MA8502 Numerical solution of partial differential equations

Spectral discretization in $\mathbb{R}^1$

Fall 2012

Department of Mathematical Sciences
NTNU, N-7491 Trondheim, Norway

1 Spectral discretization based on high order polynomials

We consider the discretization of the Poisson problem in one space dimension. The domain is assumed to be $\Omega = (-1, 1)$. The boundary conditions are homogeneous Dirichlet conditions at the two end points, i.e., the solution $u(x)$ is here $u(\pm 1) = 0$. The given right hand side for the Poisson problem is denoted as $f(x)$, which we assume to be square integrable, i.e., $f \in L^2(\Omega)$. Our objective is to solve this Poisson model problem numerically using a spectral method based on high order polynomials.

We recall that the strong form of the Poisson problem can be expressed as:

$$\begin{aligned} -u_{xx} &= f \quad \text{in } \Omega = (-1, 1), \\ u(-1) &= u(1) = 0. \end{aligned}$$

The weak form of the Poisson problem can be stated as: Find $u \in X$ such that

$$a(u, v) = (f, v) \qquad \forall v \in X. \tag{1}$$

Here the solution space $X = \{v \in H^1(\Omega),\ v(\pm 1) = 0\} \equiv H_0^1(\Omega)$ (which is infinite-dimensional), $a(\cdot, \cdot)$ is the symmetric, positive-definite bilinear form

$$a(w, v) = \int_{-1}^{1} w_x v_x \ \mathrm{d}x, \tag{2}$$

and (f, v) is the usual L^2 innerproduct

$$(f, v) = \int_{-1}^{1} f v \ \mathrm{d}x. \tag{3}$$

Our discretization will be based on the weak form. We define our discrete space X_N as

$$X_N = X \cap \mathbb{P}_N(\Omega), \tag{4}$$

where $\mathbb{P}_N(\Omega)$ is the space of all functions which are polynomials of degree N or less over Ω.

Note that $\mathbb{P}_N$ is a finite-dimensional space with dimension $N+1$ (one needs $N+1$ parameters to define a polynomial of degree N), while X_N is a finite-dimensional subspace of X of dimension $N - 1$ (we "loose" two degrees-of-freedom compared to $\mathbb{P}_N$ due to the essential boundary conditions at $x = \pm 1$).

A discrete problem can then be stated as: Find $u_N \in X_N$ such that

$$a(u_N, v) = (f, v) \qquad \forall v \in X_N. \tag{5}$$

1.1 *A priori* analysis

Already at this point we can say something about the properties of the numerical solution. First, we recall that the bilinear form $a(\cdot,\cdot)$ is symmetric and positive definite. Symmetry means that

$$a(w,v) = a(v,w) \quad \forall v,w \in X,$$

while positive definiteness means that

$$a(w,w) > 0 \quad \forall w \in X, w \neq 0.$$

Since $a(\cdot,\cdot)$ is symmetric and positive definite, we can use $a(\cdot,\cdot)$ to define an inner product and an associated norm. We denote this norm the *energy norm*; for all $v \in X$, the energy norm of v is defined as

$$|||v||| = (a(v,v))^{1/2}. \tag{6}$$

Note that the energy norm is *problem dependent*; for our Poisson problem, we observe that

$$|||v|||^2 = a(v,v) = \int_1^1 v_x^2 \,\mathrm{d}x = |v|^2_{H^1(\Omega)}. \tag{7}$$

Hence, for our problem, the energy norm is equal to the H^1 semi-norm.

When $a(\cdot,\cdot)$ is symmetric and positive definite (as in our case), we can express this in a more specific way. We say that $a(\cdot,\cdot)$ is *coercive* (or elliptic) if there exists a positive constant α such that

$$a(v,v) \geq \alpha ||v||^2_{H^1}, \qquad \forall v \in X. \tag{8}$$

The constant α is sometimes referred to as the coercivity constant.

