methods are called parameter uniform or ε-uniform methods.

[Farrell et al., 2000, p10]

If a method is ε-uniform, the error between the exact solution, u, and the numerical solution U, satisfies an estimate of the following form: for some positive integer N_0, all integers $\mathrm{N} \geqslant \mathrm{N}_0$, and all $\varepsilon \in (0, 1]$, we have

$$\|\mathrm{u} - \bar{\mathrm{U}}\|_{\bar{\Omega}} \leqslant \mathrm{C}\mathrm{N}^{-\mathrm{p}}.$$

where C, N_0 and p are positive constants independent of ε and N.

Here U is taken to be a *mesh function* defined on some set of (mesh) points in the domain $\bar{\Omega}$, and $\bar{\mathrm{U}}$ is its piecewise linear interpolant. The norm $\|\mathrm{u} - \bar{\mathrm{U}}\|_{\bar{\Omega}}$ is the maximum norm.

In the above discussion,

- the emphasis on the maximum norm comes from the fact that other norms, particularly energy norms for simple Galerkin FEMs, are not strong enough to identify layers.
- the interpolant of the numerical solution features since, as we have seen, if there are no mesh points within the layer, the solution can appear highly accurate.

So [Farrell et al., 2000] propose that methods for SPPs should be

(1) **global**: yielding an approximation that can be evaluated at all points in the domain;

(2) **point-wise accurate**,

(3) **parameter uniform** (independent of ε and computational effort)

(4) **monotone** (discrete operator respects key qualitative properties of the continuous operator).

The reminder of this section of the presentation is given over to the mathematical analysis of solutions to *one-dimensional reaction-diffusion equations*, and **finite difference** methods for approximating them.

Notation

The following notation applies in the remainder of this section.

- $\Omega := (0,1)$.
- $\bar{\Omega}^N = \{0 = x_0, x_1, \ldots, x_N = 1\}$ is a (possibly arbitrary) mesh with N intervals.
  Ω^N denotes the interior of this mesh, i.e., $\Omega^N = \{x_1, \ldots, x_{N-1}\}$.
- $\|\cdot\|$ is the maximum norm on $\mathcal{C}(\bar{\Omega})$. That is $\|u\| := \|u\|_{\infty,\bar{\Omega}} = \max\limits_{0 \leqslant x \leqslant 1} |u(x)|$.
- $\|\cdot\|_{\bar{\Omega}^N}$ is the discrete max norm for mesh functions on Ω^N.

Always: C is a constant that is independent of ε and N. It can take different values in different places – even in the same expression.

Definition (Differential operator $\mathcal{L}$)

Let $\Omega := (0,1)$. Given the function $b \in \mathcal{C}^4(\bar{\Omega})$, subject to $b(x) \geqslant \beta^2 > 0$, and parameter $\varepsilon \in (0,1]$, the differential operator $\mathcal{L}$ is defined, for all $\psi \in \mathcal{C}^2(\bar{\Omega})$ as

$$\mathcal{L}\psi = -\varepsilon^2\psi'' + b\psi$$

Definition (Reaction-diffusion equation)

Let u be the solution to

$$\mathcal{L}u = f \text{ on } (0,1), \text{ with boundary conditions } u(0) = u(1) = 0, \qquad (1)$$

where $f \in \mathcal{C}^4(\bar{\Omega})$.

The imposition of homogeneous Dirichlet boundary conditions is only to simplify the exposition. Results are easily extended to more general situations.

Maximum principles

The following result is very standard, and elementary, but worth considering in detail in order to be able to generalise later.

Lemma

Maximum Principle If $\phi(0) \geqslant 0$, $\phi(1) \geqslant 0$, *and* $\mathcal{L}\psi(x) \geqslant 0$ *for all* $x \in \Omega$, *then* $\phi(x) \geqslant 0$ *for all* $x \in \bar{\Omega}$.

