# CS 450: Numerical Anlaysis[1]

## Boundary Value Problems for Ordinary Differential Equations

University of Illinois at Urbana-Champaign

[1] *These slides have been drafted by Edgar Solomonik as lecture templates and supplementary material for the book "Scientific Computing: An Introductory Survey" by Michael T. Heath (slides).*

# Boundary Value Problems for ODEs

- Often we seek to solve a differential equation that satisfies conditions on its values and derivatives on parts of the domain boundary. Consider a first order ODE $\boldsymbol{y}'(t) = \boldsymbol{f}(t, \boldsymbol{y})$ with general *linear boundary conditions*:

$$\boldsymbol{B}_a \boldsymbol{y}(a) + \boldsymbol{B}_b \boldsymbol{y}(b) = \boldsymbol{c}$$

  - *for IVP, simple case of Dirichlet (value) condition* $\boldsymbol{B}_a = \boldsymbol{I}$, $\boldsymbol{B}B_b = \boldsymbol{0}$
  - *conditions are separate if* $\boldsymbol{B}_a \neq \boldsymbol{0}$ *and* $\boldsymbol{B}_b \neq \boldsymbol{0}$

- High-order boundary conditions can be reduced to first-order like ODEs themselves:

  *The Neumann boundary condition gives* $\boldsymbol{y}'(a)$*, which simply implies we have* $\boldsymbol{u}_2(a) = \boldsymbol{y}'(a)$ *if we use* $\boldsymbol{u}_1 = \boldsymbol{y}$ *and* $\boldsymbol{u}_2 = \boldsymbol{y}'$*.*

# Boundary Value Problems for ODEs

- Can derive the solutions to a linear ODE BVP $\boldsymbol{y}'(t) = \boldsymbol{A}(t)\boldsymbol{y}(t) + \boldsymbol{b}(t)$ from solutions to homogeneous linear ODE $\boldsymbol{y}' = \boldsymbol{A}(t)\boldsymbol{y}(t)$ IVPs:

  *Let the solution to the homogeneous ODE IVP with $\boldsymbol{y}_i(a) = \boldsymbol{e}_i$ be given by the $i$th column of $\boldsymbol{Y}(t)$, so*

  $$\boldsymbol{Y}(t) = \boldsymbol{I} + \int_a^t \boldsymbol{A}(s)\boldsymbol{Y}(s)ds.$$

  *Now, we seek to derive a solution of the form $\boldsymbol{y}(t) = \boldsymbol{Y}(t)\boldsymbol{u}(t)$ with $\boldsymbol{u}(0) = \boldsymbol{y}(a)$, which satisfies the given boundary condition,*

  $$\boldsymbol{B}_a\boldsymbol{Y}(a)\boldsymbol{y}(a) + \boldsymbol{B}_b\boldsymbol{Y}(b)\boldsymbol{y}(a) + \boldsymbol{B}_b\boldsymbol{Y}(b)\int_a^b \boldsymbol{u}'(s)ds = \boldsymbol{c}$$

  $$\underbrace{(\boldsymbol{B}_a\boldsymbol{Y}(a) + \boldsymbol{B}_b\boldsymbol{Y}(b))}_{\boldsymbol{Q}}\boldsymbol{y}(a) = \boldsymbol{c} - \boldsymbol{B}_b\boldsymbol{Y}(b)\int_a^b \boldsymbol{u}'(s)ds,$$

  *the existence of solution $\boldsymbol{y}$, thus generally depends on invertibility of $\boldsymbol{Q}$.*

# Boundary Value Problems for ODEs

- Can derive the solutions to a linear ODE BVP $\boldsymbol{y}'(t) = \boldsymbol{A}(t)\boldsymbol{y}(t) + \boldsymbol{b}(t)$ from solutions to homogeneous linear ODE $\boldsymbol{y}' = \boldsymbol{A}(t)\boldsymbol{y}(t)$ IVPs:

  *To determine* $\boldsymbol{u}(t) = \boldsymbol{y}(a) + \int_a^t \boldsymbol{u}'(s)ds$*, we use the differential equation*

$$\begin{aligned}\boldsymbol{A}(t)\boldsymbol{y}(t) + \boldsymbol{b}(t) = \boldsymbol{y}'(t) &= \boldsymbol{Y}'(t)\boldsymbol{u}(t) + \boldsymbol{Y}(t)\boldsymbol{u}'(t)\\ &= \boldsymbol{A}(t)\underbrace{\boldsymbol{Y}(t)\boldsymbol{u}(t)}_{\boldsymbol{y}(t)} + \boldsymbol{Y}(t)\boldsymbol{u}'(t),\end{aligned}$$

$$\boldsymbol{u}'(t) = \boldsymbol{Y}^{-1}(t)\boldsymbol{b}(t).$$

  *Thus, given* $\boldsymbol{u}(t) = \boldsymbol{y}(a) + \int_a^t \boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)ds$*, the overall solution is*

$$\begin{aligned}\boldsymbol{y}(t) &= \boldsymbol{Y}(t)\boldsymbol{u}(t)\\ &= \boldsymbol{Y}(t)\underbrace{\boldsymbol{Q}^{-1}\left(\boldsymbol{c} - \boldsymbol{B}_b\boldsymbol{Y}(b)\int_a^b \underbrace{\boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)}_{\boldsymbol{u}'(s)}ds\right)}_{\boldsymbol{y}(a)} + \boldsymbol{Y}(t)\int_a^t \underbrace{\boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)}_{\boldsymbol{u}'(s)}ds\end{aligned}$$

# Linear ODE BVP Green's Function

- We now express our solution (with form $\boldsymbol{y}(t) = \boldsymbol{Y}(t)(\boldsymbol{u}(a) + \int_a^t \boldsymbol{u}'(s)ds)$) in the form $\boldsymbol{y}(t) = \boldsymbol{s}(t) + \int_a^b \boldsymbol{G}(t,s)\boldsymbol{b}(s)ds$ where $\boldsymbol{G}$ is the *Green's function*:
  *Using the fact that,*

$$\boldsymbol{Y}(t)\boldsymbol{Q}^{-1}\underbrace{(\boldsymbol{B}_a\boldsymbol{Y}(a) + \boldsymbol{B}_b\boldsymbol{Y}(b))}_{\boldsymbol{Q}}\int_a^t \boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)ds = \boldsymbol{Y}(t)\int_a^t \boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)ds,$$

  *we can express our previous solution in the form,*

$$\boldsymbol{y}(t) = \boldsymbol{Y}(t)\boldsymbol{Q}^{-1}\left(\boldsymbol{c} + \boldsymbol{B}_a\boldsymbol{Y}(a)\int_a^t \boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)ds - \boldsymbol{B}_b\boldsymbol{Y}(b)\int_t^b \boldsymbol{Y}^{-1}(s)\boldsymbol{b}(s)ds\right),$$

  *which allows us to derive the Green's function,*

$$\boldsymbol{G}(t,s) = \boldsymbol{Y}(t)\boldsymbol{Q}^{-1}\boldsymbol{I}(s)\boldsymbol{Y}^{-1}(s), \quad \boldsymbol{I}(s) = \begin{cases} \boldsymbol{B}_a\boldsymbol{Y}(a) & : s < t \\ \boldsymbol{B}_b\boldsymbol{Y}(b) & : s \geq t \end{cases}$$

# Conditioning of Linear ODE BVPs

- For any given $\boldsymbol{b}(t)$ and $\boldsymbol{c}$, the solution to the BVP can be written in the form:

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(t)\boldsymbol{c} + \int_a^b \boldsymbol{G}(t,s)\boldsymbol{b}(s)ds$$

  $\boldsymbol{\Phi}(t) = \boldsymbol{Y}(t)\boldsymbol{Q}^{-1}$ *is the fundamental matrix, which like the Green's function is associated with the homogeneous ODE as well as its linear boundary condition matrices* $\boldsymbol{B}_a$ *and* $\boldsymbol{B}_b$*, but is independent* $\boldsymbol{b}(t)$ *and* $\boldsymbol{c}$*.*

- The absolute condition number of the BVP is $\kappa = \max\{||\boldsymbol{\Phi}||_\infty, ||\boldsymbol{G}||_\infty\}$:

  *This sensitivity measure enables us to bound the perturbation* $||\hat{\boldsymbol{y}} - \boldsymbol{y}||_\infty$ *with respect to the magnitude of a perturbation to* $\boldsymbol{b}(t)$ *or* $\boldsymbol{c}$*.*

# Shooting Method for ODE BVPs

- For linear ODEs, we constructed solutions from IVP solutions in $\boldsymbol{Y}(t)$, which suggests a method for solving BVPs by reduction to IVPs:

  *The* *shooting* *method iteratively (for* $k = 1, 2, \ldots$*) constructs approximate initial value guesses* $\hat{\boldsymbol{y}}^{(k)}(a) \approx \boldsymbol{y}(a)$*, solves the resulting IVP, and checks the quality of the solution at the new boundary,*

  $$||\boldsymbol{B}_b \hat{\boldsymbol{y}}^{(k)}(b) - \boldsymbol{B}_a \hat{\boldsymbol{y}}^{(k)}(a) - \boldsymbol{c}||,$$

  *the initial conditions for the next shot,* $\hat{\boldsymbol{y}}^{(k+1)}(a)$ *can be constructed by a 1D root finding technique on*

  $$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{B}_a \boldsymbol{x} + \boldsymbol{B}_b \boldsymbol{y}_{\boldsymbol{x}}(b) - \boldsymbol{c}, \text{ where } \boldsymbol{y}_{\boldsymbol{x}}(b) \text{ is the IVP Solution with } \boldsymbol{y}_{\boldsymbol{x}}(a) = \boldsymbol{x}.$$

- *Multiple shooting* employs the shooting method over subdomains:

  *Conditioning of shooting method depends on stability of IVPs, which can be worse than conditioning of the BVP. However, the shooting problems are interdependent, as they must satisfy continuity conditions on boundaries between them, leading to a system of nonlinear equations.*

# Finite Difference Methods

- Rather than solve a sequence of IVPs that satisfy the ODEs until they (approximately) satisfy boundary conditions, we can refine an approximation that satisfies the boundary conditions, until it satisfies the ODE:

  *Finite difference methods works by obtaining a solution on points* $t_1, \dots, t_n$*, so that* $\hat{\boldsymbol{y}}_k \approx \boldsymbol{y}(t_k)$ *by finite-difference formulae, for example*

$$\boldsymbol{f}(t, \boldsymbol{y}) = \boldsymbol{y}'(t) \approx \frac{\boldsymbol{y}(t+h) - \boldsymbol{y}(t-h)}{2h} \Rightarrow \boldsymbol{f}(t_k, \hat{\boldsymbol{y}}_k) = \frac{\hat{\boldsymbol{y}}_{k+1} - \hat{\boldsymbol{y}}_{k-1}}{t_{k+1} - t_{k-1}}$$

  *the resulting system of equations can be solved by standard methods and is linear if* $\boldsymbol{f}$ *is linear.*

- Convergence to solution is obtained with decreasing step size $h$ so long as the method is consistent and stable:

  *Consistency implies that the truncation error goes to zero, while stability ensures input perturbations have bounded effect on solution.*

# Finite Difference Methods

- Lets derive the finite difference method for the ODE BVP defined by

$$u'' + 1000(1+t^2)u = 0$$

with boundary conditions $u(-1) = 3$ and $u(1) = -3$.

*Using a discretization with points $t_1, \dots, t_n$, $t_{i+1} - t_i = h$, and a centered difference approximation for $u''$ we obtain*

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + 1000(1+t_i)u_i = 0.$$

*We can rewrite the above using linear equations with matrices*

$$\boldsymbol{A} = \begin{bmatrix} 1 & & & & \\ 1/h^2 & -2/h^2 & 1/h^2 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/h^2 & -2/h^2 & 1/h^2 \\ & & & & 1 \end{bmatrix} \quad \textit{and} \quad \boldsymbol{B} = \begin{bmatrix} 0 & & & \\ 0 & 1000(1+t_2) & & \\ & & \ddots & & \\ & & & 1000(1+t_{n-1}) & 0 \\ & & & & 0 \end{bmatrix}$$

*and solve the system* $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{u} = \begin{bmatrix} 3 & 0 & \cdots 0 & -3 \end{bmatrix}^T$.

# Collocation Methods

- *Collocation methods* approximate $\boldsymbol{y}$ by representing it in a basis

$$\boldsymbol{y}(t) = \boldsymbol{v}(t, \boldsymbol{x}) = \sum_{i=1}^{n} x_i \boldsymbol{\phi}_i(t).$$

*To construct equations, consider approximation for a set of collocation points $t_1, \ldots, t_n$ with $t_1 = a$ and $t_n = b$,*

$$\forall_{i \in \{2,\ldots,n-1\}} \quad \boldsymbol{v}(t_i, \boldsymbol{x}) = \boldsymbol{f}(t_i, \boldsymbol{v}(t_i, \boldsymbol{x})),$$

*with two more equations typically obtained from boundary conditions at $t_1, t_n$.*

- *Spectral methods* use polynomials or trigonometric functions for $\phi_i$, which are nonzero over most of $[a, b]$, while *finite element* methods leverage basis functions with local support (e.g. B-splines).