Let us now recall a standard orthogonality result. Since (1) is true for any $v \in X$, it is certainly true for any $v_N \in X_N$ since $X_N \subset X$, i.e.,

$$a(u, v_N) = (f, v_N) \qquad \forall v_N \in X_N. \tag{9}$$

We now subtract (5) from (9) and use the linearity in the first argument of the bilinear form $a(\cdot,\cdot)$ to arrive at

$$a(u - u_N, v_N) = 0 \qquad \forall v_N \in X_N. \tag{10}$$

In orther words, the discretization error $u - u_N$ is othogonal to all element in X_N, using the symmetric, positive definite bilinear form $a(\cdot,\cdot)$ to measure orthogonality (also denoted as a-orthogonality).

Next, consider the function $w_N = u_N + v_N \in X_N$, i.e., the discrete solution u_N perturbed by any function v_N in X_N. Then, $\forall v_N \in X_N$, $v_N \neq 0$,

$$\begin{aligned}
a(u - w_N, u - w_N) &= a(u - (u_N + v_N), u - (u_N + v_N)) \\
&= a(u - u_N, u - u_N) - 2a(u - u_N, v_N) + a(v_N, v_N) \\
&= a(u - u_N, u - u_N) + a(v_N, v_N) \\
&> a(u - u_N, u - u_N).
\end{aligned}$$

We have here exploited the othogonality result (10). Another way to state this result is

$$|||u - w_N||| \geq |||u - u_N||| \quad \forall v_N \in X_N, \tag{11}$$

or

$$|||u - u_N||| = \inf_{w_N \in X_N} |||u - w_N|||. \tag{12}$$

In terms of words, we can say that the discrete solution u_N represents the closest we can come to the exact solution u over all possible elements in the discrete space X_N when we measure "distance" in the energy norm. Hence, this result is also referred to as an *optimality* results. In fact, this type of result is true for any choice of discrete space as long as the discrete space is a subspace of X.

Let us now try to quantify how the error $u - u_N$ behaves as a function of our discretization parameter, which in our case is the polynomial degree, N. We start from the optimality result (12) and observe that

$$\begin{aligned}
|||u - u_N||| &= \inf_{w_N \in X_N} |||u - w_N||| & (13) \\
&\leq |||u - I_N u||| = |u - I_N u|_{H^1} \leq ||u - I_N u||_{H^1}, & (14)
\end{aligned}$$

where $I_N u$ is a high order polynomial interpolant of u ($I_N u \in \mathbb{P}_N$). The first inequality follows from the fact that $I_N u$ cannot approximate u better than u_N. In addition, we have used the equivalence between the energy norm and the H^1 semi-norm for our Poisson problem. If we now choose a good interpolant, e.g., the high order interpolant based on the Gauss-Lobatto Legendre points, we expect quite sharp bounds. In particular, we know that

$$||u - I_N u||_{H^1} \leq c N^{1-\sigma} ||u||_{H^\sigma} \tag{15}$$

when $u \in H^\sigma$; here, c is a positive constant. We can also use the coercivity condition (8) to arrive at the *a priori* error estimate

$$\alpha^{1/2} ||u - u_N||_{H^1} \leq |||u - u_N||| \leq ||u - I_N u||_{H^1} \leq c N^{1-\sigma} ||u||_{H^\sigma} \tag{16}$$

or

$$||u - u_N||_{H^1} \leq \frac{c}{\alpha^{1/2}} N^{1-\sigma} ||u||_{H^\sigma}. \tag{17}$$

Hence, we expect rapid convergence to the exact solution as long as u has a high degree of regularity. In the particular case when u is analytic ($\sigma \to \infty$), we expect exponential convergence as N increases. Note that this *a priori* estimate is only a constant away from the best approximation error, i.e., the closest we could come to u over all candidates in X_N (measured in the H^1 norm), even if we new the exact solution u.