[*arguments on the board*]

Maximum principles

There are many useful consequences of this. For example:

It easily follows that, if $\mathcal{L}u = f$ and $u(0) = u(1) = 0$, then

$$u(x) \leqslant \|f\|/\beta^2.$$

[*arguments on the board*]

(Here $\psi = \|f\|/\beta^2$ is called a *barrier function*; we'll see more of these).

This shows that u is bounded. But for the analysis of a finite difference method, we need bounds on derivatives of u.

Bounds on derivatives

Lemma (Lemma 6.1 of [Miller et al., 2012])

$$\|u^{(k)}\| \leqslant C(1+\varepsilon^{-k}).$$

[*arguments on the board*]

Bounds on derivatives

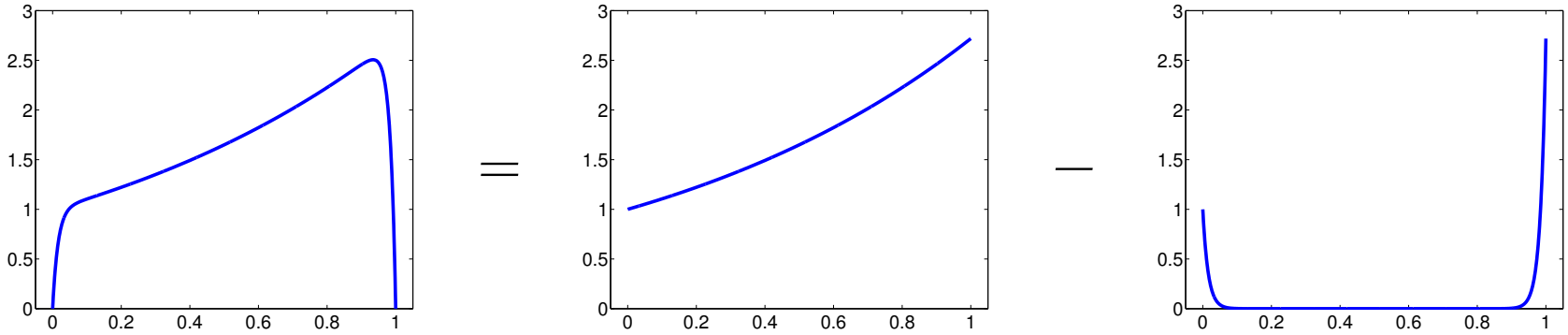
The above bounds are correct, but not at all sharp: one expects that $|u''(x)| \approx \varepsilon^{-2}$ only near the boundary.

Our numerical analysis will require sharper, point-wise bounds obtained via a *decomposition* of the solution of the DE as the sum of "regular" and "layer" components:

$$u = \underbrace{v}_{\text{regular}} + \underbrace{w}_{\text{layer}}.$$

Solution decomposition

$$u = \underbrace{v}_{\text{regular}} + \underbrace{w}_{\text{layer}}.$$



Roughly,

- v represents the solution away from the boundaries, but any layers present are only weakly expressed.
- w accounts for the boundary conditions and, thus, the boundary layers, but decays rapidly away from the boundaries.

Definition (Reduced problem)

To get the reduced problem, set $\varepsilon = 0$ in the reaction-diffusion equation (1), and neglect the boundary conditions. That is, let v_0 be the solution to

$$b(x)v_0(x) = f(x).$$

Regular part

Let $v = v_0 + \varepsilon^2 v_1$, where v_0 is the reduced solution, and v_1 solves

$$\mathcal{L}v_1 = v_0'', \qquad v_1(0) = v_1(1) = 0.$$

From our earlier lemma, it is clear that $|v_1^{(k)}(x)| \leqslant C(1 + \varepsilon^{-k})$.
It follows that $|v^{(k)}(x)| \leqslant C(1 + \varepsilon^{-k+2})$.