  *Eigenfunctions of differential operators are typically trigonometric functions or polynomials, hence the name "spectral methods".*

# Solving BVPs by Optimization

- We reformulate the collocation approximation as an optimization problem: *Consider the simplified scenario* $\boldsymbol{f}(t,y) = \boldsymbol{f}(t)$ *with residual equation,*

$$\boldsymbol{r}(t,\boldsymbol{x}) = \boldsymbol{v}'(t,\boldsymbol{x}) - \boldsymbol{f}(t) = \sum_{j=1}^{n} x_j \boldsymbol{\phi}_j'(t) - \boldsymbol{f}(t)$$

  *and minimize it using the objective function,*

$$F(\boldsymbol{x}) = \frac{1}{2}\int_a^b ||\boldsymbol{r}(t,\boldsymbol{x})||_2^2 dt.$$

- The first-order optimality conditions of the optimization problem are a system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{aligned}\mathbf{0} = \frac{dF}{dx_i} &= \int_a^b \boldsymbol{r}(t,\boldsymbol{x})^T \frac{d\boldsymbol{r}}{dx_i} dt = \int_a^b \boldsymbol{r}(t,\boldsymbol{x})^T \boldsymbol{\phi}_i'(t) dt \\ &= \sum_{j=1}^{n} x_j \underbrace{\int_a^b \boldsymbol{\phi}_j'(t)^T \boldsymbol{\phi}_i'(t) dt}_{a_{ij}} - \underbrace{\int_a^b \boldsymbol{f}(t)^T \boldsymbol{\phi}_i'(t) dt}_{b_i}\end{aligned}$$

# Weighted Residual

- *Weighted residual methods* work by ensuring the residual is orthogonal with respect to a given set of weight functions:
  *Rather than setting components of the gradient to zero, we instead have,*
  $$\int_a^b \boldsymbol{r}(t, \boldsymbol{x})^T \boldsymbol{w}_i(t) dt = 0, \forall i \in \{1, \dots, n\},$$
  *which again yields a system of equations of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where*
  $$a_{ij} = \int_a^b \boldsymbol{\phi}_j'(t)^T \boldsymbol{w}_i(t), \quad b_i = \int_a^b \boldsymbol{f}(t)^T \boldsymbol{w}_i(t).$$
  *The collocation method is a weighted residual method where $\boldsymbol{w}_i(t) = \boldsymbol{\delta}(t - t_i)$.*
- The Galerkin method is a weighted residual method where $\boldsymbol{w}_i = \boldsymbol{\phi}_i$.
  *Linear system with the stiffness matrix $\boldsymbol{A}$ and load vector $\boldsymbol{b}$ is*
  $$\boldsymbol{0} = \sum_{j=1}^{n} x_j \underbrace{\int_a^b \boldsymbol{\phi}_j'(t)^T \boldsymbol{\phi}_i(t) dt}_{a_{ij}} - \underbrace{\int_a^b \boldsymbol{f}(t)^T \boldsymbol{\phi}_i(t) dt}_{b_i}.$$

# Linear BVPs by Optimization

- Lets apply the Galerkin method to the more general linear ODE $\boldsymbol{f}(t, y) = \boldsymbol{A}(t)\boldsymbol{y}(t) + \boldsymbol{b}(t)$ with residual equation,
  *First, choose basis functions $\{\phi_i\}_{i=1}^n$ to satisfy the boundary conditions, so solution automatically satisfies them, then minimize the residual,*

$$\boldsymbol{r} = \boldsymbol{v}' - \boldsymbol{A}\boldsymbol{v} - \boldsymbol{b}, \text{ so that } \boldsymbol{r}(t, \boldsymbol{x}) = \sum_{j=1}^{n} x_j(\boldsymbol{\phi}_j'(t) - \boldsymbol{A}(t)\boldsymbol{\phi}_j(t)) - \boldsymbol{b}(t).$$