2 Implementation

As a basis for the space $\mathbb{P}_N(\Omega)$ we choose the high-order Lagrangian interpolants $\ell_j(x)$, $j = 0, \ldots, N$ through the $N + 1$ Gauss-Lobatto Legendre (GLL) points ξ_j, $j = 0, \ldots, N$, i.e.,

$$\mathbb{P}_N(\Omega) = \mathrm{span}\{\ell_0(x),\, \ell_1(x), \ldots, \ell_N(x)\}. \tag{18}$$

We recall that

$$\ell_j(x) \in \mathbb{P}_N(\Omega), \qquad j = 0, \ldots, N, \tag{19}$$

and that

$$\ell_j(\xi_i) = \delta_{ij}, \qquad i = 0, \ldots, N, \quad j = 0, \ldots, N. \tag{20}$$

See Figure 1 for a plot of one of these basis functions.

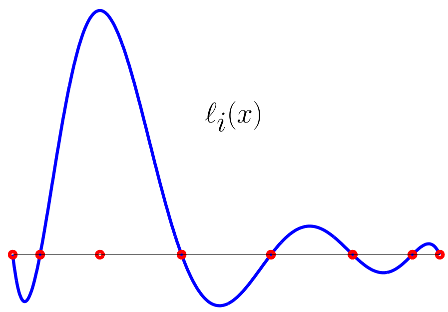

Figure 1: A plot of one of the Lagrangian interpolants, $\ell_i(x)$, $i = 0, \ldots, N$, defined on the interval $[-1, 1]$. The basis function is a polynomial of degree N; the plot shows the particular case when $N = 7$ and $i = 2$. All the associated $N + 1 = 8$ Gauss-Lobatto Legendre points ξ_j, $j = 0, \ldots, N$, are indicated as red circles. Note that $\ell(\xi_i) = 1$, while $\ell_i(\xi_j) = 0$ for $j \neq i$.

A natural basis for X_N is then given as

$$X_N = \mathrm{span}\{\ell_1(x),\, \ell_2(x), \ldots, \ell_{N-1}(x)\}, \tag{21}$$

i.e., we have removed the first and the last basis functions in $\mathbb{P}_N(\Omega)$.

With this basis for X_N, we can now express the discrete solution $u_N(x)$ as

$$u_N(x) = \sum_{j=1}^{N-1} u_j \ell_j(x), \tag{22}$$

where u_j is the basis coefficient associated with the basis function $\ell_j(x)$. Given the properties of our basis functions (Lagrangian interpolants), it follows that the basis coefficient u_j is equal to the value of the discrete solution $u_N(x)$ at the GLL point (or node) ξ_j, i.e., $u_j = u_N(\xi_j)$. Due to this property, we also say that we are using a *nodal* basis.

Inserting (22) into (5), we obtain

$$a(\sum_{j=1}^{N-1} u_j \ell_j(x), v) = (f, v) \qquad \forall v \in X_N. \tag{23}$$

In order to fulfil the condition "$\forall v \in X_N$", we choose systematically $v(x)$ to be one of the basis functions in X_N, i.e., we choose $v(x) = \ell_i(x)$, $i = 1, \ldots, N-1$. In this way, we get the $N-1$ conditions

$$a(\sum_{j=1}^{N-1} u_j \ell_j,\, \ell_i) = (f, \ell_i), \qquad i = 1, \ldots, N-1. \tag{24}$$

Using the bilinearity of $a(\cdot,\cdot)$ (linearity in the first argument and linearity in the second argument), as well as the symmetry property of $a(\cdot,\cdot)$, we can express the discrete problem as: Find $\underline{u} \in \mathbb{R}^{N-1}$ such that

$$\sum_{j=1}^{N-1} a(\ell_i,\, \ell_j)\, u_j = (f, \ell_i), \qquad i = 1, \ldots, N-1, \tag{25}$$

where $\underline{u}$ is a vector of dimension $N-1$ comprising the unknown basis coefficients u_j, $j = 1, \ldots, N-1$. Note that introducing a specific basis for the discrete space X_N has transformed the discrete problem into an algebraic problem. We could, of course, have chosen a different basis for X_N (the basis is not unique); in this case, the algebraic system of equations would have been different, the basis coefficients would have been different, however, the solution $u_N(x)$ would be the same (the solution only depends on the discrete space X_N). Our particular choice of basis is related to numerical convenience, numerical stability, and computational cost.