Layer part

Since $u = w + v$, we have that w satisfies the homogeneous DE

$$-\varepsilon^2 w'' + bw = 0 \text{ on } \Omega, \qquad w = u - v \text{ on } \partial\Omega$$

Define the boundary layer function

$$\mathcal{B}_\varepsilon(x) := \exp(-x\beta/\varepsilon) + \exp(-(1-x)\beta/\varepsilon).$$

Also define the barrier function

$$\psi^\pm(x) = C\mathcal{B}_\varepsilon(x) \pm w(x)$$

for some suitable large C so that $\psi^\pm$ is non-negative on the boundary. Then the Maximum Principle shows that $|w(x)| \leqslant \mathcal{B}_\varepsilon(x)$ on $\bar{\Omega}$.

Now a minor variant on the argument that was used to bound $|u^{(k)}|$ will give that

$$|w^{(k)}(x)| \leqslant C\varepsilon^{-k}\mathcal{B}_\varepsilon(x).$$

The above solution decomposition tells use that the derivatives of u *are* large, but decay rapidly away from the boundaries. So it will be of no surprise to learn that there is a good strategy for solving these problems that involves a mesh condensing near those boundaries.

First, for an arbitrary mesh $\Omega^N = \{x_0, x_1, \ldots, x_n\}$ the standard second-order finite-difference operator becomes

$$(\delta^2 v)_i = \frac{2}{h_{i+1} + h_i} \left(\frac{v_{i+1} - v_i}{h_{i+1}} - \frac{v_i - v_{i-1}}{h_i} \right).$$

where $h_i = x_i - x_{i-1}$.

It satisfies

E.g., [Miller et al., 2012, Lemma 4.1]

$$\left| (\delta^2 - \frac{d^2}{dx^2}) \phi(x_i) \right| \leqslant \frac{1}{3} (x_{i+1} - x_{i-1}) |\phi|_3.$$

Then the FDM is

$$U_0 = 0,$$
$$L^N U_i := -\varepsilon^2 \delta^2 U_i + b(x_i) U_i = f(x_i), \qquad i = 1, \ldots, N-1,$$
$$U_N = 0.$$

This operator satisfies a *discrete maximum principle*, and an ε-uniform stability result: if $U_0 = U_N = 0$, then

$$|\Phi_i| \leqslant \frac{1}{\beta} \max_{i \leqslant j \leqslant N} |L^N \Phi_j|.$$

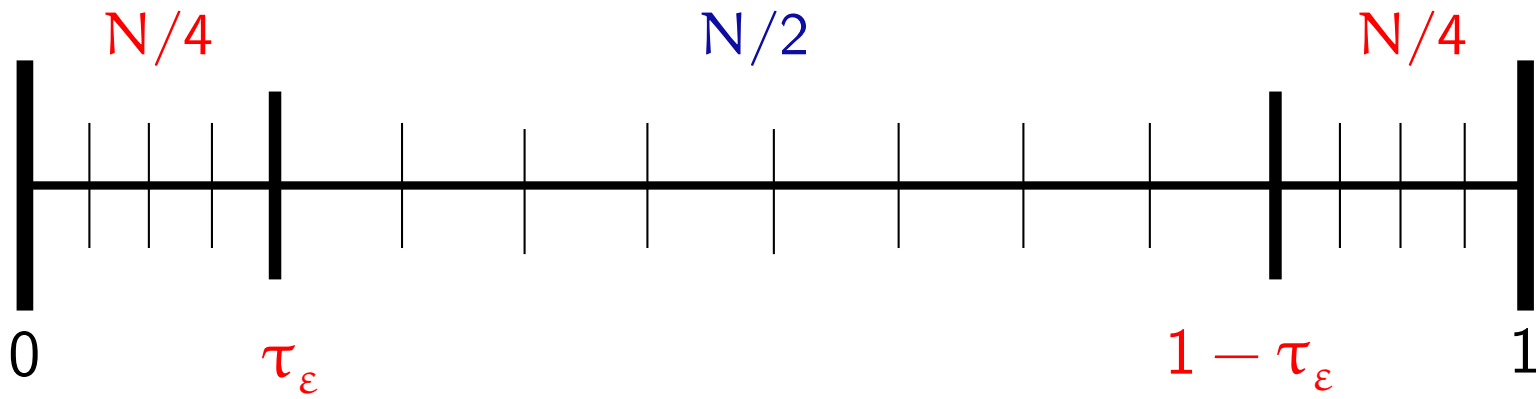
[*arguments on the board*]