  *The Galerkin method, minimizes the residual by orthogonality with respect to a set of test functions that is the same as the set of basis functions,*

$$\begin{aligned} \boldsymbol{0} &= \int_a^b \boldsymbol{r}(t, \boldsymbol{x})^T \boldsymbol{\phi}_i(t) dt \\ &= \sum_{j=1}^{n} x_j \int_a^b (\boldsymbol{\phi}_j'(t) - \boldsymbol{A}(t)\boldsymbol{\phi}_j(t))^T \boldsymbol{\phi}_i(t) dt - \int_a^b \boldsymbol{b}(t)^T \boldsymbol{\phi}_i(t) dt \end{aligned}$$

# Nonlinear BVPs: Poisson Equation

In practice, BVPs are at least second order and its advantageous to work in the natural set of variables.

- Consider the Poisson equation $u'' = f(t)$ with boundary conditions $u(a) = u(b) = 0$ and define a localized basis of hat functions:

$$\phi_i(t) = \begin{cases} (t - t_{i-1})/h & : t \in [t_{i-1}, t_i] \\ (t_{i+1} - t)/h & : t \in [t_i, t_{i+1}] \\ 0 & : \text{otherwise} \end{cases}$$

  where $t_0 = t_1 = a$ and $t_{n+1} = t_n = b$.
  *Trying to define the residual equation as usual, we obtain*

$$r = v'' - f, \text{ so that } r(t, \boldsymbol{x}) = \sum_{j=1}^{n} x_j \phi_j''(t) - f(t).$$

  *However, $\phi_j''(t)$ is undefined, since $\phi_j'(t)$ is discontinuous at $t_{j-1}, t_j, t_{j+1}$.*

# Weak Form and the Finite Element Method

- The finite-element method permits a lesser degree of differentiability of basis functions by casting the ODE in *weak form*:
  *For any solution $u$, if test function $\phi_i$ satisfies the boundary conditions, the ODE satisfies the weak form,*

$$\int_a^b f(t)\phi_i(t)dt = \int_a^b u''(t)\phi_i(t)dt = u'(b)\underbrace{\phi_i(b)}_{0} - u'(a)\underbrace{\phi_i(a)}_{0} - \int_a^b u'(t)\phi_i'(t)dt$$

$$= -\int_a^b u'(t)\phi_i'(t)dt.$$

  *Note that the final equation contains no second derivatives, and subsequently we can form the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ with,*

$$a_{ij} = -\int_a^b \phi_j'(t)\phi_i'(t)dt, \quad b_i = \int_a^b f(t)\phi_i(t)dt.$$

  *The finite element method thus searches the larger (once-differentiable) function space to find a solution $u$ that is in a (twice-differentiable) subspace.*

# Finite Element Methods in Practice

- Hat functions are linear instances of *B-splines*:

  *B-splines of degree $k$ are $k$-times differentiable. For higher-order ODEs or high-order convergence with $h$, its necessary to use $k > 1$.*

- Finite-element methods readily generalize to PDEs:

  *In its most basic form each element corresponds to a triangle (2D) or quadrilateral (3D).*

# Eigenvalue Problems with ODEs

- A typical second-order scalar BVP eigenvalue problem has the form

$$u'' = \lambda f(t, u, u'), \quad \text{with boundary conditions } \; u(a) = 0, u(b) = 0$$

*Lets first consider $f(t, u, u') = g(t)u$, in which case we can approximate the solution at a set of points $t_1, \dots, t_n$ using finite differences,*

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = \lambda g_i y_i,$$

*which corresponds to a tridiagonal matrix eigenvalue problem $\boldsymbol{A}\boldsymbol{y} = \lambda \boldsymbol{y}$ via*

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{g_i h^2} = \lambda y_i.$$

# Eigenvalue Problems with ODEs

- Generalized eigenvalue problems arise from more sophisticated ODEs,

$$u'' = \lambda(g(t)u + h(t)u'), \quad \text{with boundary conditions } u(a) = 0, u(b) = 0$$

*We can approximate each of the derivatives at a set of points* $t_1, \ldots, t_n$ *using finite differences,*

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = \lambda\left(g_i + \frac{y_{i+1} - y_{i-1}}{2h}\right)y_i.$$

*which can be expressed as a generalized matrix eigenvalue problem*

$$\boldsymbol{A}\boldsymbol{y} = \lambda\boldsymbol{B}\boldsymbol{y}$$

*where both* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are tridiagonal.*