At this point, we could succinctly express the discrete problem (25) in matrix form as

$$\underline{A}\,\underline{u} = \underline{b}, \tag{26}$$

where $\underline{A} \in \mathbb{R}^{(N-1)\times(N-1)}$ represents the discrete Laplace operator with matrix elements

$$A_{ij} = a(\ell_i, \ell_j), \qquad i = 1, \ldots, N-1, \quad j = 1, \ldots, N-1, \tag{27}$$

and $\underline{b}$ is a right hand side vector with elements

$$b_i = (f, \ell_i), \qquad i = 1, \ldots, N-1. \tag{28}$$

However, in the following, we will choose to evaluate the bilinear form $a(\cdot,\cdot)$ and the L^2 innerproduct $(\cdot,\cdot)$ (the right hand side) using numerical quadrature. In particular, we will use Gauss-Lobatto Legendre (GLL) quadrature. For an integrand $g(x)$, the integral

$$\int_{-1}^{1} g(x)\,\mathrm{d}x \tag{29}$$

is approximated as

$$\sum_{\alpha=0}^{N} \rho_\alpha g(\xi_\alpha), \tag{30}$$

with exact integration obtained for all $g(x) \in \mathbb{P}_{2N-1}(\Omega)$. Note that the GLL integration rule is realized by sampling the integrand $g(x)$ at the $(N+1)$ GLL points and expressing the integral as a weighted sum of these samples. The GLL weights are denoted as ρ_α, $\alpha = 0, \ldots, N$.

Instead of the discrete problem (25) we will solve the problem: Find $\underline{u} \in \mathbb{R}^{N-1}$ such that

$$\sum_{j=1}^{N-1} a_N(\ell_i,\, \ell_j)\, u_j = (f, \ell_i)_N, \qquad i = 1, \ldots, N-1, \tag{31}$$

where $a_N(\cdot,\cdot)$ and $(\cdot,\cdot)_N$ denote integration of $a(\cdot,\cdot)$ and $(\cdot,\cdot)$ by GLL quadrature, respectively. Hence, our matrix elements and right hand side are given by

$$A_{ij} = a_N(\ell_i, \ell_j) = (\ell_i', \ell_j')_N, \qquad i = 1, \ldots, N-1, \quad j = 1, \ldots, N-1, \tag{32}$$

and

$$b_i = (f, \ell_i)_N, \qquad i = 1, \ldots, N-1, \tag{33}$$

respectively.

Let us now discuss the use of GLL quadrature in our problem. First, consider the matrix element A_{ij},

$$\begin{aligned} a(\ell_i,\ell_j) &= \int_{-1}^{1} \underbrace{\ell_i'(x)\ell_j'(x)}_{\in\mathbb{P}_{2N-2}(\Omega)} \,\mathrm{d}x = a_N(\ell_i,\ell_j) \\ &= \sum_{\alpha=0}^{N} \rho_\alpha \ell_i'(\xi_\alpha)\ell_j'(\xi_\alpha) \\ &= \sum_{\alpha=0}^{N} \rho_\alpha D_{\alpha i} D_{\alpha j} \\ &= A_{ij}. \end{aligned} \tag{34}$$

We note that the GLL quadrature is exact here since the integrand is a polynomial of degree $(2N-2)$ which is less than or equal to $(2N-1)$. Each matrix element in *the derivative matrix* $\underline{D}$ is the derivative of one of the basis functions at one of the GLL points, i.e.,

$$D_{mn} = \ell_n'(\xi_m), \quad 0 \le m,n \le N. \tag{35}$$

For our Poisson problem, due to the dimension of the discrete space X_N, we only need to invoke $N \times (N-2)$ of these matrix elements. We also remark that the matrix $\underline{A}$ is sometimes referred to as *the stiffness matrix*; here, $\underline{A} \in \mathbb{R}^{(N-1)\times(N-1)}$.