The piecewise uniform mesh of *Shishkin* is constructed as follows[1]:

1. Choose a **mesh transition point**

$$\tau_\varepsilon = \min\left\{\frac{1}{4}, \frac{\varepsilon}{\beta}\ln N\right\}.$$

2. Divide the domain into three sub-regions: $[0, \tau_\varepsilon]$, $[\tau_\varepsilon, 1-\tau_\varepsilon]$ and $[1-\tau_\varepsilon, 1]$.
3. Subdivide those sub-regions to obtain the mesh.



[1] This choice is for the purpose of exposition. It is better to choose $\tau_\varepsilon = \min\{1/4, 2\varepsilon/\beta \ln N\}$.

Theorem (Theorem 6.4 of [Miller et al., 1996])

There is a constant C that is independent of ε and N such that

$$\|u - U\|_{\bar{\Omega}^N} \leqslant CN^{-1} \ln N.$$

The proof proceeds by constructing a decomposition of the discrete solution $U = V + W$ that is analogous to the decomposition of the continuous solution:

$$\begin{aligned} V \text{ solves} \qquad & V_0 = v(0), \quad L^N V_i = f(x_i), \quad V_N = v(1), \\ W \text{ solves} \qquad & W_0 = w(0), \quad L^N W_i = 0, \quad W_N = w(1). \end{aligned}$$

We analyse these terms separately, i.e., we estimate $\|v - V\|$ and $\|w - W\|$.
For $\|v - V\|$... [*arguments on the board*]

Analysis

On the region $(0, \tau_\varepsilon) \cup (1 - \tau_\varepsilon, 1)$ the analysis for $\|w - W\|$ is analogous...
[*arguments on the board*]

Analysis

Finally, on the region $[\tau_\varepsilon, 1-\tau_\varepsilon]$ one can exploit the fact that w and W have decayed. In particular, for any $x_i \in [\tau, 1-\tau]$

$$\mathcal{B}_\varepsilon(x_i) \leqslant \mathcal{B}_\varepsilon(\tau) \leqslant 2\exp(-\tau\beta/\varepsilon) \leqslant 2\exp(-\ln N) \leqslant CN^{-1}.$$

[*arguments on the board*]

Wrap up

- The significance of the above result is that we have designed a scheme for which we can prove that the pointwise error is independent of ε.
- The result is easily extended to show that $\|u - \bar{U}\| \leqslant CN^{-1} \ln N$.
- The result is correct, but not sharp. Can the scheme and analysis be improved so that $\|u - U\|_{\bar{\Omega}^N} \leqslant CN^{-2} \ln^2 N$? [Discuss!]
- The approach here, of using a piecewise uniform mesh, is very elementary. A more sophisticated mesh, such as the graded mesh of Bakhvalov, can yield a fully second-order scheme.

References

Farrell, P. A., Hegarty, A. F., Miller, J. J. H., O'Riordan, E., and Shishkin, G. I. (2000).
Robust Computational Techniques for Boundary Layers.
Number 16 in Applied Mathematics. Chapman & Hall/CRC, Boca Raton, U.S.A.

Miller, J. J. H., O'Riordan, E., and Shishkin, G. I. (1996).
Fitted numerical methods for singular perturbation problems.
World Scientific Publishing Co., Inc., River Edge, NJ.

Miller, J. J. H., O'Riordan, E., and Shishkin, G. I. (2012).
Fitted numerical methods for singular perturbation problems.
World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, revised edition.

Protter, M. H. and Weinberger, H. F. (1984).
Maximum principles in differential equations.
Springer-Verlag, New York.
Corrected reprint of the 1967 original.