The evaluation of the right hand side using GLL quadrature gives

$$\begin{aligned} b_i = (f,\ell_i)_N &= \sum_{\alpha=0}^{N} \rho_\alpha f(\xi_\alpha)\ell_i(\xi_\alpha) \\ &= \sum_{\alpha=0}^{N} \rho_\alpha f(\xi_\alpha)\delta_{i\alpha} \\ &= \rho_i f(\xi_i) \\ &= \rho_i f_i, \end{aligned} \tag{36}$$

where we have defined $f_i = f(\xi_i)$, $i = 0,\ldots,N$. However, note that the right hand side vector elements b_i are only defined for $i = 1,\ldots,N-1$. In principle, we should include information about $f(x)$ over the entire domain including the boundary points because there are no boundary conditions imposed on the given data (the boundary conditions are only imposed on the solution $u(x)$). The only assumption we have made is that $f(x) \in L^2(\Omega)$, i.e., that $f(x)$ is square integrable. Nonetheless, the information about $f_0 = f(\xi_0) = f(-1)$ and $f_N = f(\xi_N) = f(1)$ do not enter into the right hand side (36). This is just related to our choice of basis functions for X_N as Lagrangian interpolants through the GLL points, and the fact that we are also using GLL quadrature.

Let us revisit the derivation of the discrete right hand side vector $\underline{b}$ with elements $b_i = (f, \ell_i)_N$, $i = 1, \ldots, N-1$. Let us now first approximate $f(x)$ as the high order interpolant $I_N f(x)$ through the GLL points, that is, $f_N(x) = I_N f(x) \in \mathbb{P}_N(\Omega)$ and $f_N(\xi_m) = f(\xi_m)$ for $m = 0, \ldots, N$. From the properties of the interpolant and the chosen basis for $\mathbb{P}_N$, it follows immediately that we can write

$$f_N(x) = \sum_{j=0}^{N} f_j \ell_j(x). \tag{37}$$

Note that the sum here goes from 0 to N (there are no restrictions on $f(x)$ at the boundary).

Next, we will show that

$$b_i = (f, \ell_i)_N = (f_N, \ell_i)_N. \tag{38}$$

To this end, we evaluate the expression,

$$\begin{aligned} (f_N, \ell_i)_N &= (\sum_{j=0}^{N} f_j \ell_j, \ell_i)_N \\ &= \sum_{j=0}^{N} (\ell_i, \ell_j)_N f_j \\ &= \sum_{j=0}^{N} (\sum_{\alpha=0}^{N} \rho_\alpha \delta_{\alpha i} \delta_{\alpha j})\, f_j \\ &= \sum_{j=0}^{N} \rho_i \delta_{ij} f_j \\ &= \rho_i f_i, \end{aligned} \tag{39}$$

which is the same result as we derived earlier. Hence, there is no difference in the discrete right hand side whether we use $f(x)$ or the interpolant $f_N(x)$. This result is due to the fact that we are using GLL quadrature instead of exact integration of the L^2 innerproduct.

In the following, we will denote the innerproduct $(\ell_i, \ell_j)_N$ as B_{ij}. We also introduce *the mass matrix* $\underline{B}$ with matrix elements B_{ij}. From the above derivation, we see that

$$B_{ij} = \rho_i \delta_{ij}. \tag{40}$$

The mass matrix is thus diagonal, with the positive GLL weights along the diagonal. As a comparison, the mass matrix when using linear finite elements in one space dimension is a tridiagonal matrix.

To summarize, the discrete Poisson problem is based on the weak form given by (5). By choosing a nodal basis for the discrete space X_N based on the GLL points, and choosing GLL quadrature, we arrive at the algebraic system

$$\underline{A}\,\underline{u} = \underline{B}\,\underline{f}, \tag{41}$$

where the elements in the stiffness matrix $\underline{A}$ are given as

$$A_{ij} = a_N(\ell_i, \ell_j) = (\ell_i', \ell_j')_N = \sum_{\alpha=0}^{N} \rho_\alpha D_{\alpha i} D_{\alpha j}, \tag{42}$$

and the elements in the mass matrix $\underline{B}$ are given as

$$B_{ij} = \rho_i \delta_{ij}. \tag{43}$$

The stiffness matrix is full, while the mass matrix is diagonal. Here $\underline{f}$ is a vector of the given data $f(x)$ sampled at the (internal) GLL points.

Once the algebraic system has been solved for the unknown basis coefficients $\underline{u} \in \mathbb{R}^{N-1}$, the discrete solution $u_N(x)$ is given at *any* value of x through the basis

$$u_N(x) = \sum_{j=1}^{N-1} u_j \ell_j(x). \tag{44}$$

3 Discretization error including quadrature

Without quadrature errors, the discrete solution should only be a constant away from the best approximation error, i.e., we should have the following *a priori* error estimate (see (17))

$$||u - u_N||_{H^1(\Omega)} \leq \tilde{c}\, N^{1-\sigma} ||u||_{H^\sigma(\Omega)} \tag{45}$$

for $u \in H^\sigma(\Omega)$ (here $\tilde{c}$ is a constant).

With GLL quadrature, the left hand side of (31) is exact, while the error in the right hand side can be interpreted as an interpolation error of $f(x)$. The discretization error will in this case have two contributions, one being a constant times the best approximation error associated with approximating $u(x)$ using high order polynomials, and one being a constant times the interpolation error associated with interpolating $f(x)$ using high order polynomials. The *a priori* error estimate can in this case be expressed as

$$||u - u_N||_{H^1(\Omega)} \leq c_1\, N^{1-\sigma} ||u||_{H^\sigma(\Omega)} + c_2\, N^{-\rho} ||f||_{H^\rho(\Omega)} \tag{46}$$

for $u \in H^\sigma(\Omega)$ and $f \in H^\rho(\Omega)$ (here c_1 and c_2 are constants).

Exercises

1. Consider the Poisson problem (1) with $f(x) = x^3$. We discretize this problem using a discrete space X_N as defined in (4) with $N = 5$ and evaluation of the bilinear and linear forms using GLL quadrature as described in the notes. What will the discretization error be in this case?

2. Prove that the stiffness matrix $\underline{A}$ and the mass matrix $\underline{B}$ both are symmetric and positive definite.

3. Consider the eigenvalue problem $-u_{xx} = \lambda u \quad \text{in } \Omega = (-1, 1)$, with $u(-1) = u(1) = 0$. Derive the weak form for this problem and discretize it using high order polynomial approximations of the eigenfunctions. Express the discrete problem in matrix form. What do you expect to get for the smallest eigenvalue? Will it be positive?

4. Consider the Helmholtz problem $-u_{xx} + u = f \quad \text{in } \Omega = (-1, 1)$, with $u(-1) = u(1) = 0$. Derive the weak form for this problem. In particular, identify the bilinear and linear forms for this problem. What is the energy norm associated with this problem?

5. Solve the Poisson problem $-u_{xx} = f \quad \text{in } \Omega = (-1, 1)$, with homogeneous Dirichlet boundary conditions $u(-1) = u(1) = 0$. Choose for example the exact solution to be $u(x) = (1 - x^2)\, e^x$. From $u(x)$, find the corresponding "source term" $f(x)$. Solve the Poisson problem using a spectral discretization as discussed in class. You need to construct the right hand side ($b_i = \rho_i f_i$) as well as the system matrix ($A_{ij} = \sum_\alpha \rho_\alpha D_{\alpha i} D_{\alpha j}$). You can use the available MATLAB functions to compute the quadrature weights and the differentiation matrix. Solve $\underline{A}\,\underline{u} = \underline{b}$ for $\underline{u}$. The basis coefficient u_i will be an approximation to the exact value $u(\xi_i)$. Compute the maximum pointwise error at the internal GLL points: $E = \max_{i=1}^{N-1} |u(\xi_i) - u_i|$. Plott $\log(E)$ as a function of N. Do you observe exponential convergence?